

ADVISORY COMMITTEE ON EVIDENCE RULES

**SYMPOSIUM ON FORENSIC EXPERT
TESTIMONY, *DAUBERT*, AND RULE 702***

MODERATOR

Professor Daniel J. Capra

CHAIR OF RULES COMMITTEE

Hon. David Campbell

CHAIR OF ADVISORY COMMITTEE

Hon. Debra A. Livingston

FORMER CHAIR OF ADVISORY COMMITTEE

Hon. William Sessions, III

PANELISTS

Dr. Thomas Albright

Prof. Ronald J. Allen

Susan Ballou

Dr. Itiel Dror

Chris Fabricant, Esq.

Anne Goldbach, Esq.

Andrew Goldsmith, Esq.

Hon. Paul W. Grimm

Zachary Hafer, Esq.

Ted R. Hunt, Esq.

Dr. Alice Isenberg

Dr. Karen Kafadar

Prof. David H. Kaye

Prof. Jonathan J. Koehler

Hon. Alex Kozinski

* This Symposium was held on October 27, 2017, at Boston College School of Law under the sponsorship of the Judicial Conference Advisory Committee on Evidence Rules. The transcript has been lightly edited and represents the panelists' individual views only and in no way reflects those of their affiliated firms, organizations, law schools, or the judiciary.

Dr. Eric Lander
Lori Lightfoot, Esq.
Hon. K. Michael Moore
Prof. Jane Campbell Moriarty
Prof. Erin Murphy
Hon. Jed S. Rakoff
Prof. Stephen A. Saltzburg
Dr. Jeff Salyards
Hon. Patti B. Saris
Laura M. Shamp, Esq.
Thomas M. Sobol, Esq.

WELCOME AND INTRODUCTORY REMARKS

DEAN ROUGEAU: Well, good morning, everyone. Welcome. My name is Vincent Rougeau. I'm the Dean here at Boston College Law School. On behalf of our entire community, I just wanted to tell you how very, very pleased we are to welcome this Symposium to Boston College. We have as one of our core values at this institution a deep, deep commitment to public interest and public service, and this meeting is a great way for us to embody that value and to serve it in a way that we hope will give many people an opportunity to see some of the important work that you're doing, and certainly it's giving our students and our faculty a real opportunity to engage with your important work.

So I wish you all the best for this day. If there is anything you need, there are plenty of people around who are anxious to be of service. But please accept all my best wishes for a productive day. Thanks so very much.

PROFESSOR CAPRA: Thank you. [Applause] I'd like to introduce Judge David Campbell, who is the Chair of the Standing Committee on Rules of Practice and Procedure, for some opening remarks.

JUDGE CAMPBELL: Well, I just want to thank all of you for being here. On behalf of the Rules Committee's process, I can say that we are here to learn and to listen. For me, one of the best parts about being involved in the Rules Committee is that it is a genuine effort at fully informed decision-making. It's a good faith effort to make the system better. And this kind of an event is invaluable to us in evaluating potential changes and making recommendations to others who will make final decisions.

So thank you all for being here, especially to the scientists and the legal experts who will be speaking to us. This is a very valuable event for us, and we are anxious to learn.

PROFESSOR CAPRA: Thank you, Judge Campbell. And I would like to turn it over to the former Chair of the Evidence Rules Committee, Judge Sessions, whose idea it was to start this conference.

JUDGE SESSIONS: Well, thank you, Dan. And I really appreciate everyone being here. Our role in the Advisory Committee is to continue to study and review all of the rules of evidence, and then hopefully, when we

see a difficulty, we make efforts to remedy that. And the best way we have found to do that is first to begin by learning.

So, in 2009, the National Academy of Sciences raised the specter of some difficulty with forensic sciences.¹ Obviously, the President's Council on Scientific and Technical Advisors followed suit.² And, as a result, with those particular articles in mind, we decided this must be a significant area in which review is necessary.

So we are all here to listen. We're all here to learn, although I'm retired, but I still can listen. [Laughter] And so, with that, thank you very much for coming today. [Applause]

PROFESSOR CAPRA: I'm the Reporter to the Committee, and I just want to make a couple introductory remarks. First, I think I think that it's poetic justice that we're here today, and thanks to Dan Coquillette we're here today, because, in this room, the amendments to Rule 702 were approved by the Standing Committee in 1999. So that's coming full circle I guess I would say.

And then I would like to talk for a minute about how this conference kind of came to be. Dr. Eric Lander, who was the cochair of the PCAST Committee, contacted me through Professor Charles Fried of Harvard Law School, about the possibility of some kind of input from the Advisory Committee in PCAST's work, and what the Advisory Committee might be able to do in terms of rulemaking or public comment regarding the concerns about the reliability of forensic evidence. I told Judge Sessions about this, and then Judge Sessions's idea was let's get on it. Let's learn about it.

So, we're here to learn today, and we're here to discuss today. And so our first speaker today is Dr. Eric Lander.

PANEL ONE: FORENSIC EVIDENCE

DR. LANDER: Okay. Fantastic. Thank you, Professor Capra. I really appreciate the fact that the committees have come together to think about this question. I want to thank the committees and their chairs for thinking about this.

So, as Professor Capra already said, I cochaired the President's Council Advisors on Science and Technology, which is the science and technology advisory council to the White House. We worked on a wide variety of different topics over the course of eight years: spectrum policy, cybersecurity, bioweapons, antibiotic resistance, and nanotechnology. But one of the important requests that we got from the White House was to look

1. *See generally* NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009), <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf> [<https://perma.cc/CLW3-Y6VQ>].

2. *See generally* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf [<https://perma.cc/R76Y-7VUJ>].

at forensic science in light of this national academy report, to do a follow-up on it.

So we made a set of recommendations through a variety of federal agencies—NIST, OSTP, FBI, DOJ—but also to the judiciary, and in particular we made a recommendation in Recommendation 9 of our report to this Committee that it was important to look at Rule 702 and understand how science and law interacted.³

So the goal today is not to talk about the details of a 170-odd-page report with 400 footnotes. The goal is really to focus on this one question about forensic feature-comparison methods. Briefly, the methods are this: I have two samples. I look at some features. Are they indistinguishable with respect to the test I did? And if so, how often are samples that come from different sources indistinguishable? Without those two things, a laboratory test and a quantitative interpretation of how surprised I should be that I didn't see a difference, I don't have a method. That's what the method is.

So the question is, what does it mean for such a method to be reliable? Before we get to that, we might ask, why does it matter that a method is reliable? Well, it matters for several reasons. First, forensic feature-comparison evidence is seen by jurors as highly probative. When you ask from studies about juror expectations, they attach numbers like one in a million.

Moreover, the presenters of such evidence usually assert that it's highly probative. Sometimes they say, "I'm not going to attach a number to it, but it doesn't matter." In fact, not attaching a number is worse because people will attach a highly probative number to that.

Also, it can't be independently evaluated by lay jurors because it requires knowledge of the complex methods and knowledge of databases. I can't draw a conclusion about how surprising it is that two fingerprints were indistinguishable without knowing the frequencies of how often that happens, and there's no chance that a lay juror can know that.

In addition, it is not a theoretical problem solely because we know that unreliable forensic feature-comparison evidence has led to wrongful convictions.

Moreover, fourth reason, reliability is required by Rule 702. Rule 702 is explicit in saying that, for admissibility, testimony must be based on reliable principles and methods.⁴ The U.S. Supreme Court has said that the inquiry is flexible, number one.⁵ I'll come back to that. Its overarching subject is scientific validity. And in a case involving scientific evidence, evidentiary reliability will be based on scientific validity.

So the rule sets up the court as a gatekeeper. Relevance is not enough. You can't just bring in evidence because it's relevant. You can't simply rely on cross-examination because the rule requires courts to ensure that the

3. *Id.* at 142–45.

4. FED. R. EVID. 702.

5. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 594 (1993).

evidence is based on reliable methods, and reliability is based on science. In any case, that's how I read the simple language.

And a "flexible inquiry" as I understand it means appropriate to the kind of evidence being examined. It does not mean, as occasionally suggested, whatever the judge feels like.

Yet Rule 702 is not working. Why? Both the leading advisory group to the executive, that is PCAST, and the leading advisory group to the legislature, the National Academy of Sciences, both reached the same conclusion that Rule 702 as applied by the courts is not working to exclude unreliable feature-comparison testimony.

These feature-comparison methods were devised as rough heuristics for criminal investigations. They were not grounded in the practices of scientific validation. Totally fine. And, in fact, we fully endorse the idea that for investigation to formulate a hypothesis, use whatever you want. This discussion refers only to the fact that when you come to court, the method must be reliable.

Hunches are fine for an investigation. But these methods have been accepted by courts despite their validity and accuracy never having been empirically tested in some cases with respect to some methods. Historically, I could characterize the evidence or the claims about these forensic methods as going through three stages: stage one, data-free claims; stage two, contrived shooting-fish-in-a-barrel tests; and stage three, empirical tests of actual accuracy under conditions resembling those for intended use.

The FBI gets the biggest gold star ever for developing the black-box tests of fingerprints that established through real empirical evidence that the method had a measurable reliability. It made mistakes, but it was—their estimate was about 1 in 600. The error bars on that could be as high as 1 in 300, but you could measure it.

Here's roughly the history—and I won't go through details. This is not the PCAST interpretation. I've simply noted what has been said. With regard to fingerprints in 1984, the FBI said in its official document, which it has since disavowed, that fingerprints were infallible.⁶ They have disavowed that since. As late as 2009, a former FBI lab director said the error rate for fingerprints was less than 1 in 11 million, because the FBI had done eleven million fingerprints, and it wasn't aware of any errors. [Laughter] It's cited in the PCAST report.⁷

And relevant to empirical tests, the FBI's big gold star, it did a test. There is a number. You know it, about 1 in 600. But the claims were off by at least 17,000-fold.

For firearms, the early studies that were done were contrived studies. They were shooting-fish-in-a-barrel studies. They came out at a 1 in 5000 error rate. And when actual tests commissioned by the Department of Defense were done, they came out at 1 in 50. That's not a trivial difference.

6. PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 2, 45.

7. *Id.* at 45.

Hair analysis claims in 1974, which the FBI still cites today as the foundation for hair analysis, claimed a 1 in 40,000 error rate. But, actually, PCR DNA tests on hair in 2002 found 1 in 9 times that the FBI had asserted a match that was erroneous.⁸

Then there is bite-mark identification, a 1 in 6 trillion error rate according to the founding articles of the field. Recent studies that actually empirically try merely to ask can you distinguish amongst four possible sources—not everybody, four sources—find a 1 in 6 error rate. That was off by about a trillion-fold.

Footwear, famously, there's a quote of a 1 in 683 billion error rate.⁹ There's actually not a single empirical study that's been done to validate a number. That's bothersome to us, and it should be bothersome to you if these methods must be reliable.

Our key conclusion is that a forensic feature-comparison method can be considered reliable only if it has been empirically tested under conditions appropriate to its intended use and found to have accuracy appropriate to its intended use. This is not a radical conclusion.

Nothing else can substitute for an empirical test. That's what science is about. There are many alternatives people would like to invoke. Well, the method is based on science. Examiners have good professional practices. They have training, certification, accreditation. There are professional organizations, best practices manuals. There's extensive experience in using it, and there are public papers in peer-reviewed journals.

All of these statements are also true for psychics. Psychics have professional organizations, peer-reviewed journals, extensive experience, and they're all good things to have. Don't get me wrong. I like all those things. But they do not establish scientific validity. Over the last 400 years we have all come to agree that scientific validity comes from an empirical test.

This is scientifically obvious, but it somehow meets resistance among some. So I want to pose a question to each and every speaker who comes up. I would ask whether you agree or disagree: "Is empirical evidence of accuracy necessary for a forensic feature-comparison method to be reliable?" I'd ask each speaker to align themselves with yes or no.

If no, explain how in the world something can be considered reliable without testing it and measuring its accuracy. If yes, given what we've seen, Rule 702 has failed in practice to live up to its intended purpose and needs some fixing. We need to clarify our imperative for empirical evidence of sufficient accuracy for intended use, and it must be accompanied by an accurate statement concerning the level of accuracy.

That's it, nine minutes and fifty-eight seconds. And I will stop there.

PROFESSOR CAPRA: That's worth a handshake. [Laughter] Our next speaker is Dr. Itiel Dror.

8. *Id.* at 3.

9. *Id.* at 115.

DR. DROR: I am not a forensic scientist. I am a scientist, but a cognitive neuroscientist, which I won't explain what it is because I won't have much time left. [Laughter]

But I want to point out that I work in multiple expert domains. I work a lot with the U.S. Air Force on pilot decision-making. I work with frontline police, in the medical domain, and in forensic science too. I'm interested in how people make decisions, error rate, reliability, and bias.

Now, if you want reliability—and I have to say I'm talking about reliability from a legal point of view, Federal Rule of Evidence 702, which is very different from reliability when we talk about pure science. What you need to have is objectivity. So we need to talk about instrumentation, objective criteria, quantification.

Some forensic science has instrumentation like toxicology and drug analysis. In terms of objective criteria, there aren't any in many domains. If it's firearms or handwriting or fingerprint, they compare two patterns and decide if they're similar or not to conclude that they come from the same source. And nobody can stand here or in front of PCAST or NAS and tell you what this criteria is. It's subjective and in the eye of the beholder. And that is part of what PCAST said, and good luck with trying to find objective criteria for all the forensic science.

However, and a big however, until and if we find this objective criteria, I don't think we need to throw away the baby with the bathwater. So we can have objectivity or subjectivity. Subjectivity is fine as long as it's unbiased, or we should talk about minimally biased subjectivity so we keep subjectivity as good as it can be. There are many expert domains for which there's no objective instrumentation and quantification, but nevertheless, the decisions are very accurate and reliable.

So, as a cognitive scientist, I'm saying that subjectivity is fine as long as it's based on the actual evidence and data—for example, that there's no circular reasoning, no suspect target bias. In other words, often what people do is look for the suspect in the evidence. It's not the evidence that's driving the forensic comparison. They're working backwards from "here is the suspect. I'm going to look for the suspect in the evidence."

It's the way I do archery. I learned it from the Romans. I take a bow and arrow and I shoot it. It hits a tree, and then I draw the target. Always bull's-eye. I never miss. [Laughter]

And that's how they do it in mind works and rationalization and backward reasoning. So that's an example, and we can use, for example, a procedure called linear sequential unmasking (LSU), which makes the forensic examiner stick to the evidence and see the evidence first. They're not allowed to see the DNA profile of the suspect or the fingerprint of the suspect. They have to see the evidence first and characterize it, and then they see the suspect. So the evidence is driving the process. And, again, if you're doing the sequential unmasking as an example, then subjectivity is fine.

I'll give one more example. We want to make sure that the forensic examiner, when making a subjective judgment, is basing it on relevant information and what they need to do their forensic work. We're not acting

like scientists, and we're not really contributing to the core when the forensic examiners are exposed to a lot of irrelevant information, irrelevant context. They're comparing firearms, and they know that this firearm killed a kid, a white kid by an African American, or whatever context they know. They need to have their information limited to the firearm itself. The problem is that forensic evidence and expert evidence in general can really help the court to administer justice, but it's mainly, in my view, misused and abused because of these issues.

The context causes the examiners to see the data differently. You have two fingerprint examiners not agreeing on the minutiae of the fingerprint. They're not even comparing fingerprints but actually are comparing what the data is if impacted by this bias and cognitive factors.

An example of the hierarchy of expert performance will be in DNA misinterpretation. We hear a lot about firearms, bite marks, and fingerprints, but even with DNA interpretation, there is subjectivity and bias if you're exposed to irrelevant information.

I came out with a study back in 2011, and NIST now has replicated—my coauthor from NIST has replicated it. We're talking here the same laboratory, same statistics, same proficiency, same training. And yet because of cognitive bias, they are going to get different results. And science is not about that, and justice is not about that. They're not going to have the results of my scientific experiment depending on which scientists have done the study. And I'm not going to go to jail or not depending on the luck of the draw, which forensic examiner examines the data. That's not science. That's not justice.

And there are new papers. I read a paper a few weeks ago on "Strengthening Forensic DNA Decision Making Through a Better Understanding of the Influence of Cognitive Bias."¹⁰ And what's nice about it is the coauthor is Bruce Budowle, who was the head of the FBI DNA lab and wrote most of the protocols in DNA, and Amy Jeanguenat. She was the director of Bode Cellmark DNA Laboratory, one of the biggest DNA laboratories in the United States.

So the community, the leadership is moving forward, is buying into it, and this is just one example. The problem is, and excuse the language, that without a kick in the ass, excuse me, nothing is going to change. We've been going in circles and circles. We had the NAS report back in 2009. NIST has excellent expert working groups in different forensic domains, making great recommendations, moving forward.

But the courts are not moving on forensic evidence. And I work a lot with experts across the United States and other countries. And you hear the experts, when you push them a lot, they say the court accepts it anyway so it doesn't matter. And if the court accepts what they are doing, it's not going to change because the courts are the real drivers in making the change.

10. See generally Amy M. Jeanguenat et al., *Strengthening Forensic DNA Decision Making Through a Better Understanding of the Influence of Cognitive Bias*, 57 *SCI. & JUST.* 415 (2017).

So back to Rule 702. How do we make things better in Rule 702? I don't know if you need to change it or not. The wise women and men of the Committee will decide. Maybe the Rule the way it is, is fine. It's misapplied, so we need to make sure it's applied correctly. Maybe you need a new rule. I don't know. But you can think about changes that require expert evidence—they must be based only on the relevant information.

Another possible change is to require that experts take active steps to minimize cognitive bias. These are some ideas to try to make the court enforce such a requirement, and perhaps the court will not need to enforce because the minute the forensic examiners hear that their current methods are going to be scrutinized by the courts, changes will be made. The rules committees and the courts are the drivers to make the change because we've been going in circles and circles for many, many years, and without your input, nothing will change.

Thank you very much.

PROFESSOR CAPRA: Thank you, Dr. Dror. [Applause] Our third speaker is Dr. Karen Kafadar. [Pause] But first I would like to ask a question. Dr. Lander, in requiring objective procedures, are you throwing out the baby with the bathwater, as Dr. Dror had to say?

DR. LANDER: No, no, no. Dr. Dror and I completely agree. He's saying that even if a method is not objective, but it could be done by a machine and all that, and it's being done subjectively, it can be admissible. But you need empirical evidence. So fingerprints are subjective. That's okay. I'd love to have an artificial intelligence just do it and tell you, and maybe we'll have that someday. We have an examiner do it. No problem, as long as there's empirical evidence that they're accurate and we can measure their accuracy.

Firearms, subjective. The PCAST report was entirely about the subjective areas and about what's necessary to believe a subjective method as opposed to an objective method, which is kind of easy. If it's objective, the examiner doesn't matter.

PROFESSOR CAPRA: And how does "biasability" factor into it?

DR. LANDER: So biasability will factor in with respect to the question of "has the method been demonstrated under conditions relevant to its use?" So, if, in fact, it's been demonstrated under a set of conditions that are relevant to the use, these resemble the real samples, the kinds of background information, for example, might be there, you would argue that's relevant. If it's been tested in a situation that's not really relevant, and eventually a judge will make a gatekeeping decision about whether or not it was tested under relevant conditions. We have many shooting-fish-in-a-barrel things that were not really relevant to the condition of use. And so, I endorse everything Dr. Dror says.

PROFESSOR CAPRA: Okay. Thank you. Dr. Karen Kafadar.

DR. KAFADAR: So thank you very much for inviting me. And I hope what I say today will be useful. My key message is, measurements are not perfect. There's uncertainty. There's errors in interpretation. Was it bronze, was it tan, was it brown? There are reporting errors. Peer review does not guarantee that the method is right or has been proved.

Scientific testimony should acknowledge and quantify the effects of various sources of uncertainty in the data. Different conclusions may or may not lead to different approaches to data. To the extent that different approaches to data all lead to the same conclusion, that's good, but one ought to look. And this is probably the most important. According to Merriam-Webster, statistics is a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of data.¹¹

Statisticians should not, indeed are not allowed to, testify as experts in chemistry, biology, physics, geology, et cetera. Rule 702 says a witness is qualified as an expert by knowledge, skill, experience, training, or education.¹² So what are nonstatisticians doing testifying about the collection, analysis, and interpretation of data?

I'm going to have an actual example where that's what went wrong—that they were letting into the testimony chemists who were interpreting data that, unfortunately, was misinterpreted. And then, as presently written, in my opinion, Rule 702 does not achieve this goal of preventing experts from testifying about things that they didn't have the ability to do, especially when it comes to serious statistical analysis.

I'm going to let Jeff Salyards talk about sensitivity and specificity. He's a scientist. I'm looking to see how well—if you're trying to decide if two specimens come from the same source—how well do the features help to identify that they are from the same source versus distinctive, indicating that they're not the same.

But I will say that in order to do that, you need very well-designed studies where the truth is known, and the FBI black-box studies that Eric Lander referred to is an example of that.

I'll also just mention that a lot of times in analyzing data, we don't have enough data to really confirm. It's simply too costly to collect enough data under a number of different circumstances, different levels of evidence quality, different examiner skill level, instrument types, et cetera.

And if it's too costly, more often than not, we tend to rely on convenient models. And in DNA, for calculating random match probabilities, they rely on a multinomial model, often in trace evidence, assuming that the measurements or some function of the measurements comes from a nice, pretty, normal distribution.

So you've all seen this bell-shaped distribution before. [Referring to a PowerPoint presentation] And the black curve is actually the bell-shaped distribution. There are two alternative models on that plot that are also often very plausible in that the data cannot distinguish between them. And yet the error rates that you would conclude from those three models could be very, very different.

11. See *Statistics*, MERRIAM-WEBSTER'S DICTIONARY, <https://www.merriam-webster.com/dictionary/statistics> [<https://perma.cc/EQG5-F5BC>] (last visited Feb. 14, 2018).

12. FED. R. EVID. 702.

So the example that I'm going to use to explain this is from compositional analysis of bullet lead.¹³ So this was a committee on which I served. This was actually five years before the 2009 report. The FBI had asked us to look at their procedures, which were basically statistical procedures for comparing whether the concentrations of seven elements in the lead found at the crime scene matched those from bullets found from the potential suspect's possession.

So they had a method for doing this, and they said that the conclusion was that the method was extremely good at being able to detect when two bullets came from the same source, and also they said they had the false-match rate. So how did they do this? They collected 1837 bullets, and they wanted these to be as different as possible, right? They wanted to represent the space of all possible bullets, okay? Then they used their match rule on all possible pairs of bullets, and they said, look, we got only 693 false matches. Well, they were different. They were intended to be different. I'm surprised they got any.

So, we said, well, let's use your data and actually try to model what they would look like, what would be the error rate under different conditions. What if the bullets were this far apart, this far apart, this far apart? And needless to say, the error rates were considerably higher.

Are we going down the same road right now with glass evidence? Right now, there are three standards involving the measurement of glass—three different ways of doing it. Two of those actually have some data on them, and we've looked at them. Two of them on LA-ICP-MS were getting quite different answers, depending on whether you look at one dataset or the other. There's no data on a third method. That, however, is the one that has been approved already to be posted on the NIST registry, but it's the same idea about their statistical method, so-called statistical method, that they're using. And, again, they're using it on data where they've tried to collect a big representative sample of different kinds of glass. I'm surprised they get any matches at all.

So there's a lot of day-to-day variability. And the actual false-positive rate, we suspect, could be a lot higher. The elements are more correlated. They're bouncing all over the place. And I mentioned that two data sets were giving two different sets of answers.

The real question is, why are nonstatisticians being allowed to testify on this stuff? Allowing chemists to testify on the collection, analysis, interpretation, and presentation of data makes about as much sense as asking me to talk about ICP-MS, okay? It certainly makes no sense.

So I just will mention with latent fingerprints, for years they were going around saying there's a 0 percent error rate. I mean, no statistician would say there's a 0 percent error rate. We could have saved them a lot of time.

13. See generally R.D. Koons & J. Buscaglia, *Forensic Significance of Bullet Lead Compositions*, 50 J. FORENSIC SCI. 341 (2005).

I will agree that the Ulery study on latent fingerprinting was quite well done.¹⁴ I will mention that further studies need to be done. Those were 169 latent print examiners who agreed to participate. It is just one study. Sir Ronald Fisher is known to have said that he is more likely to trust a result that was demonstrated at the P equals .05 level in ten studies than a single study where it looked like it was even more significant. So, in other words, replication of the findings in the Ulery study is necessary.

Statistical issues should involve different scenarios, levels of evidence, quality, et cetera, and also blinding. You know, they now do medical studies where, of course, they're always doing double blind. Latent print examiners have been saying that's impossible, you can't do double blind. Actually, Dr. Peter Stout, who's the director of the Houston Forensic Science Center, believes it is possible.

Courtroom testimony right now does not comport with statistical reliability. Here is an example of courtroom testimony from Brendan Max, who's the forensic science division chief at the Chicago public defender's office. And he questioned somebody and said, "You can't identify somebody on the basis of fingerprints through the exclusion of all others, right?" And the witness said, "oh, yes. Oh, yes, I can." And Mr. Max is frustrated because he says, "Try as we might, we can't get their testimony excluded pursuant to Rule 702. We can't even get judges to conduct pretrial hearings to assess the qualifications."

So, final remarks. Data analysis and interpretation do require some statistical expertise, more than a little. One or two courses in statistics no more makes a statistical expert than my two or three courses in chemistry would make me a chemist. I don't believe that nonstatisticians should be allowed to be providing expert testimony when it comes to the analysis of data. Relying on models is almost always necessary, but it can be tricky, and so you have to look at several of them. Opinions are not database conclusions, and experts, in my opinion, should be disclosing all the facts and data related to how they came up with their opinions. And that's all.

PROFESSOR CAPRA: Thank you, Doctor. [Applause] Our next speaker is Dr. Jeff Salyards. But first, a question for Dr. Kafadar.

You have the opinion that any kind of statistical analysis done by a forensic examiner without grounding in statistical expertise should be excluded? Is that the innovation that the Rules Committee should be thinking about?

DR. KAFADAR: Yes, and I don't know how to implement a change to do it because Rule 702 already says that an expert should not be allowed to testify beyond their expertise. And so I'm thinking that maybe it's just that courts don't realize that the statistical aspect is an area of expertise that goes beyond what chemists or biologists or ballistics experts can helpfully testify about.

PROFESSOR CAPRA: Right.

DR. LANDER: Dan, if I can add a point to that?

14. See generally Bradford T. Ulery et al., *Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners*, PLOS ONE, Mar. 2012, at 1.

PROFESSOR CAPRA: Yes.

DR. LANDER: I think that's a wonderful third-order consideration. The issue right now is in most cases, there's not even agreement that we have to discuss the accuracy at all.

DR. KAFADAR: That's right. That's a good point.

DR. LANDER: That the idea is we can present evidence without speaking about its accuracy. It's great that Karen is on this wonderful point that the people who speak about accuracy should be qualified.

I would just love it to be a requirement that if we're saying two things match and that's remarkable in some way, we must speak something about a level of accuracy.

PROFESSOR CAPRA: Right. Dr. Salyards.

DR. SALYARDS: My name is Jeff Salyards, and as everyone said, there's not a lot of time for formalities. But because I do work for the Department of Defense, it's important that I tell you, although I hope I uphold their finest standards, I'm not here representing them today. This is not the opinion of the Army or the Department of Defense.

We're going to talk about a few words: uncertainty, error, mistake, and the difference between validation with a little "v" and a big "V." It's part of a host of words that I think are really important. I promise this isn't just semantics. How we make each one of these things better is a different recipe. And we're bad, as a community, at talking about these words. But you know what? We're bad as physical scientists talking about these words.

We sometimes say error when we mean uncertainty. We sometimes say mistake when we mean error. So it's important that we separate those out. Why? Well, I think Eric hit it on the head. The substitutes for what really needs to happen if we abandon empirical proofs of accuracy is that you could let a Ouija board into your courtroom, and you don't want to do that, right?

And that's the dinner menu for the Titanic. [Referring to a PowerPoint presentation] The bottom there is the recipe for Waldorf pudding. If you follow this on the internet, there's a big debate in the culinary community of what was that recipe. I'm sure that evening there was some chef who was really working hard on that pudding.

As forensic-science thought leaders, we need to stop that quest. We're not trying to make Waldorf pudding on the Titanic. We need to patch the ship up. There are much bigger things afoot. So I'll give you a few examples.

So uncertainty is the first one, and this is important. Uncertainty occurs when we measure something. That's Heisenberg right there, and he came up with this concept that you can't know both the location and the momentum of a particle. It's a quantum-mechanical consequence. How that should apply to us is if I just asked you how long is that red line there, and you start staring at that ruler, Dr. Kafadar and I could probably talk to you for an hour about this really powerful exercise. How long is it? You know, is it 3.63 centimeters, or is it 3.62 centimeters, or is it 3.635794287? And you start realizing, if I start tacking on decimal places, that's nonsensical.

And so uncertainty is all about that. It's not about how sure I am of my final result. It's about a measurement. If we're measuring something, and there are some equations that govern that, it's got to be one of these physical quantities. There are not very many things that we can measure, right?

So I love this. We didn't plan this, but not to throw the baby out with the bathwater, I've picked a very appropriate example. Once we measure something, we're going to make a conclusion. We're going to reach a conclusion. And in this case, the conclusion that we're going to work on is, is this woman pregnant? [Referring to a PowerPoint presentation]

Now I think there's many mothers-in-laws out there that would claim they're expert at this. They're measuring something, Dr. Dror, that's very subjective, the glow, the glow of their daughter-in-law, of, "does she look pregnant?" And we've all done this, right? You can look and say, she's pregnant, right? And then you look at that young lady there and go, she's not, right? And then you look at her and go, she's pregnant, right? And then look at her, she's not, right? And then you look at her and go, hmm. [Referring to a PowerPoint presentation]

I'm sure we've all had the awkward social situation of encountering a friend, and you think for a moment, is she pregnant, or has she just gained a few pounds? Do I say congratulations, right? So that's not about how confident I am that she's pregnant. It's about how comfortable I am telling her she's pregnant.

So how does this begin to play, the words we're going to use here? We really want, as forensic scientists, to get towards confidence, and there's two ways we would do that. The first is on that X-axis, we'd establish what the heck it is we're measuring. Is it just this subjective "glowness?" Maybe it is. And there would be two populations—this red population of unpregnant women who have less glow than the blue population of pregnant women. [Referring to a PowerPoint presentation]

But as scientists, I would think I would make a more reliable assessment by checking the beta-human chorionic gonadotropin in their urine. There's something I could objectively measure. And then it would make—it would move us down and it would make those bell curves separate more.

But once I have that, I'm going to declare some cutoff and say if they have more glow than that, I'm going to call them pregnant. If they have less glow than that, I'm going to call them not pregnant. Wherever I move that line, I get some errors, and they come in two flavors: type one errors, false positives; and type two errors, false negatives.

And that curve on the right is called a ROC curve, a receiver operator characteristic. [Referring to a PowerPoint presentation] The DoD actually used it in looking at radar of enemy and friendly aircraft. The medical community uses this a lot, and they'll often just use it as a cartoon. The more that that line deflects up into the left, the better the method is. So they can just look at it and go, "that's a good ROC curve."

So we want something like that, but what's important is if I have the method, it's going to have an error rate. That's inherent to the method. It

doesn't mean anybody did anything wrong. It's just inherent to the method. It's important as forensic scientists that we talk about it that way.

Now, to Dr. Dror's point, what if we don't have something we can measure? What if it is subjective? Well, the FBI did this wonderfully with the black-box study. We can measure how good the evaluators are. We just show you a lot of pregnant people and let you pick where we have ground truth. And if you're actually good at it, I don't care that it's subjective, but there has to be some empirical demonstration that you're good at it.

So an error is not a mistake. In fact, the only way to measure an error rate is to assume that you're in a mistake-free environment, right? Now this is important because we really rely on this in forensic science. There are a lot of things we can do to minimize our mistake rate. We can surveil the activities in the laboratory. We can put in a robust quality system. The FBI is really good at this. We can do testimony monitoring. We can look at our notes.

But that does not lower the error rate. The only way we lower the error rate is to have a better method. Laboratories are the perfect place to deal with mistakes. Universities and thought centers are the best place to deal with errors.

All right. Validation, capital "V," lowercase "v." ISO 17025 governs what we do in laboratories says that our methods have to be valid, and it also says in a paragraph later on that our equipment has to check out.¹⁵ And this is what gets us into some problems.

So, if we take a look at DNA, the gold standard, and we got a new 3500 instrument at the lab, we take it out of its box, we plug it in, we turn it on. We make sure the thing works, it wasn't damaged. And then we begin taking some ground-truth samples, and we start making electropherograms, and the manufacturer says they better look like this.

And then we try it, maybe we make the room hotter or colder, or we, you know, play with the instrument. We turn a few knobs, and we make sure we're making those. And we start calling that, yeah, I validated the instrument. Well, that's a validation. It's the one with the lowercase "v." We really just check the performance of the instrument because work that Dr. Lander did early on asks, "did we pick the right DNA stuff in the first place? What if that marker just says you have a human liver? What if that one just says you have bilaterally symmetric ears?" Then we just failed capital "V" validation. It's not valid. It works, but that can get you to a Ouija board in the courtroom.

John Butler out at NIST did a pretty profound study that found out that we're very uneven across the community about how we handle DNA mixtures. I'm not sure that we show any success about what may be a three-person mixture. I think we might be pleasantly surprised when we dig in that the data are there to tweeze out that information. But we have got to be

15. See generally *ISO/IEC 17025*, ISO, <https://www.iso.org/obp/ui/#iso:std:iso:guide:31:en> [<https://perma.cc/27J4-Y43U>] (last visited Feb. 14, 2018).

careful. Even the stuff that is valid doesn't mean we can extend it to an area where we haven't shown its expertise.

Okay. So I'm going to just give you one more example. Gunshot residue. We're good at looking for lead-bearing antimony particles. Also, we have a group of scientists that are looking for organic residues of how this might work, and we're looking at which instruments will help us find all this stuff.

I would argue that we're making Waldorf pudding because what we really need is just a very simple test that says, "hey, if we swab those people, and then we swab those people who maybe have never touched a firearm, and then we swab these people who were near a firearm, but it wasn't fired, they didn't fire it, can we tell them apart?" And if we can't, the rest of that's Waldorf pudding. It doesn't help.

But we can sure come in and sound very convincing in your courtroom. We can show you molecules and data. And so that's what we've got to avoid.

There's an apple and an orange. [Referring to a PowerPoint presentation] If you take an infrared spectrum of that apple and orange, they're identical, right? So even though infrared spectroscopy is a great method, it's not a great method if you're trying to tell apples and oranges apart.

So thank you so much.

PROFESSOR CAPRA: Thanks very much. [Applause]

Most of what we've been talking about today and what started this whole event was feature-comparison methods. That's the PCAST report. But I also thought it would be useful to talk about other areas in which experts are being challenged and the process of validating the expertise. One important topic is the potential unreliability of eyewitness identification. On this topic, I have the honor of introducing Dr. Tom Albright.

But I have a question from Dr. Salyards while Dr. Albright is setting up: What about false negatives, do we care about false negatives at trial?

DR. SALYARDS: Yeah, I think we care about it, but we care about it way before it ever gets to final. A false negative is not going to get to trial.

PROFESSOR CAPRA: But do we care about a process if it has a lot of false negatives, but the false positives are low?

DR. SALYARDS: You know, that's certainly been the folklore of forensic science—as long as we don't put an innocent person in jail it's fine.

PROFESSOR CAPRA: Right.

DR. SALYARDS: As a scientist, I would say we are worried about false negatives. What we want is a method that is proved reliable, and there's no way to do that without lowering both error rates.

PROFESSOR CAPRA: Would you draw an inference from a large number of false negatives that there must be false positives you can't even find? I mean, something that's wrong on one side has to be wrong on the other?

DR. SALYARDS: It actually trades off. It has to do with how you slide that cutoff. And if you make it really hard to have a false positive, you make it much easier to have a false negative.

PROFESSOR CAPRA: Right. Dr. Albright.

DR. ALBRIGHT: Thank you, thank you. It's a great pleasure to be here to participate in this discussion. I'm going to talk about eyewitness identification. Eyewitness testimony has long played an important role in both criminal investigations and prosecutions. But sometimes eyewitnesses fail. They identify the wrong people. And probably the most familiar evidence for this, the one that's most highly cited, is this. [Referring to a PowerPoint presentation] This is from the National Registry of Exonerations,¹⁶ a page from the National Registry. These are DNA-based exonerations. And you can ask, well, "what were the reasons why these people were convicted to begin with?" And it turns out that in about 70 percent of the cases it was eyewitness misidentification that was a major factor leading to conviction.

This is totally unacceptable in a civilized society with a scientific reckoning. We have to ask, "what the heck is going on here? Why are people identifying the wrong suspects?" And to address this problem, the National Academy of Sciences convened a committee of experts.

I chaired this committee together with Judge Jed Rakoff. And the committee was composed of experts from fields of neuroscience, cognitive science, law, and law enforcement. And we had a resident statistician, Karen Kafadar.

The committee looked at three aspects of this problem. We looked at the science that's relevant to the problem of eyewitness identification—what science says about what things we can identify well and what things we can't identify well. We looked at the current and best practices used by law enforcement. And we looked at how this evidence is used in the courtroom.

The committee produced a report that was published in 2014.¹⁷ The report is freely available online. It's a rich source of information on this topic. I encourage you to take a look at it. The report has already begun to have a positive effect on this field, and one of the most remarkable manifestations of that is that one of the last things that Sally Yates did in her role as the Deputy Attorney General was to promulgate guidelines for the conduct of lineups and eyewitness identification, and these guidelines are almost verbatim taken from the National Academy's report.¹⁸

Now what I'm going to talk about today in my remaining minutes is the science. I'm a neuroscientist. And I'm going to talk about what science says about what eyewitnesses can do and what they can't. And I'll begin with a formal description of the problem. Some of you are probably familiar with this way of looking at things.

16. NAT'L REGISTRY EXONERATIONS, <https://www.law.umich.edu/special/exoneration/Pages/about.aspx> [<https://perma.cc/T328-8Z72>] (last visited Feb. 14, 2018).

17. NAT'L RESEARCH COUNCIL, IDENTIFYING THE CULPRIT: ASSESSING EYEWITNESS IDENTIFICATION (2014), <https://www.nap.edu/read/18891/chapter/1> [<https://perma.cc/8LG7-2B5S>].

18. Memorandum from Sally Q. Yates, Deputy Attorney Gen., Eyewitness Identification: Procedures for Conducting Photo Arrays (Jan. 6, 2017), <https://www.justice.gov/file/923201/download> [<https://perma.cc/C4H3-XT76>].

There are two states of the world. The person in a lineup is either guilty or they're not guilty. And the witness issues one of two types of responses based on looking at the witnesses in the lineup or the individuals in the lineup.

There are two of those categories of responses that are correct, that is, hits and correct rejections, and two are incorrect. And we can quantify the performance of an eyewitness, which is what's been done in many applied studies in the field of eyewitness identification over the years. We can quantify that using numbers, derived from a ratio of the correct responses to the incorrect responses, and there are a variety of different statistical approaches you can take to that.

And in these studies that have been conducted over the past roughly three decades, there have been many variables manipulated to see how they affect eyewitness performance. And I'll come back to some of those variables in a moment. But before I do that, I first want to make a general point, which is that there's nothing particularly unusual about the problem of eyewitness identification in a legal context other than the fact that the stakes are very high.

We encounter visual classification problems all the time. These are ubiquitous in our daily life. Find your car in the parking lot. You can identify—sometimes it's hard, but you can identify your luggage on the luggage carousel. You can find the book you were reading in the library. You can identify your dog amongst a big pack of dogs at the dog park.

On the bottom half of the slide are some more specialized forms of visual classification problems. [Referring to a PowerPoint presentation] The one on the lower left is the TSA problem, identifying bad things in the luggage. This one here is actually an infamous problem. This is the problem known as sexing a baby chick. And if you think about the financial constraints on chicken producers, you realize that there's a very strong financial incentive to know the gender of the baby chick at a very early age. And this is a really hard problem. People get good at it with practice, but it's a visual classification problem.

The one down here, of course, is finding bad stuff in body tissue, tumors and whatnot. And the one on the lower right here is the forensic fingerprint feature-comparison problem. All of these are visual classification problems, and they all have in common the fact that a person, a human being, makes a decision that's informed by visual perception and memory.

And so we can then ask, well, what do we know about visual perception and memory. And I'll first make a general claim, which is that eyewitness testimony reflects a decision made by a human observer based on things seen and remembered. A causal and predictive understanding of why eyewitnesses make errors depends upon an appreciation of how human vision and memory work.

Well, lucky for us, over the past few decades, we've learned an enormous amount about how human vision and memory work. I could talk about this all day. This is what I do for a living. But I'm going to narrow it down to three things that are fundamentally important. We've heard a little bit about these factors already this morning.

The three things are uncertainty, bias, and confidence. Uncertainty refers here to the fact that vision in general is far from perfect, and the reason is visual noise. Vision is plagued by noise from many natural sources, and these include such things as dim illumination, poor optics of the eye, and distracting features in the visual scene. And in the presence of this noise, we're often faced with a great deal of uncertainty about what we're actually looking at such that any decision we might make has a high likelihood of being wrong.

Now, if uncertainty can be likened to a breakdown of accurate sensory communication, then bias is the patch. Bias is what holds everything together in the face of noise. We fill in the blanks with what we believe is likely to be out there based on our prior experiences with the world. Confronted with a noisy visual world, as we typically are, these memory-based expectations or biases grant us the perceptual certainty that we need to survive.

But here is the catch. The same system that grants perceptual certainty under conditions of noise is also capable of filling in the blanks with the wrong information. Or, to put it differently, misinformed biases cause us to perceive things that don't exist. And to make matters worse, oftentimes we're absolutely confident about it. And, of course, this is what magicians aim for. They create conditions of uncertainty, they introduce bias, and they leave us with a strong sense of confidence about something that didn't actually happen.

But if the goal is truth, which, of course, it is in the criminal justice system, then uncertainty, bias, and confidence are the hidden enemies. Now we can get a little bit more quantitative about this. We can identify the variables that lead to uncertainty, bias, and confidence in an eyewitness context. Some examples are lighting, viewing distance, the duration of the events, crowding, whether there are other things crowding in the scene, distracting features, and the presence of a weapon.

There are administrative factors that will introduce bias affecting the decision made by the eyewitness. These relate to the type of lineup conducted. There's a longstanding discussion about the merits of what's called a simultaneous versus a sequential lineup, and the differences there actually tend to bias the witness in one direction or the other. The way fillers are selected and their appearance relative to that of the suspect is an important thing that will bias the outcome.

An important factor is whether the person who is administering the lineup knows the status of the people in the lineup. Instructions given to the witness, and the race of the witness relative to the lineup participants—these are all factors that bias the outcome. And then we have the problem of confidence.

Confidence is probably the most pernicious reason why eyewitnesses fail because a witness that comes into court and states something with a high degree of confidence is generally very persuasive. And confidence will increase with the passage of time. Often the reason why that happens is they gain additional information. They talk to their friends. They talk to the press. They talk to their counsel. And in doing so, they gather additional information that's a sort of confirmation-bias strategy. And so we want to

restrict access to information from reinforcing sources, and we want to know the time of the initial identification.

So these are all factors that can influence the outcome by influencing uncertainty, bias, and confidence. And we can, in principle, get even more quantitative by using a general linearized modeling approach. We're not there yet, but this will happen at some point. We can quantify uncertainty as some function of the variables I showed you a moment ago, bias as some function of the variables I showed you a moment ago, and confidence also as some function.

And the probative value, a scalar measure of probative value, is proportional, inversely proportional, to some function of those three measures. And so we'll be able to quantify that at some point. And then you might ask, "well, what does all this have to do with Rule 702?" This is part of Rule 702—the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence to determine a fact at issue.¹⁹

The problem with eyewitness identification is it's an unusual form of evidence. Our intuitions about what eyewitnesses say are generally wrong. And part of this has to do with the fact that we generally have a very high degree of confidence in our own views of the world. I saw something. I know I saw it. I know I'm right.

If I ask you to recall what you were doing and what you saw on September 11, you could probably all tell me something in great detail with a high degree of confidence, but odds are you're wrong. There have been many studies which have actually looked at this kind of thing.

So understanding the ways in which eyewitness evidence can go wrong is absolutely essential. And the committee that I worked on, the National Academy's eyewitness committee, made this recommendation:

The committee finds that a scientific framework describing what factors may influence a witness's visual experience of an event and the resolution and fidelity of that experience, as well as factors that underlie and influence subsequent encoding, storage, and recall of memories of an event, can inform the fact-finder in a criminal case.²⁰

Many scientifically established aspects of eyewitness memory are counterintuitive and may defy expectations. Jurors will likely need assistance in understanding the factors that may affect the accuracy of the identification. An important distinction here between the types of expert testimony that we've been talking about is the expert in this case is making evaluations of human performance. The expert is not making an evaluation of a particular method or a technique for doing forensic analysis. They're making decisions.

I would qualify this by also saying that conclusions about the factors leading to unreliable identifications are not really opinions. They're not any

19. FED. R. EVID. 702.

20. NAT'L RESEARCH COUNCIL, *supra* note 17, at 111.

more opinions than gravity is an opinion. These are well-documented facts about human decisions that are informed by vision and memory. And this is critical for going forward—for making sure that the triers of fact have adequate information to make a decision.

And I'll stop there. Thank you.

PROFESSOR CAPRA: Thanks, Tom. [Applause] So I guess, Tom, that you don't agree with the case law which says you don't need expert testimony on the unreliability of eyewitness identification because the jury can understand that point.

DR. ALBRIGHT: That's right.

PROFESSOR CAPRA: As many of you know, there are those opinions out there holding that expert testimony on the unreliability of identifications is not helpful, and this report might hopefully lead courts to reassess those opinions. Another thing that struck me when I was listening to Dr. Albright is that there is a circuit court opinion about handwriting identification, and they compared it to eyewitness identification, and the idea was, well, eyewitness identification is unreliable, and we allow that, so why wouldn't we allow handwriting identification?²¹ [Laughter]

Okay. So our next presenter is Dr. Alice Isenberg.

DR. ISENBERG: Thank you. I am the Deputy Director of the FBI laboratory, and I have a Ph.D. in chemistry. So my primary role at the FBI laboratory is to serve as the chief scientist there.

Today I want to talk to you about the culture of scientific excellence and continuous improvement that already exists in the forensic-science community, and I wanted to start by addressing accreditation.

Many groups have established that accreditation is important for forensic laboratories, including the National Commission on Forensic Science, the National Academy of Sciences, and the Department of Justice. In fact, the Department of Justice, through the Attorney General, directed all DOJ labs in November 2015 to obtain or maintain accreditation.²²

Accreditation is an external assessment of a laboratory's adherence to procedures which are fit for a purpose and based on sound science. Accreditation demonstrates a high level of confidence in a laboratory's work product for its customers and provides credibility and public confidence in the laboratory.

An accredited laboratory's quality-assurance system must include standard operating procedures, proficiency testing, a training program, technical review of reports, testimony monitoring, and many other requirements.

The Bureau of Justice Statistics has stated and published that 88 percent of the 409 publicly funded forensic crime labs are accredited as of 2014, and that's up from 70 percent of those labs in 2002. For those labs that are not

21. See generally *United States v. Herrera*, 704 F.3d 480 (7th Cir. 2013).

22. Memorandum from Loretta E. Lynch, Attorney Gen., Recommendations of the National Commission on Forensic Science (Nov. 23, 2015), <https://www.justice.gov/opa/file/797541/download> [<https://perma.cc/WGU4-GCAQ>].

accredited, they tend to be very small with fewer than ten people, and they also tend to offer fewer services or forensic disciplines.

Validation, which has also been called technical merit of methods, is also a requirement for accredited laboratories. And when I speak about validation, I'm not speaking about equipment checks, as Dr. Salyards mentioned. What I'm talking about as tests for validation are tests that are set up to determine whether a method gives the right answer or not when the right answer, otherwise known as the ground truth, is known. These are empirical tests, as Dr. Lander described.

Validation does not produce a discipline-wide error rate, but it does determine the conditions under which a method is reliable and will give the right answer. Validation tests the limitations of a method prior to its use in casework by analyzing things such as precision and accuracy, robustness of the method when used with degraded samples such as you would find at a crime scene, and it tests the limit of detection.

Validation also produces caveats as to when the method can be performed reliably, and these caveats should be shared in a court setting. An example of a caveat might be that an examiner cannot state or imply that a level of numerical certainty is calculated to support the identification of blood or semen through serology exams.

In addition to accreditation, forensic science research also demonstrates a culture of continuous improvement by developing new capabilities and improving upon existing capabilities. As mentioned previously, the FBI has done a significant amount of research on decision analysis or black-box studies.

We embarked on those studies as a result of the Mayfield fingerprint identification error,²³ and continued it with the support through the 2009 National Academy of Sciences report,²⁴ which thought that this was a good type of research for us to be doing.

We have continued work in this important area through doing black-box studies on firearms, just the firearms discipline, shoe-print discipline, and handwriting disciplines. Black-box studies are a good way to study one particular type of error rate, but they should not be mistaken as a way to determine an error rate for an entire discipline.

It is problematic when the validation of the methods used by scientists to test the reliability and limitations of a method is confused with the legal standard for scientific validity related to the admissibility of testimony.

The PCAST report²⁵ suggests one way to assess the validity of a forensic science discipline, but there are, in fact, many different ways to assess scientific validity. The entire body of research and testing that has been

23. See Press Release, FBI, Statement on Brandon Mayfield Case (May 24, 2004), <https://archives.fbi.gov/archives/news/pressrel/press-releases/statement-on-brandon-mayfield-case> [https://perma.cc/TGF2-58S5].

24. See generally NAT'L RESEARCH COUNCIL, *supra* note 1.

25. See generally PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

performed on a particular forensic method goes to support the legal scientific validity or lack thereof, not just one black-box test, as suggested by PCAST.

PCAST confuses four separate issues: first, whether the discipline itself is scientifically valid and is based upon sound scientific principles; second, whether there is a known error rate for a discipline as practiced across the universe of practitioners; third, whether individual practitioners can get to the right answer when the right answer is known, in other words, a personal error rate; and they confuse, fourth, whether two different practitioners will arrive at the same answer when reviewing the same materials and data—in other words, is the method reproducible.

Examiners also go through extensive training to be qualified to perform casework in accredited laboratories. That training can last up to one or two years. It requires that examiners perform analysis on samples for which the correct answer is known. They must demonstrate an understanding of the basic science behind the method. They must demonstrate an understanding of policies, lab procedures, legal procedures, evidence handling, and the like. They undergo mock trials. And they have competency tests with real-world samples where the right answer is known.

They then have continued competency monitoring with proficiency tests at least once per year for the remainder of their career. All of this training and testing occurs before an examiner contributes to real casework. Examiners must demonstrate that they can apply the processes accurately and reliably to get the right answer before they ever get to do casework because, in casework, we don't know the right answer.

Finally, testimony monitoring is also a requirement for accredited laboratories. The FBI lab requires that our examiners request a transcript after each of their testimonies for review. We also have approved standards for scientific testimony and reports which document the acceptable range of conclusions expressed in both laboratory reports and during testimony so as not to allow them to provide overstatements of their conclusions.

The Department of Justice is working on similar documents, called Uniform Language for Testimony and Reports, and they are also working on a testimony-monitoring program which will apply to all of the DOJ component laboratories. Each of these testimony-monitoring activities, which are required by accreditation, work to prevent testimony from going beyond that which the science can support.

In summary, accreditation, validation, research, training, and testimony-monitoring activities are all important for demonstrating the reliability of forensic science. The PCAST recommendations for demonstrating foundational validity of forensic disciplines are not the only ways to demonstrate that forensic conclusions are the product of reliable principles and methods.

In the end, we all have the same goal for forensic science—to get the right answer, communicate that answer to a layperson, and to continuously improve our ability to get higher quality and quantity information from evidence. Thank you. [Applause]

PROFESSOR CAPRA: I had a question, Dr. Isenberg. The black-box studies you are referring to—they're in process now?

DR. ISENBERG: Yes, that's correct.

PROFESSOR CAPRA: Is there a projected date? How long does it take?

DR. ISENBERG: They take years actually. To gather the evidence and make sure it is blinded to all the appropriate people, we have to contract it out to other agencies, academic agencies, and then to physically send that evidence out to examiners across the country takes a good bit of time. And then, of course, the data analysis is a pretty big lift too.

PROFESSOR CAPRA: And then your reference to examiners, what's the scope of that? Does it include, for example, bite-mark examiners or latent fingerprinting? Out of all the feature-comparison methods, are we talking about all of them that are basically used in a trial?

DR. ISENBERG: I wasn't referring to bite-mark examiners because none of the Department of Justice laboratories and—I'm not aware of any lab that has a bite-mark examiner on staff. But I am referring to all the different other kinds of examiners, like chemistry examiners, DNA examiners, latent fingerprints, firearms, and the like.

PROFESSOR CAPRA: Thanks.

MS. GOLDBACH: I have a question. You indicated that—

PROFESSOR CAPRA: This is a question from Anne Goldbach, who is also on the panel coming up.

MS. GOLDBACH: Accreditation requires monitoring of testimony. Is that under [ISO] 17025?

DR. ISENBERG: Yes, it is.

MS. GOLDBACH: And is there supposed to be documentation of that monitoring?

DR. ISENBERG: Yes, there is. Each laboratory develops a methodology through which they will monitor testimony of examiners, and they have a standard operating procedure for how that will work in their methodology.

MS. GOLDBACH: And the evaluation of their testimony, is there a record of that kept as well?

DR. ISENBERG: Yes.

MS. GOLDBACH: Thank you.

PROFESSOR CAPRA: Our next speaker is Susan Ballou, from the National Institute of Standards and Technology, or NIST.

MS. BALLOU: So, Dan, thank you for the invitation, and thank you for the organizers for everything that they've done to pull this together. Talking with everyone last night, it sounded like it was a real heroic effort to get this in place. And as Jeff Salyards had mentioned before, please allow me to read my disclaimer to you. The following are my points of view and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

Science is based on principles that have been repeatedly confirmed through observation, experiment, centuries of use, and the basis of forensic analysis in many disciplines. However, the public perception is slightly skewed. This

is what we're dealing with in the courtroom, analysis complete in ten minutes. Everything is a match. The public believes that we drive Humvees from the lab to the crime scene and that, at that time, we already found the suspect and we've interrogated him and brought all the evidence back to the lab for confirmation.

However, this is far from what's actually taking place. The public believes that when a scientific-method result is provided to them, that it's fact, when, in reality, change is the scientific method. Experimentation answers questions. Experimentation is needed. And as we know in the laboratory, usually more questions arise from that experimentation.

Therefore, keep in mind that science is constantly changing. New instrumentation is changing. Vendors are out there now looking for new ways of bringing technology into the crime labs. New processes are established.

Think of the 1960s. Was an avocado-green dress okay to wear? It was at the time. Was plastic the way to go in your homes? That was the thing to do. It was acceptable. It was agreed to. Nowadays, if you see some place with a shag carpet, you kind of think about who would have picked that in the first place. But it was accepted back when.

So what was the science back then? At that time, serology. We did not have DNA. What was the point? We tried to discern out of a blood stain left at a crime scene as many factors as possible to get more information out of that particular blood stain.

What I have for you right there is the most common blood markers, ABO-type protein markers of that time. [Referring to a PowerPoint presentation] Now, if I found this information on a crime-scene stain, the best that I would be able to tell you or testify to you is that 1 out of 100 could have contributed to that stain. Not such great results, but that is what we had. That's what was accepted in the courts at the time.

So it didn't mean it was wrong. Now that we're finding all of these cases where innocent people were incarcerated because of the science that was used back then, we're learning. Science is evolving.

As was mentioned earlier, on crime scene evidence, you do not know the ground truth. That is the part of the examiner, to start tweezing out what exists. To determine statistics, you need to know ground truth. And, unfortunately, they're finding out there's a lot of information that we don't have available to give this ground truth—to determine uncertainty or probabilistic information.

So the difficulty I see is if we get too prescriptive with law or legal rules, it might hinder the progress of science. At the National Institute of Standards and Technology, our mission is to promote innovation and industrial competitiveness. We're known for our measurement-science techniques, standards, and technology. We have some of the latest equipment that we evaluate, look at their capabilities, and give a result as to what the limitations or expectations of that technology are.

Some of the things that we've taken on at NIST—and here we have forensic science—is going into the venture of the human factors. This is an

area you've already heard about this morning. It's a concept that the forensic scientists had not even considered in the past. Science is science. I'm giving you a result in court as to what my analysis was. Am I biased? No. I just followed the science.

It wasn't until we decided to pull this working group together at NIST, which was difficult because people didn't want to talk about any possible error or misrepresentation in the field. We pulled together a group looking at latent prints and determined that there are many human factors that are involved in the interpretation of the evidence.²⁶ And I thank the FBI for taking a strong stand on this and making revisions in accordance with this document. And because of that action, numerous crime labs across the United States saw what was being done and have changed or reviewed or instigated new methods and procedures.

Another example to show human factors, just to show the different disciplines that we're looking at, is a corporation that produces comparison units for firearms, bullets, and casings. That company specifically lists in their protocol that when you're capturing an image set up on the stage, one should run the image through the computer and leave all the standards or the metrics of that instrumentation alone—do not change it. It's capturing an image, putting it into a database, and then you would capture another, say, striations on another bullet and ask for a comparison for something similar in the database.

Examiners thought it looked washed out. Some agencies will allow their examiners to change the lighting. Unfortunately, the way the system was set up, some of the darker modes would capture the image differently. And, therefore, you would have a sample in the database, but when you did a search to see if it would match, again, another run, it wouldn't match because someone had changed the lighting.

We found this out by doing research at NIST on some of the nuances that were taking place. And to understand that we could not do it all within the National Institute of Standards and Technology, the upper administration realized that forensic science is one of the most progressive and interesting areas right now. So we decided to make it as one of our centers of excellence.

The Center for Statistics and Applications in Forensic Science was established in May 2015. Karen Kafadar's university [the University of Virginia] is part of this. Iowa State has the lead.

So we have approximately thirty projects, and I'm just going to read some titles to let you know the areas that we're looking at. "Statistical models for the generation interpretation of shoe print evidence." It was mentioned before—there's something lacking in shoe prints. We're addressing that. "Evaluating lay perceptions of forensic evidence and forensic statistics." "Improving the statistical validity of forensic science databases." How large

26. See generally NAT'L INST. OF STANDARDS & TECH., LATENT PRINT EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH (2012), http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=910745 [<https://perma.cc/3P4C-SK5L>].

do you need a database? What is the information that should be in it to make it usable? “Statistical and algorithmic approaches to shoe-print analysis.”

The answers found from these scientific ventures and the questions that will be raised are part of the scientific process. So, yes, there’s international documentary standards that are in place—a basis to build from—but these standards are written to allow the natural flow of science. Science involves change, new instrumentation, and new processes. Thank you.

PROFESSOR CAPRA: Thank you, Susan. [Applause] So because everybody adhered to their time limits—thank you so much—we have time for a general discussion, if anybody’s got questions or comments, either from the panel, from the Advisory Committee, from whoever is here. We’ll take it. And I guess we’ll start with Dr. Lau.

DR. LAU: Thanks. I’m from the Federal Judicial Center, but I speak only for myself. And I have a question for Dr. Kafadar about your claim that chemists cannot talk on the collection, acquisition, and analysis of chemical data.

So, you have to have a limiting principle, Dr. Kafadar. For example, are you going to tell Dr. Isenberg, a chemist, that she is not competent to testify about her using a sample because that is related to that analysis of data. So, what is the limiting principle to your statement that only statisticians should be trusted to comment on the acquisition, collection, and analysis of data?

DR. KAFADAR: So it might have been a little bit strong, and I certainly wouldn’t object if Dr. Lander testified about statistics and mathematics because he does have a degree in mathematics. But I think—

DR. LANDER: But not statistics.

DR. KAFADAR: Yeah, but not statistics. But I think the point that surprised me, you know, at the time, I kept thinking why is it that they were asking chemists to develop statistics procedures for a match. They could ask me to develop procedures for measuring the concentrations. I’ve had three chemistry classes. But it certainly doesn’t make a lot of sense. I mean, you’d ask a chemist to do that.

And likewise, it was such a simple problem that if they’d simply asked a statistician what would be the appropriate procedure, that would have been fine. Then they could have gone into the courtroom and said this is the procedure that’s been developed by statisticians.

So the statistician does not necessarily have to be in the courtroom, but one would hope that they would have consulted statisticians before they developed the procedure.

PROFESSOR CAPRA: And Dr. Lander.

DR. LANDER: Great. I’d love to engage in a colloquy with Dr. Isenberg because I have enormous respect both for her and the FBI. When PCAST wrote about foundational validity, that was one of many things that are good to have.²⁷ There was also a discussion of validity as applied, that the expert actually applied the things reliably.

27. See PRESIDENT’S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 47–56.

So PCAST wrote about foundational validity,²⁸ the concept that's meant to correspond to Rule 702(c), that the testimony is based on reliable principles and methods.²⁹ And then there's Rule 702(d), that they have been reliably applied by the expert.³⁰

So, when Dr. Isenberg says there are alternative ways other than doing a black-box study, let me be clear. A black-box study means a test, any test properly done where you have samples, where you know the ground truth, and you see how often they get it right. There's no alternative to that. That's called foundational validity because it must be a foundation.

If you haven't even tested the method to find out whether folks get it right when they're presented things where we know the right answer and they don't, you don't have a method. You can't do anything.

Dr. Isenberg refers to many other good things to do, and I support all of those good things she's talking about. We need to know about the variation between examiners, variations between conditions. Those are not alternatives to the foundation of actually having measured the accuracy of the test.

They are things that are great to have. They are follow-ons. How might it vary under different conditions amongst different examiners? But there is no way that any method should even be discussed until there is a foundational demonstration of its accuracy.

The other point that I'd make has to do with the thing I very much support, that there should be rules about testimony. The Department of Justice has issued draft guidelines for this uniform language on testimony and reports. I'm deeply disturbed by the fact that those draft guidelines forbid, in many cases, examiners from making any reference to quantitative accuracy.

I completely support the idea that there should not be overstatement. The DOJ wrote them to avoid overstatement. They wrote them so well to avoid overstatement that they forbid any statement about quantitative accuracy. Without quantitative accuracy, again, you have nothing. I totally respect what the FBI has done. They are doing these studies. Call them black box if you will. We do because it's the term the FBI adopted.

They're great things to have, and they will begin to put these fields on solid foundations where we know what their accuracy is. But I don't think there's any alternative to simply doing a test and measuring something. It isn't the end of the story, but it is the foundation of any story.

PROFESSOR CAPRA: Dr. Isenberg.

DR. ISENBERG: With regard to the first part of your question, Dr. Lander, about black-box studies not being the only way to assess foundational scientific validity, in my opinion, the PCAST report³¹ discarded a lot of research that did have empirical studies that analyzed the disciplines in a variety of different ways in a systematic fashion. And the FBI repeatedly

28. *See id.*

29. FED. R. EVID. 702.

30. *Id.*

31. *See generally* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

brought those studies to the attention of PCAST but they were not used to support the scientific validity of certain disciplines because they did not meet the definition of a black-box study defined by PCAST.

DR. LANDER: Could you just elaborate? Because I'm very interested.

DR. ISENBERG: Sure.

DR. LANDER: It's an important colloquy here.

DR. ISENBERG: I don't think we should get into a detailed analysis of the papers.

DR. LANDER: Not the specifics, but what kind of studies are you talking about?

DR. ISENBERG: In particular, the firearms discipline. There are a number of studies which did provide empirical data, and for a variety of reasons, because the particular parameters of those studies did not meet PCAST's definition for an appropriate black-box study, they were discarded. But yet, in my opinion, that's throwing out data that does show that you can get the right answer when the right answer is known to test the person who has developed the experiment.

DR. LANDER: If I can respond to that. Those were the studies that produced an error rate of 1 in 5 as opposed to the later rate of 1 in 500. We said those were inappropriate because these were studies in which there is a bag of bullets to be compared and the people doing the comparison knew that the right answer was present. That's a huge piece of information. If you know the right answer must be found there, that doesn't resemble a blind black-box study. So, we discarded a bunch of studies.

PROFESSOR CAPRA: That's the bias that Dr. Dror talks about.

DR. LANDER: That's a huge bias. You know that the right answer is present. The same was true in many of the early hair studies. We discarded those because if the examiner knows, for example, that all samples come from different sources, or the right answer must be present, it's a problem. And it's not just a theoretical problem. Look at the results. We said when they had that, they only made a mistake 1 time in 5. When they actually didn't know if the right answer is present, they made a mistake at the rate of about 1 in 500.

We discarded it not because we don't like black-box studies or we don't consider that empirical evidence. We consider that they're a shooting-fish-in-a-barrel kind of test. And so, we rejected shooting fish in a barrel. It has to be appropriate to the intended use. It's valuable empirical evidence. It's good to know. But it isn't what you do in a case. So, we said no. Do a test. DoD did a test. You're doing a test.

I don't think we truly disagree that we just want empirical evidence relevant to it. And I have huge respect, and I'm thrilled you guys are just doing it. You will set the standard. And we shouldn't be arguing about the need for a test or quibbling over whether studies done earlier truly got it right or not.

I realize the Department of Justice is concerned because PCAST feels there wasn't enough evidence. If all we're doing is arguing about whether firearms

is over the line or not, that's a minor point. You and I both agree there is a line. Without empirical evidence of the sort you guys produce, it has no business in court.

DR. ISENBERG: I just want to say that even though those studies might not have been designed or might not have analyzed their results in a way that PCAST defined as the proper way, it's still information that supports that it's a valid science because—and I don't agree with shooting fish in a barrel as an analogy. Those people still had to look at samples without knowing the true answer. And even if it was a fill-in-the-blank test or a multiple-choice test, you can still get the answer wrong. But they got it right. So, I don't agree that just because the study wasn't designed the way you defined it that it's data that doesn't support the validity of the science.

PROFESSOR CAPRA: So, for a court, the issue might be the weight of such a study as opposed to whether you consider it or not.

DR. ISENBERG: True, true. Yes.

DR. LANDER: That's right. These are good, fine points that I disagree on.

PROFESSOR CAPRA: And I'm sure that all judges will be able to figure out this discussion. Yes. So that is our challenge today, right, to figure how this applies into Rule 702.

DR. LANDER: Right. I mean, the judge will have to ask whether the evidence of validity, of reliability, relevant to the situation. And I think judges can figure that out.

PROFESSOR CAPRA: Well, we'll see because we're going to turn to judges now. [Laughter] And our first judge on the panel is Judge Kozinski from the Ninth Circuit, and we seek his advice and opinion.

JUDGE KOZINSKI: Well, I'm not really an expert. I stumbled into this field. I'm an appellate judge. We don't know much of anything, unlike trial judges. Well, we deal with facts, and I actually have on-the-ground experience dealing with experts. We sort of look down.

But I happened to stumble into it, when I wrote an article in the *Georgetown Law Journal*,³² just by accident. I was invited to do this article. I thought it was easier to do the article than say no. And when I started writing it, I realized I had over the years developed serious doubts about various aspects of the criminal justice system, and one of them had to do with forensic evidence, which I express in the article. And I was invited on the PCAST panel to be one of the advisors. Again, I, unlike Jed Rakoff, who has done a lot of work in the field and knows quite a bit, I came to this as more of a layman.

But I found the PCAST report to be quite persuasive, and I was disappointed by the swiftness with which the U.S. Justice Department and the FBI rejected the report on what I thought were insubstantial grounds. So

32. See generally Hon. Alex Kozinski, *Criminal Law 2.0*, 44 GEO. L.J. ANN. REV. CRIM. PROC. iii (2015).

I guess I'm going to have to part company with Dr. Lander and his great tribute to the FBI lab and his statement of respect.

And what I want to talk about is one of the recommendations of the PCAST report, specifically the recommendation on the independence of forensic-science labs. And I'm going to question whether or not you really can have a truly scientific office that is tied at the ankle with the prosecution office. And, unfortunately, we have had any number of reports in recent years about forensic labs gone bad—that have done terrible things.

People send me articles now, so just this month I got an article about a drama unfolding right here in Massachusetts involving the Massachusetts crime lab. I want to be careful what I say about it because it's an ongoing case. But, as I understand it, there was an examiner by the name of Sonja Farak who was found to be consuming the very drugs that she was testing, and defense lawyers raised objections and asked for information as to whether perhaps their clients' cases had been affected. And the prosecutors in Massachusetts stonewalled, basically said, oh, it's an isolated problem, don't worry about it.

I will skip over some of the details. But the bottom line is that in an opinion written just earlier this month by Superior Court Judge Richard Carey, whom I don't know, he made findings that the Attorney General's office and essentially two of its lawyers committed serious misconduct by misrepresenting to the court what the AG's office knew about the misconduct.³³

In fact, Sonja Farak was not simply an occasional user in a few isolated cases. She started using in 2005. Then eventually she started fabricating evidence. And by 2012, she was cooking meth in the crime lab, in the Massachusetts state crime lab.

We're not in Arkansas, folks, and we're not in—this is Massachusetts, very close to New York, where Judge Rakoff reigns. And this is not the first time, of course, that the Massachusetts crime lab has been implicated. We all remember Annie Dookham of Boston, who was convicted of similar misconduct. Unfortunately, this is not isolated. This is not limited to Massachusetts. This is not limited to state crime labs. We have seen tremendous problems with the FBI lab, Dr. Lander's bouquets and compliments notwithstanding.

In 1997, there was an Inspector General's report³⁴ raising serious questions about the series of FBI examiners that collectively were involved in 7600 state and federal convictions, including fifty-three death-penalty cases. This was twenty years ago. The FBI was supposed to go back and reexamine those cases. Last we heard, only a tiny fraction have been reexamined.

33. *See generally* Commonwealth v. Cotto, No. 2007770, 2017 WL 4124972 (Mass. Super. Ct. June 26, 2017).

34. *See generally* U.S. DEP'T OF JUSTICE, THE FBI LABORATORY: AN INVESTIGATION INTO LABORATORY PRACTICES AND ALLEGED MISCONDUCT IN EXPLOSIVES-RELATED AND OTHER CASES (1997), <https://oig.justice.gov/special/9704a/> [<https://perma.cc/ZT49-728K>].

Now these are people who are in there for life, some of them convicted of the death penalty. There has been no real haste in trying to undo what could be a serious injustice.

So, again, I just want to raise the flag about whether it is necessary to have forensic labs that are an arm of the prosecution or that in any way connect with the prosecution. I'm a capitalist. I believe in the free market. And there's nothing about forensic testing that requires government employees and certainly nothing that requires government employees whose career path depends on how well they please prosecutors.

I mean, you do not succeed, you do not advance as a prosecutor by the number of times you get not-guilty verdicts, even though in prosecuting, you're probably going to get not-guilty verdicts. You know, you tried the case and justice was done. But let's face it. Prosecutors want to get convictions, and if they have crime labs that are part of their organization, the implicit and explicit incentives will be to reward evidence that supports the prosecution.

Now we've talked about—and I commend—I do commend the fact the FBI and other crime labs are now moving to more objective methods. But as we have heard, there continue to be subjective elements in almost all aspects of testing, questions of lighting, questions of alignment, questions of how many comparison points you use in fingerprint testing.

Those are all judgments. And so long as you don't have what Dr. Lander referred to as the purely mechanical method, anything that requires human judgment, it seems to me that if it is done by an organization that has an inherent bias, it ought to be mistrusted.

I don't want to go so far as to suggest a rule at this point by the Advisory Committee, but I do think that a study is worthy of the question of whether or not this kind of work could be done by independent labs and whether it is even possible to squeeze out the inherent bias of having labs that are captive.

PROFESSOR CAPRA: Thank you, Judge.

JUDGE KOZINSKI: Thank you. [Applause]

PROFESSOR CAPRA: Our next speaker is Judge Rakoff.

JUDGE RAKOFF: So I'm always very happy to speak after my close and admired friend Judge Kozinski, because it always makes me appear so much more moderate. [Laughter]

Some of what I wanted to say has already been covered, but I want to say a word about the way forensic science in general and match science in particular tends to be overstated in the courts and what might be done about that.

The testimony in cases of matching fingerprints, bite marks, firearms, and so forth comes down in essence to saying this evidence gathered from the crime scene matches this evidence associated with the defendant, and the chances that it was anyone else are remote, to say the least.

The problem is that we've increasingly learned that the error rate in many of these methodologies is considerable, and in other methodologies, we just simply don't know what the error rate is. The FBI, for example, determined

it was about 11 percent in the case of hair samples. Bite marks, most of the studies show, are even a considerably higher error rate.

And this is a function of the fact, as was pointed out, that many of these methodologies were developed for investigative purposes, not as evidence to be used in court. But they have morphed into evidence that's used in court, even though there is a substantial subjective element in most of these methodologies.

I agree with the point that was made earlier, that subjectivity per se should not be a reason for excluding expert testimony. But it does open the door to factors like bias that are not present when you have a mechanical kind of test. So, it can easily result in a higher error rate than would otherwise be the case.

But the problem I want to talk about is the way this is all presented. Sometimes, remarkably, the experts say that the error rate is zero. I had this in my own court. By the way, even in federal court, forensic testimony is often from non-FBI folks, either in civil cases or in criminal cases, where for one reason or another they're using local people. So, this was a New York City lab person testifying in a toolmark case, and I asked him what his error rate was, and he said zero. And I said I'm talking about the methodology. What's the error rate for this methodology? And he said zero. And I said, how could it be zero? He said, well, not only in my case but in all my friends' cases, every time we testify, the guy's convicted. [Laughter]

So that's the kind of testimony you sometimes get. The formulation that historically has been used most commonly and indeed in some states is required is that the opinion be "to a reasonable degree of scientific certainty." That's a terminology that no scientist would ever use, and it's very misleading.

Much of this is either not science or not firmly grounded in science. But, more importantly, virtually none of this is certain. And what is a reasonable degree of scientific certainty? No one knows what that means. Different states define it differently, but the jury is never told what the definition is within that state.

In the federal courts, while the reasonable degree of scientific certainty standard is not required, it is very frequently encountered. Particularly in civil cases, you hear it all the time. There are circuit decisions approving that formulation. And yet it is, in my view, a very misleading formulation. The National Commission on Forensic Science, by a very substantial, really overwhelming majority, put out a recommendation that that formulation not be used in a forensic science case.³⁵ And the Department of Justice essentially adopted that position,³⁶ and my understanding is they are now

35. See generally Nat'l Comm'n on Forensic Sci., *Recommendations to the Attorney General Regarding Use of the Term "Reasonable Scientific Certainty,"* U.S. DEP'T JUST. (Jan 16, 2016), <https://www.justice.gov/archives/ncfs/page/file/816021/download> [<https://perma.cc/M5UF-BZRP>].

36. See generally Memorandum from Attorney Gen., *Recommendations of the National Commission of Forensic Science* (Sept. 6, 2016), <https://www.justice.gov/archives/ncfs/page/file/841861/download> [<https://perma.cc/7YV2-TZHR>].

trying to implement that. It takes a while to, if you will, filter down to the troops.

But I had a case not very long ago, a civil case, where it was used again or attempted to be used—I cut it off—and I think that and similar formulations could be the subject of this Committee’s rulemaking function. There was a draft of an amendment passed around to us—and I’m not sure who was the draftsman of this—

PROFESSOR CAPRA: That was me.

JUDGE RAKOFF: Ah, so—

PROFESSOR CAPRA: Just for fun, though.

JUDGE RAKOFF: Yes. Well, that’s sad.

PROFESSOR CAPRA: That’s how I have fun.

JUDGE RAKOFF: You didn’t have to mention it was for fun. I knew it was from you. [Laughter]

But anyway, this proposed rule, which would have, for testimony and for forensic expert witnesses, a provision, subsection C—the witness must actually state, on the basis of adequate empirical evidence, the probative value of any similarity or match between the evidentiary sample and the source sample. I like that a lot. And I thought the commentary to that could well say, for example, that we disapprove of formulations like “to a reasonable degree of scientific certainty.”

But, I mean, this is all for the Committee. My point simply is that the effect of the formulations presently used in a great many cases are to greatly exaggerate in the jurors’ minds just how error free or not error free this testimony is.

PROFESSOR CAPRA: So you, in the *Glynn* case, you chose to allow the ballistics expert to conclude that it was more likely than not the same gun.³⁷

JUDGE RAKOFF: I did, and that was special to the case. I wouldn’t say that would be appropriate in every case.

PROFESSOR CAPRA: Like not a standard for every circumstance.

JUDGE RAKOFF: Yes.

PROFESSOR CAPRA: Okay. Thanks so much, Judge Rakoff. [Applause] And Judge Moore will be our next speaker. And the floor is yours.

JUDGE MOORE: As you might expect, the Southern District of Florida, once home to the Cocaine Cowboys, sees its fair share of narcotics cases. Although we continue to see a lot of cocaine and marijuana prosecutions, we also are seeing synthetic drug cases on our docket.

Synthetic drugs, also known as new psychoactive substances, are rapidly becoming the drugs of choice. Fentanyl, flakka, spice, MDMA, cannabinoids, and bath salts. The list goes on as new drugs are created and introduced to the market. According to the Drug Enforcement Administration, law enforcement has encountered over 300 types of synthetic drugs.

37. *United States v. Glynn*, 578 F. Supp. 2d 567 (S.D.N.Y. 2008) (Rakoff, J.).

The Center for Disease Control and Prevention reported that between 2014 and 2015, the synthetic opioid death rate increased by a staggering 72 percent. Most synthetic drugs are produced in labs in China, Southeast Asia, and Mexico, and new synthetic drugs are commonly the result of ever so slightly altering the chemical compounds of existing or naturally produced narcotics.

This is an ever-changing dynamic, and it seems just as one hole is plugged another appears. The Controlled Substance Act of 1970³⁸ provided that certain drugs and chemicals that are considered controlled substances are divided into schedules. There are five schedules, with Schedule I substances being the most dangerous class of drugs and Schedule V the least.³⁹

In 1986, Congress enacted the Controlled Substance Analogue Enforcement Act,⁴⁰ which identifies substances that are substantially similar to those listed on the federal controlled substances schedule. So a substance that is not listed as a Schedule I or II controlled substance receives the same treatment as a Schedule I or II controlled substance if it shares a substantially similar chemical nature, it has a stimulant, depressant, or hallucinogenic effect on the central nervous system, and that is substantially similar to or greater than the controlled substance, and with respect to a particular person who represents or intends that it has a stimulant, depressant, or hallucinogenic effect on the central nervous system that is substantially similar.

In 2011, the Attorney General, through the Food and Drug Administration, placed five synthetic cannabinoids and three synthetic stimulants on Schedule I. The Synthetic Drug Abuse Prevention Act of 2012⁴¹ added five structural categories of synthetic cannabinoids and eleven synthetic stimulants and hallucinogenics to Schedule I. It also expanded the time frame of temporary scheduling.

Essentially, the Attorney General has the discretion to, by order, add a substance to Schedule I on a temporary basis if it is necessary to avoid an imminent hazard to the public safety. The sentencing of individuals convicted of possession or distribution of synthetic drugs, particularly analogues that are not specifically listed on the schedule, is also working its way through the federal courts.

The U.S. Sentencing Commission is presently conducting a two-year study of synthetic drugs. In April 2017, the commission held a public hearing to receive testimony on the prevalence and effect of synthetic drugs, and since that hearing, a study of specific categories of synthetic drugs, including fentanyl, has begun.

The commission is focusing on the chemical structure, pharmacological effects, potential for addiction, legislative and scheduling history, as well as other related issues. There are numerous efforts to streamline and simplify

38. 21 U.S.C. § 801 (2012).

39. *Id.* § 812.

40. Pub. L. No. 99-570, 100 Stat. 3207 (codified at 21 U.S.C. § 813 (2012)).

41. 21 U.S.C. § 812(c) (2012).

the identification and criminalization of new designer drugs. The reality is that flexibility is critical in this area.

In terms of *Daubert*⁴² and Rule 702,⁴³ I believe that maintaining the status quo and keeping the rule as is, bestowing a gatekeeper with discretion regarding the admissibility of expert testimony, is particularly important in these cases as well.

Rule 702 provides that an expert opinion is admissible if it is based upon sufficient facts or data, the testimony is the product of reliable principles and methods, and the witness has applied the principles and methods reliably to the facts of the case.

In *Daubert*, the Supreme Court set forth factors for courts to consider in determining whether an expert's methodology is sufficiently reliable, whether it can and has been tested, whether it has been subjected to peer review and publication, what its known or potential rate of error is, and general acceptance in the field.⁴⁴

Because these synthetic drugs are being altered at a rapid rate, it comes as no surprise that a defense of a criminal defendant will likely involve challenging the admissibility of expert testimony that a particular substance was a barred synthetic drug. A chemist's methodology might be sound in other respects, but there could be questions about reliability with testing and opining on particular chemical structures that have not been tested repeatedly before. *Daubert* provides much-needed flexibility to gatekeepers in the emerging area of synthetic drugs.

In 2015, I presided over a six-day jury trial that resulted in the conviction of a defendant on one count of conspiring to possess with intent to distribute controlled substances and controlled substance analogues. The defendant appealed his conviction, arguing in part that the court clearly erred in finding that the most closely related substance referenced in the drug equivalency tables in the sentencing guidelines was THC, not marijuana.

The defendant argued that the government's expert opinions were not reliable because they were based on human testing. The expert whose testimony I admitted opined on the effects of the alleged analogues on the human central nervous system and represented that his opinions were based on the structure of the chemicals, in vitro testing, in vivo testing in rodents, and case reports.

The Eleventh Circuit affirmed and noted that the defendant's objections went to the weight of the testimony, not the admissibility. This was a fairly straightforward case, and because of the flexibility provided in *Daubert*, I considered the expert's methodology and reliability without being constrained due to the lack of a particular kind of study or the lack of peer-reviewed articles, for example.

42. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

43. FED. R. EVID. 702.

44. *Daubert*, 509 U.S. at 593–95.

However, it is not always this straightforward. In 2013, in *United States v. Fedida*,⁴⁵ Judge Dalton, of the Middle District of Florida, addressed the reliability of prosecution experts. There, the defendant was indicted for knowingly possessing and conspiring with others to knowingly possess UR-144, a controlled substance analogue.

The government proffered testimony of expert witnesses who intended to opine on the pharmacological effects of UR-144 and the substance to which it is analogous under the schedule. At a hearing, the government's experts opined that UR-144 and XLR-11 have similar effects on the central nervous system. The court stated that these opinions were based on "little more than the experts' review of the available scientific literature, which was limited to a few articles containing reports on UR-144's binding affinity to cannabinoid receptors."⁴⁶

The court also noted that the experts' opinion had not been subject to peer review or publication and that there was no evidence concerning the known or potential error rates of its experts' methodology, which allowed it—which amounted to little more than the deduction of a working hypothesis supported by a general knowledge of chemistry and biochemistry.

The court declined to rule on the admissibility at that particular juncture, but it noted that it was not inclined to permit this expert testimony. These are just a few examples of the *Daubert* issues that can arise in this context.

In conclusion, courts routinely confront novel legal issues, and there are different ways to address these issues when they are the result of the production, possession, or distribution of yet unregulated substances. One way is for the legislature to enact appropriate legislation as quickly as the need for regulation or criminalization arises, a lofty if not unrealistic goal.

The other is to guide the court with general principles of applicability. The Analogue Act was Congress's answer to the growing and evolving controlled substance problem facing the country, and the attempts at playing catchup have proved challenging. This small but unfortunately quickly expanding area of federal criminal law supports the notion that providing the court with flexibility is necessary and ultimately consistent with the legislative approach in the field as well.

PROFESSOR CAPRA: Thank you, Judge Moore. [Applause] Yes, Dr. Lander.

DR. LANDER: So I strongly endorse what you're saying, Judge Moore. I think the question is "what does flexibility mean?" Your notion of flexibility is one I very much endorse. It says that in different types of evidence, like synthetic drug evidence, the relevant question of reliability might be a different framework than in forensic feature-comparison, and within that framework, there must be flexibility.

I don't understand you to mean, though, that for synthetic drug evidence, one would have the flexibility to admit evidence that lacked any empirical basis whatsoever. So, as I understand it, the challenge here is in bringing in

45. 942 F. Supp. 2d 1270 (2013).

46. *Id.* at 1281.

expert testimony in different fields, there must be flexibility within a framework.

But, of course, the frameworks have to exist. With forensic feature-comparison evidence, we're like ten steps behind you. There are folk out there who are saying we don't ever have to have run an empirical test. We don't have to have any data about reliability. I wish we were in your case.

The challenge for the Committee will be to figure out how we avoid messing that up. I believe the kind of statements made today about the need for empirical evidence wouldn't mess up anything regarding synthetic drugs. But it's possible that you may carve off and say through Advisory Committee notes with respect to certain types of evidence, here are valuable things to consider.

My own guess is judges would appreciate understanding, for different types of evidence, the key things to consider. And I hope with regard to feature-comparison that we get to the point where you are right now with this drug evidence and the scourge it is.

JUDGE MOORE: Well, so the drug example is a particularly difficult one because it is rapidly changing, and I use that as an example because of the flexibility that the analogue statute is trying to get at, and accordingly the need for flexibility as to the admissibility of the evidence. But for the purposes of today's presentation, I think the point that I was trying to get across to the Committee members is that any change should not be so bright line that it divests the court of the discretion it needs to make that determination and not just go through a checklist of saying did you satisfy this, this, this, and this and, therefore, it's admissible. Or you're lacking this, and, therefore, it should be excluded.

And then, from the standpoint of appellate review, the court looks at the instruction that the trial court has given, and that generally it's up to the jury to attach whatever weight it chooses to the expert opinion, and that the jury will consider other issues such as bias. And whoever this expert is going to be is subject to cross-examination, whether he's employed by the FBI or if he's out in the private sector, whatever it may be. And that is what triers of fact are best at evaluating.

DR. LANDER: Sure. But you gatekeep the method. It must be a chemist who actually has measured molecules and actually knows how to work out the structure of molecules and compare them. So you actually are an absolute gatekeeper for that. And then, once somebody actually has a chemical method, yes, of course, these things go to weight and to juries.

JUDGE MOORE: My best gatekeeper function recently was when the chemist took the stand, started to be questioned by the prosecutor, and the prosecutor gave him the exhibit, ten pounds of cocaine, and he says, this is not my exhibit, I didn't test this. They had called the wrong expert. [Laughter]

DR. LANDER: Good gatekeeping.

PROFESSOR CAPRA: We have gone to great efforts to bring Professor Ron Allen in to speak to us today from Northwestern University. He's at Seton Hall today where they're doing, as it happens, a forensics conference.

Not one as good as ours, but that's all I can say. And so Ron is going to start our academic panel, and we are going to proceed from there.

PROFESSOR ALLEN: Well, good morning, everybody. It's a great pleasure to be here. I wish I could be with you and see some old friends, make some new friends, and be in a room with many people whom I admire, but thank you for tolerating this long-distance presentation.

It's a great privilege and honor to be asked to participate in even a small way in the work of the rules advisory process. The topics that you're looking at today are important, and you have a lot of brain power in the room, so much so that I wouldn't be surprised if you can solve some of the problems posed by forensics.

My comments will in a way, in a deep conceptual way, capture some of the concerns that were expressed this morning, but they're also going to point in a different direction. Almost all the comments today were pointed at the problem of admissibility to the judge making the *Daubert* decision, but there's a second and equally important aspect to expert testimony, and that's how it's used at trial.

Now, certainly, knowledge is advancing and generally reliable consensus on some of the issues plaguing the use of forensic science at trial may actually be decided in the sense that any issue that's decided in any scientific field more or less definitively resolving scientific controversies will have two important consequences in the real world.

First, certain issues will not really be the subject of litigation, notwithstanding the somewhat false belief that there can't be even partial directed verdicts in criminal cases because, in fact, those issues will have been decided. But second, certain other issues will continue to be the foundation upon which qualified experts offer opposing opinions. I'm going to return to these two possibilities in a moment. But first, what really is the conceptual foundation of the problem that you are addressing?

I think it was captured by a colleague of mine, who I shall call by the gender-neutral name Blake to mask identity. Blake speaks Romeyka, a variety of Pontic Greek named after what was once the Pontus region near the present day Turkish city of Trabzon, and all made famous by having been visited by Jason and the Argonauts and for being the home of the Amazons, as I'm sure you all know.

Blake is an ancient Greek specialist, and Romeyka is believed to be the closest living language to ancient Greek. He's a wonderful person or she's a wonderful person but is somewhat excitable. When startled, Blake often starts talking Romeyka, which for Blake is really the language of choice. Blake observed an accident involving a rather gruesome injury and was called as a witness. Blake was quite nervous about this and asked me to come to court with him or her.

And here's what happened. Blake was called to the stand and things went well through the preliminary questions. The plaintiff's counsel then showed Blake a picture of the crash scene and Blake's face turned white. The rather oblivious counsel asked Blake if Blake saw what happened, and Blake, with an effort, nodded yes, to which the counsel, of course, gave a perfunctory "let

the record show.” Counsel then asked Blake to describe what happened, which set off an uncontrollable hurricane in Romeyka. [Laughter]

The counsel tried to intervene but Blake was unstoppable. Finally, the judge intervened and said to Blake, “What’s going on here?” And Blake responded in English that the defendant was reckless, is totally responsible for the plaintiff’s injuries and should pay huge compensation, to which the judge responded, please tell us what happened, to which Blake responded with puzzled astonishment, “but I just did.”

Now suppose you’re the trial judge in this case. Would you let this testimony go to the jury in this form or would you do something about it? I think I know the answer to that question, so let me pose the next question. Why then do you let or do we let, or does the rules process let, a certain subset of experts speak the functional equivalent of Romeyka to the jury, followed by an equally uninformative assertion of who should win the case?

The use of expert testimony in many American trials is a reproach to the deepest aspirations that we hold for that form of dispute resolution. Check your own intuitions and knowledge and experience and see if it coincides with mine. The deepest aspiration that we have for trials is that facts will be found and the law applied by disinterested, unbiased, fair-minded, intelligent decision makers who will process and comprehend the evidence, deliberate upon it using all the cognitive tools at their disposal in a fair-minded fashion, and, by doing so, reach the most plausible appraisal of truth of the matter available to them.

None of this can happen if a witness speaks Romeyka, regardless of what label is attached to that witness, lay or expert. Take a conceptual step back from this conundrum and see clearly its origins. Trials are, in our system, almost exclusively pedagogical events. They’re educational events. A disinterested person, judge or juror, is educated about the pertinent events and assigned the task of doing the best that he or she can to figure out what actually happened at the time in question.

That can only occur if you understand the evidence, but much of the way in which so-called expert testimony—and I say so-called because it abuses the word “testimony” to apply it to some of the gibberish that is presented—only if that testimony is provided at trial in a way that they understand does it meet these aspirations. And the way in which it’s presented now makes comprehension largely impossible.

Under current practice, expert testimony is not provided to educate the fact finder. It is offered as a conclusion to be deferred to by the fact finder. The critical conceptual disaster at the heart of the American use of expert testimony centers on the difference between education and deference. Trials are supposed to be educational events but deteriorate into deference in many instances of expert testimony.

This highlights the critical flaw, at least in my opinion, of *Daubert*. *Daubert* took a good first step in essentially requiring the trial judge to become sufficiently educated about the proposed testimony to judge it rationally, but it did not take the equally critical step of requiring its presentation to the fact finder in the same fashion. And if I may point out,

that also captures almost this entire morning in talking about the relationship between experts and judges.

But if you have a jury, you have a next step, and that first step has no obvious relationship to the way we conduct our affairs now to the second. So you might resist this line of reasoning with a comforting, but I think false, belief that deference can occur rationally because it cannot. Putting aside extreme cases of charlatans that are unworthy of discussion, one can rationally choose to whom to defer only if one understands the pertinent field, in which case, obviously, the deference is unnecessary.

You also may resist this analysis with the equally false but conventional view that juries will defy being educated, a resistance formed in part by many of the ridiculous juridical practices that treat jurors like children in need of nannies, but I think this too is false.

The question is whether the jury can be educated, not whether each juror is, and we know a fair amount about the collective cognitive capacities of small groups. What we know is quite at odds with the disparaging view that animates some of those ridiculous judicial practices I just referred to. Indeed, if I had to take my chances on a complex factual issue with a randomly selected American judge, or a jury selected in a normal way, I'd choose the latter every time.

No matter how any particular juror scores on your favorite test of cognitive capacities, they collectively bring a wealth of knowledge, experience, and cognitive abilities that overwhelms that of a single individual. There are some studies that suggest to the contrary, but none of those have ever been conducted with an American advocate working within the adversary system whose case depends upon the ability to explain the expert testimony to the judge.

But tut, tut, tut, Professor Allen, some may say this is all very interesting, but it's so academic, where let me say parenthetically being academic is not meant as a compliment. How would a judge ever know if a jury were adequately informed about a matter? And like a typical academic, I haven't even thought about the cost. Shame on me.

These are rather weak excuses, in my humble opinion, for not doing the right thing. First, it's true that no one knows what is in the mind of another person, but we have lots of tools available to help make that approximation. Indeed, that is exactly what you do every time a judge is in the room or ask the judge in a peremptory motion, you are imagining what will be going on in the mind of the decision maker.

Here, like there, the question is not whether you know some state of mind exists or will exist but whether it's reasonable to so believe. And, yes, there would be some added cost, but the real significance of cost lies hidden today. One party can shift the costs of explanation to the other in contravention of the normal cost-bearing rules of our legal system simply by offering expert opinion testimony. If the testimony would lead to a wrong result, the opposing party has to explain why rather than the proffering party justify it.

It's also peculiar that the objection to the cost of education is limited to information that may actually be useful to jurors in their real lives, as

compared to the vast amount of education foisted upon them in the typical case that will be utterly useless, except perhaps as fodder for cocktail conversations, like the nine months the jurors in the O.J. Simpson case absorbed information that will be totally useless to them in the rest of their lives.

But maybe some issues do defy comprehension. I doubt it, but let's indulge the assumption. In those cases, the jury can't be educated, and those cases should not be tried. By the way, if you succeed in refining forensic science, that's exactly the result that you'll get. So this should not sound like a startling point. If you actually come to consensus so that there's no disagreement, that will end the litigation on those particular issues.

But in any event, to try a case where you, the trial judge, actually thinks comprehension is impossible is to turn what is supposed to be a rational system into something indistinguishable from trial by combat and ordeal. And you, in my, again, humble opinion, should not preside over such rituals, although how to avoid doing so, I haven't got the time now today to talk about.

So now you can see why the title of this brief talk, which I'm coming to the end of, refers to fiddling while Rome burns. Everyone should be absolutely clear that I'm not being disrespectful to the scientists in the room or the lawyers and law professors working hard to advance knowledge and prove its use at trial. No, not at all. They're doing, as my dad used to say, God's work.

I am being critical of the judges, the law professors, and whoever else that tolerates this mess and the rule process that perpetuates it. We should all be collectively a little bit ashamed of ourselves for allowing this paradigmatic example of irrationality to work its way into the heart of a process for which, as I said at the beginning, and I repeat, I think, and I think you agree—our deepest aspiration is that facts will be found and the law applied by disinterested, unbiased, fair-minded, intelligent decision makers who will process and comprehend the evidence and deliberate upon it using all the cognitive tools at their disposal in a fair-minded fashion and, by doing so, reach the most plausible appraisal of the truth of the matter available to them.

What can you do about it? My guess is that one result of this meeting is that you will be presented with a complex and maybe somewhat verbose proposed amendment to the federal rules. But that proposed amendment, my guess is also, is going to say, "We're quite serious about what we've said about these matters in the preexisting rules, so please take them seriously."

I suggest a quite different amendment that simply says expert testimony must be presented in a comprehensible manner. It's important to note that lousy science cannot be presented in a comprehensible manner and requiring experts to explain themselves adequately for comprehension would, in fact, erect a significant part of the barrier to the admissibility of lousy evidence, for which the Federal Rules Committee has been searching for a long time.

So thank you very much for listening to me. I wish I were there.

PROFESSOR CAPRA: Thank you, Ron. [Applause] Our next speaker is Professor David Kaye.

PROFESSOR KAYE: Thank you. I'm going to try to reinforce some comments we've heard, expand on some, and maybe even question a few, and I'd like to try to offer a slightly different perspective now.

By way of background, we've had a lot of reports dealing with particular scientific evidence of one kind or another over time, and it's interesting sociologically, if nothing else, to ask which ones had some impact in the courts and which ones didn't. Most of them I'd say really didn't overall. And that leads me to my first topic, which is the mechanisms by which courts have basically avoided what I think some of the clearer statements in *Daubert* are.

We all know the court listed a number of factors, but you would think that Dr. Lander's points about the need for empirical validation would be reflected in the very first factor laid out in *Daubert*—testing, whether something can be and has been tested. Unfortunately, the *Daubert* Court speaks in terms of whether the method can be, rather than has been, tested. That's great in the philosophy of science if you want to distinguish metaphysics, religion, or something from law. It's not so good for telling what is established or, in the words we've been using, reliable science, where actual testing is required.

And so courts have relied on that difference, and I'm talking about court of appeals opinions here from the Third Circuit, the Seventh Circuit, to basically say things like, well, there has not been actual testing, but still it can be tested, so *Daubert's* been satisfied with respect to testability. Adversarial testing has been substituted for the kind of empirical testing with situations in which knowing the truth or an approximation of it at least is available.

The next *Daubert* factor is peer review and publication. Courts have deemed peer review to be satisfied in the sense of having a second examiner look at a first examiner's results. That's not what the *Daubert* Court was talking about when it referred to peer review as part of a scientific process that involves publication.

The third *Daubert* factor is publication. Well, I am a member of a committee within the recently formed OSAC,⁴⁷ which Susan Ballou knows very well and is related to NIST. And I've got to say, I see technically merit-based forms of proposals that come through listing fifty publications, but there's no systematic structure for looking at which of those publications are really scientific journals, which of them are trade journals, and what they actually contain.

The fourth *Daubert* factor is the known or potential rate of error. Again, adversarial testing. We heard Dr. Lander refer to former testimony from an FBI witness that we don't know of any error, and courts have said adversarial testing is good enough; it's worked in the legal system for a long time. That may be some indication of truth, but it's not scientific truth, right?

The next *Daubert* factor is standards and controls. Now here I'm going to get to something that I think is very significant, aside from the fact that the

47. *Organization of Scientific Area Committees for Forensic Science*, NAT'L INST. STANDARDS & TECH., <https://www.nist.gov/topics/forensic-science/organization-scientific-area-committees-osac> [<https://perma.cc/VD7S-DCHD>] (last visited Feb. 14, 2018).

problem is courts are too willing to say there are standards without asking whether they are controlling—whether they control the discretion of and the judgment process that Dr. Dror was talking about, to include protective mechanisms and to limit the amount of subjectivity. Some subjectivity is always present. And we are still getting statements about no uncertainty in identifications in standards.

To reinforce a point about chemists and others, one standard that exists and that is also being discussed now within OSAC defines in respect to spectrometry comparison of infrared spectrum. This could also apply in a drug identification case that Judge Moore talked about. What is a meaningful difference? It is a property that does not—of a sample that does not fall within the variation exhibited by the comparison sample.

Well, the comparison sample has no variation. It's a sample. It's got peaks. I finally figured out what they meant was replicate measurements on the same sample. There was no number of replicate measurements, no standard for figuring out how much of that difference is meaningful. When asked to provide that information, the answer I got was, "oh, that would be difficult to do." Well, all right, so standards, just because they exist, don't necessarily prove a lot.

The final factor—degree of acceptance in the scientific community. In *Daubert*, what have courts done for forensic identification? They've said, if it is accepted by the forensic identification community itself, that will do. And Judge Posner said, an expert on art style is making a comparison between a given painting to see if it's a forgery with other paintings and we don't require any scientific showing because Rule 702 is not limited to scientific experts—so the evidence ought to come in, you know, it'll be good enough.

And that is not a ridiculous position to take, but we need to be very clear then about the fact that it's stylistic, impressionistic evidence and not "scientific" evidence.

If you're interested in trying to encourage the implementation of the rule, what Ron Allen was referring to, "we mean what we said earlier," or even to modify it, here are my suggestions.

First of all, with respect to terminology, use the words reliability and validity or less ambiguous terms as they are used in statistical and social science rather than trying to adapt them to legal purposes, as I'm afraid the PCAST report⁴⁸ does. There can be validity to theories. That doesn't mean you have the kind of empirical testing that Dr. Lander referred to necessarily, although there has to be some behind at a level of theory, a theory like there's a lot of variation. There's more intersampling. There's more variation among different samples than within repeated things. You know, those are plausible theories.

And then there's the problem that, what is it we're trying to get at when we talk about validity. Validity has to do with fitness for purpose, with using

48. See generally PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

measurements for a particular purpose, not just the quality of measurements. And if that's what it means, what is the purpose we're talking about here? Is it we've got to be 95 percent confident in some sense? Or is it that the level of uncertainty in what is relevant evidence needs to be properly presented in one way or another?

And these ideas of validity and reliability, as they are used scientifically, where reliability refers to reproducibility and repeatability of measurements typically, rather than everything that's mashed together in Rule 702⁴⁹ under the word reliability. In any event, if we think about a polygraph, for instance, it's a valid technique. You could say the measurements are valid if you're interested in knowing a level of the elevated—of the heartbeat, the systolic blood pressure, the level of sweat, and you can measure stress perhaps. For that purpose, it might be valid, but for telling whether somebody is lying, it would not be valid for that purpose.

Next, the question of conclusions is vital. This is something that I think Professor Allen was trying to get at—the process of simply giving deference to the conclusions of the analyst. Or should the analyst also be giving statements about hypotheses or about the evidence?

Now what I'm going to get at here is a whole range of writing by forensic statisticians that has not been part of the discussion here. The discussion here has been on classification. The expert is going to say it's this or it's that. And then we're going to ask “what's the error rate associated with that?” Or better yet, “what's the sensitivity and specificity?” Or better yet, if we're talking about classifications, “what's the likelihood ratio for a positive result?” If you go back to hair identification, for instance, the false positive error rate is not 11 percent, it's 35 percent.

Eleven percent is the posterior, we're flipping things around. But the likelihood ratio, that is the probability of observing hairs from—of the evidence with hairs from—the evidence being the examiner's conclusion from the same source, compared to hairs from a different source, is about 2.5.

There is some probative value, frankly, in hair comparisons based on the FBI study, and our question becomes whether or not to structure testimony to say if you stick within the limits of what's known about a likelihood ratio, that is acceptable, and if you're going to rely on the idea that the analysts aren't really doing science as in some opinions, they're just experts, it ought to come in, it's useful, it's relevant, that's okay.

But then I think it's very important to make it clear that the expert is not permitted to talk about a scientific method that's being performed.

PROFESSOR CAPRA: Thanks, David. [Applause] Our next speaker is Professor Jay Koehler.

PROFESSOR KOEHLER: All right. Ordinarily, redundancy is not a good thing in conversation, but I think here it may be a good thing. It might show that we're saying some of the same things, and maybe some of the key points

49. FED. R. EVID. 702.

are getting highlighted. I say that because I may not have anything very original to say given the eleven talks that came before me.

But here's what I do want to say. At a high level of abstraction, I would say that Rule 702⁵⁰ is a good rule. By focusing on the reliability of what an expert says, by discouraging speculative opinion statements, by requiring that all expert inferences be supported by sufficient facts and data, it embodies the good intention of only letting truly helpful information from experts reach jurors, who, studies show, tend to give great credence and weight to what those experts say. But in practice, in matters concerning forensic science evidence, Rule 702 falls short.

Forensic scientists routinely offer opinion testimony that reaches beyond the bounds of what the available data back up and that falls well outside of what could be construed as helpful to the trier of fact based on Rule 702(a). For example, when a forensic expert finds that there are features associated with a latent fingerprint, a handwriting sample, a tire track, or a bite mark that appear to correspond with those found in another fingerprint, handwriting sample, or tire or dental model, there's no basis in existing facts or data for that expert to expand on this finding by offering his or her opinion that the pairs of prints or markings share a common source. I say this because there are virtually no available scientific data that speak directly to the accuracy of these opinions.

This point was made by the PCAST report,⁵¹ of course, the 2009 National Academy of Sciences report,⁵² and many academic papers that preceded these reports. These reports and other papers spoke of the need for proficiency tests that will help provide estimates for the rate at which errors, false positive errors in particular, occur in the forensic sciences under various conditions.

But this point about the lack of scientific data that speak to the accuracy of forensic-science conclusions and opinions has been met with fierce resistance from various quarters in the justice system. Opponents of this claim argue that there have been hundreds, if not thousands, of articles published in peer-reviewed journals, many forensic-science journals, over the past century that validate the science and that forensic examiners frequently take proficiency tests and they do extremely well on the whole. And courts rely heavily on the results of these proficiency tests in many areas, such as ballistics, as proof that *Daubert's* admissibility factor that concerns known or potential error rate is satisfied.

The problem, though, is that the word "test" or more specifically the phrase "proficiency test" has been used to mean two very different things. Now, for simplicity, I've called the two types of proficiency tests type I and type II, with apologies to the statisticians in the room.

A type I proficiency test is an indicator of competency, so according to the FBI, proficiency testing is a quality-assurance measure used to monitor

50. *Id.*

51. See generally PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

52. See generally NAT'L RESEARCH COUNCIL, *supra* note 1.

performance and identify areas in which improvement may be needed. Type I proficiency tests are designed for internal purposes, such as providing information about the effectiveness of training programs, laboratory procedures, and ensuring basic competency for the examiners.

And examiners' good performance on this test—and generally, the performance has been excellent—leads courts to infer that the error rates are low. But this isn't what scientists are referring to when they speak of the need to conduct proficiency tests to identify the accuracy of forensic science claims. I think Dr. Lander made that clear, but I'm saying it again here. Instead, I think scientists are referring to type II proficiency tests—tests that are expressly designed to tell the outside world how accurate forensic scientists' conclusions really are under conditions that approximate those in casework.

The distinction between these two types of proficiency tests is crucial. A type I test tells us next to nothing about accuracy and error rate because they're not designed for this purpose. The external manufacturers of these tests say so expressly in the tests that they provide. A type II test, on the other hand, expressly takes on the error rate estimation question and is designed accordingly. With few exceptions, type II tests have not been conducted in the forensic sciences. We've heard a lot about one test today. That itself is telling that we're talking about one black-box test. There should be dozens. There should be hundreds. They should span across many of the forensic fields, but they don't, or at least the data from those is not publicly available.

Therefore, I would say we know next to nothing about the accuracy rates of our forensic methods and conclusions. When a tire-track examiner opines that a particular tire made a particular tire track based on observed similarities, we have no basis for assessing the accuracy of that opinion.

Let me hold off on that for just a minute. Framed as a Rule 702 issue, there are insufficient facts and data under Rule 702(b)⁵³ to support the reliability of the tire-track examiner's subjective source opinion. There is insufficient evidence that the underlying tire-track principles and methods promote reliable opinions, as Rule 702(c) requires,⁵⁴ and we can't say that the expert who is the first source for matching features has reliably applied the method in the instant case, as Rule 702(d) requires.⁵⁵

So, in short, the tire-track examiner's source opinion is not helpful to the trier of fact under Rule 702(a)⁵⁶ beyond the information the examiner provides about specific areas of consistency or inconsistency between a tire track and a single possible source tire.

Trial courts have consistently proven themselves unworthy of our trust when it comes to enforcing the Rule 702 mandate that speculative, unsupported forensic-science source-opinion testimony should be

53. FED. R. EVID. 702(b).

54. *Id.* r. 702(c).

55. *Id.* r. 702(d).

56. *Id.* r. 702(a).

excluded.⁵⁷ Consequently, the forensic-science community has little reason to either test its fundamental claims or to rein in strong source opinion testimony that the examiners routinely provide at trial.

So, to address this problem, Rule 702 should be amended. Specifically, I'll put something out there for criticism. Rule 702 should be amended so that it unequivocally states something like the following: Expert scientific opinions that are substantially based on subjective judgments are inadmissible. Frame it this way, inadmissible, unless what Dr. Lander wants is so, unless accompanied by reliable data from empirical tests that assess the accuracy of similar opinions offered under relevant conditions.

If the Rule is revised in this way or something like this, we would begin to see admissibility hearings with teeth. The forensic-science community would likely start taking its testing responsibilities seriously, and we would see a substantial increase in transparent, high-quality scientific studies that assess the accuracy of forensic-science conclusions. Once such accuracy data become available, judges and jurors will be in a position to assess the probative value of the conclusions and opinions of forensic examiners and other experts who offer subjective judgments and opinions at trial. Thanks. [Applause]

PROFESSOR CAPRA: And so next we have Professor Jane Moriarty to speak.

PROFESSOR MORIARTY: Thank you so much. I'm truly honored to be here. This, for an evidence scholar, this was coming to Mecca to speak to you, so thank you for the work you do, thank you for including me. My title is "Deceptively Simple—Judicial Gatekeeping of Forensic Science Feature-Comparison Evidence."⁵⁸ I chose it is I think it describes what part of the problem is about forensic feature-comparison-method matching in Rule 702. So I just want you to know that you can go on *New Scientist* and it asks you to figure out whether you can compare samples visually and match them. And, in fact, according to *New Scientist*, they do match and I can look at points of distinction, but what do I know, right?

We should trust the experts. However, intuitively, as humans, we think we can do this. We really do. We are by nature pattern matchers. It's part of how we function in the world. So it's something we think makes sense intuitively. We believe juries can evaluate these meaningfully. I've seen judges write a lot about that, jurors can reexamine this pattern-matching evidence, unlike DNA or epidemiology, which we don't believe they can really access very well. And it's also, some judges say, just soft science, which doesn't require a *Daubert* analysis.

And I keep thinking about why that is, and I think it's because it's deceptively simple, by which I mean it's not simple, and we're not recognizing the complexity. The NAS report explains to us that the

57. *Id.* r. 702.

58. See generally Jane Campbell Moriarty, *Deceptively Simple: Framing, Intuition, and Judicial Gatekeeping of Forensic Science Feature-Comparison Methods Evidence*, 86 FORDHAM L. REV. 1685 (2018).

conclusion of a match is a complicated question of both science and statistics, and it spells out exactly what is necessary to be able to declare a match.⁵⁹ It's enormously complicated. In the PCAST report, one of the conclusions is that without an appropriate estimate of its accuracy using appropriate empirical testing, which we've heard a lot about today, a metrological method is useless because no one has any idea how to interpret a match.⁶⁰

These are very serious concerns, and why are they important? Because, if we even look at a handful of the data on wrongful convictions, approximately 60 percent of them involve a faulty forensic-science conclusion. And that should make us all stop and think quite a bit about whether we're getting this right and why we aren't.

So here's my thought, and I've been thinking about Michael Saks and I wrote an article back in 2005 on forensic science for the ABA's *Judges' Journal*,⁶¹ and we expected to be inundated with complaints and questions, comments, critiques. We heard nothing.

And so I've been writing about judicial decision-making about various forms of science since that time, and I'm thinking the problem here is we believe this is a simple problem to solve and we use simple problem-solving methods. Let me explain. If we frame the problem as simple, we use faster intuitive decisions. So, for the members of the Committee, think of this as our excited utterance objection, right?⁶² We all go, I've got that, right? Right away, we know we know how to do it; we know how to evaluate it end to end.

If we frame a problem as complex, we have a slower, more deliberate response, and I'm going to show you in the cases how that's true. And, again, for the Committee, with all due respect, I give you as an analogy the prior-consistent-statement exception with a temporal requirement.⁶³ We know that's a complicated, detailed thought process to get through. And then we have *Daubert* on remand, which Judge Kozinski, in his *Brave New World* analysis, is talking about relative risk ratio.⁶⁴

It's two entirely different ways of thinking. We have thinking slow; we have thinking fast. I'm sorry, I'm going to use type I and type II in a totally different way to confuse you. [Laughter]

So this unconscious and fast thinking versus this slower, more deliberate doubting-type thinking. The latter one costs us mental energy, and we all try to conserve energy. So, when we think we have it, we use a simple mental shortcut called a heuristic.

We do it all the time. The problem is they sometimes lead to characteristic errors or biases. And by biases, I mean not that you hate people but that cognitive biases affect humans when we make quick decisions. So our

59. NAT'L RESEARCH COUNCIL, *supra* note 1.

60. PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 140–41.

61. See Jane Campbell Moriarty & Michael J. Saks, *Forensic Science: Grand Goals, Tragic Flaws, and Judicial Gatekeeping*, JUDGES' J., Fall 2005, at 16.

62. See FED. R. EVID. 803(2).

63. See *id.* r. 801(d)(1)(B).

64. *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1321 (9th Cir. 1995).

framing effect is how we look at the problem. If it's simple, we have a number of biases that can creep in. One is belief of perseverance where we continue to think the same thing, even when we're shown evidence to the contrary, to stay the course. And another is confirmation bias—we seek out information that confirms what we know.

So I'm going to put these up, you can look at them. [Referring to a PowerPoint presentation] These are all really quick decisions that you'd use to justify. They're not, for example, what Judge Kozinski did in *Daubert* on remand,⁶⁵ where he goes through a detailed analysis of teratology and epidemiology or what we saw in Supreme Court cases on medical causation, for example.

What we have here is a long history of use. It seems like these are simple things to resolve. These are normal approaches if we think it's simple. I'd also say this is what happened, by the way, when we are thinking about things like astrology in the courtroom.

Every one of these works. Or if we think about phrenology, it had its own journal in the 1850s, and it had a long history of use. We are faithful to its traditions. And we consulted with the practitioners. It's a classic mistake we all make as humans when we think something's simple.

So one of the great examples of this is the NAS report that says, in substance, there is no scientific support for the use of hair comparison.⁶⁶ You've just heard about the high error rate. This was brought to the court's attention in a case. I mean, there's tons of cases. This is just one. And it said it's scientifically reliable. Judges may even take judicial notice of it. And in that case, the court also wrote about the NAS report, but it just dismissed the concerns.

Why? They assumed the problem was a simple one and they used heuristic thinking, you know, these sort of shortcut methods, not the long, complicated sort of *Daubert*-style method. So how do we encourage judges to think about this matter as having greater complexity?

We have some proposed changes, and I very much like Jay Koehler's. I think that does get to the problem. We just switch it. It's not admissible unless. The problem, of course, is the liberal thrust of the rules, which seem to counter that. But I like the proposal very much because I think it's "a stop and then decide" mentality. So why don't we encourage judges to think of this as complex? And I think the proposed tweak to Rule 702 would do that.

I also think that if we require judges to move more slowly through the evaluation of forensic methods the way they do in teratology cases or when they decide complex causation cases, and, by the way, these are the same judges. They're all doing it. To shift methods, right? To think of this as the complex problem. And the draft of the proposed Rule 707 would certainly do it, although my guess is that we're going to hear problems from the behavioral-science experts, you know, why them, not this group, and also medical testimony, now what? What do we do with doctors?

65. *Id.*

66. See NAT'L RESEARCH COUNCIL, *supra* note 1, at 155–61.

But the better approach, I think, is simply better-grounded forensics, which meet a lot of these requirements, and then we can go back to a less rigorous analysis. So thank you all very much. [Applause]

PROFESSOR CAPRA: Thanks, Jane. Our next presenter is Professor Erin Murphy.

PROFESSOR MURPHY: I echo what Jane said about coming to Mecca. It's a real privilege to be here given the issues that we work on. I'm going to shift the conversation slightly because I want to think about a particular set of challenges that are emerging at an increasing rate, and that is the challenge of machine-generated evidence and how that interfaces with the rules.

The rule that we have is one that even in its literal terms is really thinking about human beings as the kind of ultimate vector through which information is acquired and delivered. It talks about testimony, which is something humans do. It talks about a witness, which is a human being. It's not really thinking about evidence that simply comes from a machine, is created by a machine. It might even be knowledge that is unique to the machine, even though it was ultimately programmed by humans.

The adversarial system is one that ultimately is about the human contest and what happens inside of a particular trial or a courtroom to challenge evidence. And when you think about machine-generated evidence, it may be that those human rules for uncovering malfeasance, mistakes, misunderstanding, don't function as well.

And when I say human versus machine evidence, I don't mean to suggest that these are really easily divided categories, nor that this is a particularly new problem. We can think of something as simple as a date stamp on a surveillance video in a public housing project, with a time attached to it, and think about, what is that evidence? Is that machine-generated evidence, the computer or the digital program that records created the knowledge of what time it was when the observations on that date stamp were made?

We can think within the context of the evidentiary rules about how we might try to grapple with that as an authentication question. Is this tape really what it purports to be? As a hearsay question, is this date stamp the sort of truth of the matter asserted as time? As a confrontation problem, how do we confront a date stamp on a video? Or, obviously, in this case, as a *Daubert* problem,⁶⁷ what is the reliability of this piece of information or evidence?

And you see this playing out in cases, most prominently in the confrontation context and in the hearsay context, where courts are really struggling to understand, what is that information? And you see the Supreme Court Justices dueling over whether the witness who takes the stand is just a scrivener of this electronic or digital information, or are they something else? And how should we contemplate that? And this is particularly important for machine-generated evidence because it can evade all of those questions and all of that scrutiny because sometimes you don't even have an actual witness.

67. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993).

We knew that before *Melendez-Diaz*⁶⁸ in the confrontation context, literally, there'd just be a document that's produced, and under the hearsay rules, it's admitted. And there wasn't really much thinking about how you go about a *Daubert* challenge in that context. And even after the confrontation cases where an expert witness must be produced, who that expert witness is and how they relate to the evidence that's machine-generated and presented is complicated.

Do they just need to be able to say I know how to read a machine? Do they need to understand the particularities of that machine? I know that we calibrate our date stamps all the time. Do they need to have even a deeper knowledge? I know the algorithms or the functioning of how this machine was programmed to have date stamps.

What kind of level of complexity or knowledge do we demand of the individual witness? So, when I say I am trying to give a broader picture of machine-generated evidence, I'm thinking of everything along that line of date stamps and so forth to GPS pins that are dropped or GPS knowledge that comes out, to cell phone tracking devices, to probabilistic genotyping systems and all the way down the line.

Now I want to make two distinctions or characteristics about machine-generated evidence that pose a special problem. One is that the 1.0 version of this was that the machine was a tool that was used by humans, the same way a calculator might be a tool used by a human. The human can do the thing; it's just that the machine does it faster or does it more accurately or more precisely than a human would. But increasingly, we're seeing machine-generated evidence that doesn't actually reproduce or speed up what humans do, it creates unique knowledge.

To cite an example in the area I know best in this regard, I would say probabilistic genotyping systems. These are computer software algorithms that do computer-simulated models that a human being could not do. It would take too long, it's too complex, there are too many dimensions to the variables. And so, when a computer spits out the probability of a DNA match, it's spitting out knowledge that only the algorithm and the computer have. It can't be independently verified or checked by an individual human being. And so it's unique, machine-based knowledge.

The second dimension, I think, about machine-generated evidence that's important and has a parallel to ordinary human expertise is, who creates this knowledge? Is it exclusive to the criminal justice system, or is it something that is in the market or in the general world and we're just harnessing it within criminal justice?

To just give some examples, a probabilistic genotyping system is pretty exclusive to the criminal justice system. It's using markers exclusive to criminal justice purposes. The vendors and consumers of that are all criminal justice-related agencies, as opposed to say a GPS pin in Google, where Google is a private company that's doing a lot more with its algorithms and software than simply providing criminal justice information. You're just

68. See generally *Melendez-Diaz v. Massachusetts*, 557 U.S. 305 (2009).

reaching in and harnessing it. And that distinction is important because it has, I think, a lot to tell us about the risks and also the design features of the tool that's being used.

We've heard a lot about accuracy and reliability. And David Kaye said we've packed so much into that word, reliability. And I want to try to think for a minute just of how we might unpack that as it regards machine evidence.

We might think of the method, the theory that we have that we're trying to implement through the machine. We might think of the actual implementation of that method in the source code. What is the source code doing to implement what we are theorizing about?

We might think about the validation of the machine, both as a machine that when you run that source code it operates correctly, as well as a machine validated within certain perimeters. This is the context in which it can operate, on these kinds of DNA samples but not on those. We might think of the calibration of that machine. So this machine generally works and it works in these parameters, but no one's cleaned it for the last five years, and now it's not working so great anymore because it's not regularly checked and calibrated.

We, of course, would think about the execution of the technology and the machine by an individual and whether there might be any subjective inputs or any sort of human interface with the machine knowledge. And then lastly, of course, we think about the actual human being coming into court and reporting on what they know about the machine and where the machine comes from.

Each of these components, I think, is important to test and safeguard within the context of a *Daubert*-style analysis. And yet, it's easy to avoid all of them by simply saying that the machine is a mere scrivener. Close the curtain; we're done. Or saying, you know, it's a machine; therefore, as long as the theory sounds right and as long as there was a validation test somewhere, sometime, at some point, we don't have to do all those intermediate steps along the way.

So I would encourage the Committee in thinking about this rule, and let me add one piece. Comparatively, in the United Kingdom, they do have a rule that talks about what you do with evidence where there isn't basically the human vector, there's simply the machine-generated evidence. And their rule is criticized and incomplete. I think this Committee could think more completely about how you would operationalize a *Daubert*-style test where the machine is the unique source of knowledge, where the human being might be just a vector delivering that knowledge, including elevating the ideas of transparency in each of those pieces and disaggregating all those pieces.

And the last thing I would say is that, as with other areas of the evidence rule where a Rule 403 prejudice/probative-style balance is explicitly written into the rule,⁶⁹ it also might be the case that recognizing the unique power of machine evidence to influence jurors or to have a kind of strong sway over

69. FED. R. EVID. 403.

jurors, scientific evidence in general, this would be an appropriate place to insert that kind of special balancing as well. Thank you.

PROFESSOR CAPRA: Thanks, Erin. [Applause]

Apropos of machine-generated evidence, there's an interesting opinion by Judge Kozinski about Google Maps and Google pins that I recommend to you as to whether it's hearsay or not. Look up "Kozinski Google pins" and you'll probably find it.⁷⁰

Steve Saltzburg is our last academic. Steve?

PROFESSOR SALTZBURG: It's really nice to be here. As some of you know, Dan and I are coauthors, and he made me write a paper, so I did, but I only speak for myself today because I'm going to say a couple things that may be stronger than Dan would feel comfortable addressing.

Number one, I want to talk about the proposed or the possible Rule 707, is it a good idea? No, in my judgment. Why? Number one, the word "forensic," which we've used a lot today, we have forensic everything these days, right? We've got forensic accountants. We've got forensic psychiatrists, forensic psychologists. We've got forensic anything that requires expertise. They call themselves forensic, which really means they do it for a living, I guess, for the most part.

But I think if you're talking about only matching samples, the requirements in the working draft might make sense, but they should be insisted upon by judges right now under Rule 702,⁷¹ in my judgment.

The second reason for not adopting this rule is because all that's going to happen is that the first sentence in an expert report is going to be, "my method is repeatable, reproducible, I'm capable of applying the method reliably, and I accurately state it on the basis of empirical evidence, what it is I stated." They will just quote the rule. And then you'll have the same kind of reports that you get now. I just don't think it's going to advance the ball.

And the third reason for doing it is unless you're going to end up adopting new rules for every kind of expert testimony, you just clutter things and don't leave it to judges to do their job.

The second point I want to make is, what accounts for the failures of forensic evidence or scientific evidence in the courts? Is it the judges not doing their job? I think one of the problems we have is lawyers don't understand exactly how many ways experts can be wrong. I mean, we've heard this today again and again and again. It's not just that the technique itself may be overstated. It's also whether this individual is actually capable of doing that test correctly, whether the person did it correctly in this case and interprets it correctly. All these matters are subject to possible challenge. And yet, too many lawyers out there don't seem to get the point that nobody thinks it's competent representation to fail to investigate the facts.

Everybody knows we're supposed to investigate the facts of the case. So many lawyers, though, assume that the forensics are good. And one of the

70. *See generally* United States v. Lizarraga-Tirado, 789 F.3d 1107 (9th Cir. 2015).

71. FED. R. EVID. 702.

reasons they do is they sometimes don't have their own expert. Now I talked to A.J. Kramer about this [the Federal Defender member of the Advisory Committee]. He says Federal Defenders have access to forensic experts. I know that they don't always use that access in all cases. Lawyers appointed under the Criminal Justice Act⁷² generally have to petition the judge. They have to move for the appointment of an expert.

Some judges are more liberal in granting funds to hire experts than others. It's a big problem in state courts. A typical conversation more in state court than federal goes like this. The judge says "you've made a motion for appointment of an expert, why?" "Well, because, Judge, I don't know anything about this field." And the judge goes, "well, have you had any problem with the government's expert?" And the answer is "I don't know whether I have a problem because I don't know anything about this field." And the judge says, "well, is there anything in particular this expert did that gives you some pause?" "Judge, I don't understand the particulars of what this expert did." And the judge says, "well, if I give you an expert in this case, I would have to give an expert in every case in which there's a government forensic expert." "Yeah." And my answer is "yeah, that's right." That's exactly right, shocking as it is. I believe that's a part of the system.

So what else is a problem in criminal cases in federal court that affects forensic evidence? One is that we seem to take the requirement of preparation more seriously in civil cases. I think *Daubert* is working fairly well in civil cases because the requirement of pretrial disclosure of an expert report is so detailed and really enforced by judges that if cases are going to go to trial, the experts are usually deposed, and there are competing experts. The vetting is pretty good.

In criminal cases in federal court the reports aren't the same. They're usually not as comprehensive, and so often they don't get the same kind of vetting or testing as they do in civil cases. And I think we can do better on that.

I don't like Professor Koehler's amendment. I wouldn't mind it so much if it were applied only to matching, but applied to expert testimony across the board, we wouldn't have differential diagnoses and we wouldn't have a lot of the mental-health testimony that we get in cases because that's all basically subjective.

I don't think it would be a bad idea, by the way, for the Advisory Committee and the Standing Committee to do something I've thought about and recommended over the years, that is, to amend the Advisory Committee notes, add to them from time to time, even if you're not changing the rule or—

PROFESSOR CAPRA: The relevant statute does not contemplate Committee notes without an accompanying rule change.⁷³

PROFESSOR SALTZBURG: You should be able. Then the alternative is to send an advisory out. Here's something federal judges should think

72. 18 U.S.C. § 3006A (2012).

73. 28 U.S.C. § 2072 (2012).

about. And I want to give an example of this. My opinion is the single greatest revolution under the Federal Rules of Evidence occurred four years ago, and it's been evolving ever since, and it's a long time coming. And that is for the first time, a couple of federal appellate courts said looking at Rule 404(b),⁷⁴ which I think most of the judges would say that if you look at a criminal case, the three rules that matter the most are Rules 404(b), 801(d)(2)(E),⁷⁵ and, to some extent, 702.⁷⁶

But these appellate courts under Rule 404(b) for the first time said what a trial judge needs to do is to require the proponent of the evidence to articulate how you get to the purpose for which you say the evidence is offered without going through a propensity analysis. Tell us how you get from the evidence to the purpose, knowledge, intent, without going through propensity.

It turns out if you apply that principle, there are thousands of cases that are wrongly decided and have been since 1975. It's a revolution. Is it noted? Most of the circuits continue to ignore that suggestion, but where it's applied, it's made a huge change.⁷⁷ Somebody, I think, in the Advisory Committee, Standing Committee rules-making process should be able to call attention to major developments like that.⁷⁸

And the other thing is, the Supreme Court of the United States has been slow to recognize the importance of giving an indigent defendant the assistance of an expert. The *Ake* case⁷⁹ provides a right to an expert under some conditions, but it's the rare case where federal and state judges have said there's a constitutional right to have an expert appointed. The bottom line in most of the decisions is there's got to be something that you can point to that's uniquely different about that case and that supports that requirement.

I think what we've heard today is that it's every case in which scientific forensic evidence is offered that lawyers need help. If they don't think they do, they're living in a world of arrogance that I don't know or understand. Thank you, Dan.

PROFESSOR CAPRA: Thanks, Steve. [Applause]

So that's our academic panel. We now move to practitioners, and our first speaker is Ted Hunt from the Justice Department.

MR. HUNT: All right. Good morning. My name is Ted Hunt. I'm Senior Advisor on Forensic Science at the Department of Justice. The Department, of course, shares the goals underlying the PCAST report⁸⁰ that feature-

74. FED. R. EVID. 404(b).

75. *Id.* r. 801(d)(2)(E).

76. *Id.* r. 702.

77. *See, e.g.*, *United States v. Gomez*, 763 F.3d 845 (7th Cir. 2014); *United States v. Caldwell*, 760 F.3d 267 (3rd Cir. 2014).

78. The Advisory Committee is in fact considering proposed amendments to Rule 404(b). Memorandum from Daniel J. Capra, Reporter, Advisory Comm. on Evidence Rules, to Advisory Comm. on Evidence Rules, Consideration of Possible Changes to Rule 404(b) (Apr. 1, 2017), in ADVISORY COMMITTEE ON RULES OF EVIDENCE APRIL 2017 AGENDA BOOK 306, 306 (2017), http://www.uscourts.gov/sites/default/files/advisory_committee_on_rules_of_evidence_-_spring_2017_meeting_materials.pdf [<https://perma.cc/DQX3-3VTR>].

79. *See generally* *Ake v. Oklahoma*, 470 U.S. 68 (1985).

80. *See generally* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

comparison methods used by forensic experts must be scientifically valid and that forensic results and conclusions should not exceed the scientific and epistemic limitations of the discipline.

However, the Department does reject PCAST's premise that there exists a singular and exclusive means by which to establish the foundational validity of these methods. There's also been no principled explanation that I have heard offered about why forensic science alone should be subjected to a different rule of evidence than any other applied science.

Potential revisions to Rule 702⁸¹ contemplate that an expert's testimony, if first determined by a trial judge to be forensic in nature, would then be reviewed under both the existing standard and additional admissibility criteria while the testimony of nonforensic experts who may also proffer opinions on things such as source- and activity-level propositions and who also testify to ultimate issues in both civil and criminal cases would be judged under the present standard.

Many problems will arise, not the least of which are definitional in nature, and indeed, the written materials that were submitted in advance of this symposium were at a loss to determine exactly where to draw that line, whether and under what circumstances a rule targeted at forensic identification would also entail litigated issues and other applied disciplines, and trial judges, of course, would be faced with the same conundrum.

Briefly, by way of background, the *Daubert* decision states that in a case involving scientific evidence, evidentiary reliability will be based upon scientific validity.⁸² PCAST, in its report, takes the term scientific validity from *Daubert* and divides it into two parts, which it describes as foundational validity and validity as applied.⁸³

The report then equates its phrase, foundational validity, with *Daubert*'s term, scientific validity. PCAST then uses the phrase foundational validity to describe the scientific standard that corresponds to the legal requirement in Federal Rule of Evidence 702—that that evidence must be based upon reliable principles and methods.⁸⁴

The report then, on pages fifty-two and fifty-three, purports to describe a nonseverable six-part test for establishing scientific validity,⁸⁵ which declares, “The sole way to establish foundational validity is through multiple independent black box studies that measure how often examiners reach accurate conclusions across many feature-comparison problems involving samples representative of the intended use. In the absence of such studies, the feature comparison method cannot be considered scientifically valid.”⁸⁶

Now, the department believes that the set of criteria collectively asserted by PCAST, if offered as a means by which foundational validity of forensic

81. FED. R. EVID. 702.

82. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 794–95 (1993).

83. PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 42–43.

84. *Id.*

85. *Id.* at 52–53.

86. *Id.* at 68.

feature-comparison methods could be established, is not necessarily wrong or ill advised. But what we believe is wrong and ill advised is PCAST's novel premise that the set of criteria that comprise its nonseverable six-part test collectively constitute the exclusive means by which scientific validity of a feature-comparison method can be established.

Before the release of the PCAST report last fall, September 2016, the Department knew of no scientific-discipline-specific multipart validation test published by any group that claimed to establish the exclusive means by which the foundational validity of an applied scientific discipline could be established.

The Department believes that the PCAST test for scientific validity is one view but an extremely narrow view—the means by which scientific validity for purposes of establishing evidentiary reliability can be established. Mainstream scientific thought is not so narrow and prescriptive. Instead, it's consistent with the principle that all available information, evidence, and data derived from studies of diverse and varied experimental design can be appropriately used to inform questions of method accuracy, precision, and fitness for an intended purpose.

Consistent with this mainstream approach, just last month the American Association for the Advancement of Science published a study on latent fingerprint examination,⁸⁷ and the study by that group disagreed with PCAST's premise that only those research papers that PCAST considered to be intentionally and appropriately designed could be considered when determining whether sufficient evidence supported the foundational validity of a feature-comparison method.

Instead, the association explained the concept of convergent validity—drawing conclusions from the literature as a whole, recognizing that studies will have different strengths and limitations, different strengths and weaknesses, and that the various publications, each with distinct limitations, when considered in isolation can reinforce each other and collectively support conclusions that would not be warranted on the basis of a single study.

Now the Department agrees with the mainstream approach of the association and with the consensus of scientific thought that the body of literature bearing upon issues of foundational validity must be considered as a whole and not in an atomistic fashion, as PCAST has proposed and which the American Association for the Advancement of Science has recently rejected.

This holistic approach to validation is embraced by the international scientific community. The International Organization for Standardization, or ISO, is the preeminent international body for developing and publishing consensus international standards and guides. ISO 17025 is the international

87. See generally WILLIAM THOMPSON ET AL., FORENSIC SCIENCE ASSESSMENTS: A QUALITY AND GAP ANALYSIS (2017), https://mcprod.aas.s3.amazonaws.com/s3fs-public/reports/Latent%20Fingerprint%20Report%20FINAL%209_14.pdf?i9xGS_EyMHnIP LG6INIUYzb66L5cLdlb [https://perma.cc/Q5EP-8264].

standard that governs the general requirements for the competence of testing and calibration laboratories around the world involving clinical testing, research, as well as forensic science.⁸⁸ The same rules apply to each type of lab, no matter what its specialty or whether it's testing clinical samples, groundwater, or forensic evidence for that matter.

These standards were created by the scientific community for the scientific community. ISO states that labs must confirm that their methods are fit for the intended use and requires that method validation be as extensive as is necessary to meet the needs of a given application or field of application.⁸⁹ So that's in direct contrast to PCAST's six-part test.⁹⁰ The ISO standard is not prescriptive as to how labs must validate their methods, which criteria must be included, or what experimental design must be utilized.

As a corollary to this nonprescriptive approach to validation, the Advisory Committee's materials submitted in anticipation of this Symposium echo the same thing, and I'll quote those materials. "The committee has always avoided setting forth lists of relevant factors in the text of evidence rules on the ground that a rule is not a treatise and any list is bound to be underinclusive."

The Department agrees with the wisdom of this sentiment. Legally mandating a narrow and exclusive set of factors in Rule 702 for the validation of forensic methods would wrongly deny the consideration of a rich body of studies with the diversity of experimental design, depth and breadth of focus, testing criteria, and derived data, in favor of a monolithic, one-size-fits-all approach that is in no way tailored to, as *Kumho Tire*⁹¹ said, the specific task at hand.

The PCAST approach, however, is not consistent with the scientific mainstream's holistic approach to method validation that is embodied in the currently governing international standard. The Department strongly believes that pragmatic flexibility, which is a hallmark of both the Federal Rules of Evidence and the mainstream of scientific thought, must be maintained and that the existing rule should not, therefore, be amended.

PROFESSOR CAPRA: Thank you, Ted. [Applause] Dr. Lander.

DR. LANDER: Well, I want to very much thank Mr. Hunt for his scientific analysis of this. It's great. I happen to have before me the PCAST report⁹² and the AAAS report.⁹³ Let's actually address this.

PCAST did not suggest that there's a nonseverable six-part test that should be embodied in Rule 702. PCAST said we better have empirical evidence. That's what Rule 702 ought to stand for. The Department of Justice resists the idea that we must have empirical evidence and has said so quite explicitly to PCAST—

88. See *ISO/IEC 17025*, *supra* note 15.

89. See *id.*

90. See PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 52–53.

91. See generally *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999).

92. See generally PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

93. See generally THOMPSON ET AL., *supra* note 87.

MR. HUNT: We don't reject that.

DR. LANDER: Please, I was there meeting with the DOJ staff when they explained very clearly that we would be in a lot of trouble if this went into effect now. Could you grant us a grace period before this goes into effect? This will affect ongoing prosecutions and past convictions. We concluded it was not within our roles as scientists to grant grace periods. We would say what was needed, but we went through various methods and the Department asked what the problems were.

With regard to this six-part, nonseverable test, it says you've got to have big enough samples to draw a conclusion.⁹⁴ The examiners can't know the right answer in advance. You've got to decide what the structure of the test is before you do it. You can't decide *ex post* what you want to make a significant conclusion. It should be overseen by people that don't have a stake in the outcome. The results should be available to other scientists. And you should have more than one study, called reproducibility.⁹⁵

I believe those six parts are called science. There is nothing particularly unusual about calling for those things. If the Department finds those to be strange or out of whack, it would be good to sit down and ask which of those things you do not like.

With regard to the AAAS disagreeing with the PCAST report, the statement of convergent—

MR. HUNT: The approach, the approach.

DR. LANDER: —the statement of convergent validity appears once in the main text of the report.⁹⁶ It appears in section 5 of the report, which is not about foundational validity. It's asking whether there is an adequate scientific foundation for understanding all the factors that might affect accuracy. It goes through a bunch of them. It recites what PCAST did and then it says, "We have reviewed the same literature, and our conclusions largely align with those of the PCAST report, although we do not attempt to distinguish foundational validity and validity as applied, and as a result, we consider all studies that examine the accuracy of latent fingerprint examiners."⁹⁷

There are ninety-two references to the PCAST report. The differences and disagreements are tiny. And, in particular here, since they are not attempting to distinguish between foundational validity and all the other questions, I agree with them that looking at the whole literature holistically is great.

But as to whether or not you have a method at all and whether it is reliable, there is science. It is used in every discipline. And trying to throw up smoke that somehow there's something other than "we run a test and see if it works" is unhelpful. I am sympathetic to the Department that you have situations where you are in the middle of cases where there is not currently the scientific

94. See PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2, at 52–53.

95. See *id.*

96. THOMPSON ET AL., *supra* note 87, at 44.

97. *Id.*

underpinnings, and I salute the FBI for getting those underpinnings, but let's not confuse the issue as to whether or not science has a meaning in this case.

PROFESSOR CAPRA: What I take from the Department's presentation is that it is not per se an attack on the PCAST report but about how you look at reliable forensics. Do you work from a list or do you look at it more holistically? That would be one issue. The other issue being, do you need a new rule at all? And I understand the point you made, Ted, and the point that Steve made.

One argument is that you have a rule now, and the PCAST report says you can cobble together in the existing rule on the Committee note from 2000 a perfect means of regulating feature-comparison methods. And yet, demonstrably, the rule is not being used to meaningfully regulate this evidence.

DR. LANDER: I'll note PCAST did not recommend changing Rule 702.

PROFESSOR CAPRA: Right.

DR. LANDER: We recommended changing the advisory note.

PROFESSOR CAPRA: Right, but—

DR. LANDER: We were told that's not possible.

PROFESSOR CAPRA: That's right.

DR. LANDER: And so we're discussing amending Rule 702.

PROFESSOR CAPRA: Right.

DR. LANDER: Our belief was this is what Rule 702 is meant to mean and there must be a means to make the courts do it. If an advisory note is a possibility, I'd favor it. If it's not, change a comma in the rule and then write a new advisory note. Change one word, any word and write an advisory note. Anyway, thank you.

PROFESSOR CAPRA: Does anybody on the Committee want to speak for a minute or two? Yes, Judge.

JUDGE LIVINGSTON: Well, Dr. Lander, you're correct that PCAST did not recommend the enactment of the rule, and we've heard some discussion of that and you're today suggesting maybe we change a comma in the Advisory Committee. Is that based on the comments today that this rule is meant to apply to many different areas? Why wouldn't a specialized forensic rule be a good idea?

DR. LANDER: I have no objection to a specialized forensic rule if—

JUDGE LIVINGSTON: For forensic feature comparison.

DR. LANDER: For forensic feature comparison. I believe that different classes of evidence, by now, we understand, have certain clear meanings about reliability. If this Committee were to adopt the practice that over time, as there are areas that are substantial enough to require some greater instruction, the Committee could write specific rules as to those areas. I think you would bite them off sequentially as case law evolves and it becomes clear that there's a need for greater clarity because I personally don't want to mess up drug cases and other things.

So I see two ways to do it. One is expanding the advisory note about it. The other is to write a rule for that area. I am not in a position to say which

is better. What I am in a position to say is it is desperately needed. We know there is an area right now that fails to achieve Rule 702's purpose of requiring true reliability.

We know as a scientific community what reliability means. It means empirical evidence with a measurement of accuracy in this case. And one way or the other, the Committee should accomplish that because we know it is leading to real problems. If you do it, it will not keep evidence out in the long run; it will create a period of a little bit of turmoil, I confess, for the Department of Justice, but it will then cause the forensic science community to distinguish between reliable methods and nonreliable methods.

PROFESSOR CAPRA: Judge Rakoff?

JUDGE RAKOFF: I'm wondering whether the Committee has considered one rule for scientific experts and another rule for all other experts, because I think that's the real dividing line, so I don't think it has to be limited to forensic scientists.

But science has a methodology, and as *Daubert* itself recognizes, it includes aspects that other kinds of expertise do not. And, yes, there will be close cases where it'll be unclear perhaps whether it's science or not. In science, that's a problem with any rule that tries to make distinctions, but the basic fundamental difference is that scientific evidence has a methodology established, frankly, by centuries, ever since Francis Bacon. And nonscientific expertise is a different animal. And so I just wonder whether the Committee might want to consider that possibility, two rules, one for scientific expert and one for not.

PROFESSOR CAPRA: Thanks, thanks. So did you want to respond, Ted?

MR. HUNT: Yeah, I just—I guess I want to reiterate something that Dr. Isenberg said earlier. There were over 2000 studies submitted to PCAST. There were three that they said were valid or that met those criteria. So that's an awful lot of data to just throw away and say it means nothing.

DR. LANDER: Three?

MR. HUNT: Two fingerprint studies and the firearms study, the three black-box studies.

DR. LANDER: No, no. The report considers many, many studies. Many, many.

MR. HUNT: Well, those are the three that spoke to foundational validity as you defined it.

DR. LANDER: Ah, that spoke to foundational validity.

MR. HUNT: Those are the three that you claimed spoke to foundational validity.

DR. LANDER: So we asked the Department of Justice—the Department of Justice says the report omitted many relevant scientific studies.

MR. HUNT: Yes.

DR. LANDER: We submitted a request to the Department of Justice in late September 2016 asking the Department to specify any studies that had been relevantly omitted.

MR. HUNT: Your request said appropriately designed.

DR. LANDER: In December, the Department replied that there were not.

JUDGE LIVINGSTON: Finish this remark, but we've got to move on, guys. Hearing from all of you is going to be most helpful to us.

PROFESSOR CAPRA: We're at our practitioner side of the panel now, and that's the final part of the forensics inquiry, and then we move over to the broader questions of *Daubert*. And let's go to Andrew Goldsmith, from the Justice Department.

MR. GOLDSMITH: Well, thanks, Dan. Simply put, this is a situation where I believe that the solution is to ensure that practitioners and the judiciary understand and apply the existing rules and not one that calls for changing the rules. As you all know, the Department has made it a top priority to reduce violent crime and increase public safety. And reliable forensics is required by everyone associated with the criminal justice system, both to exonerate the innocent and to convict the guilty.

The Department agrees that only reliable forensic evidence should be admitted into evidence, and we're working to strengthen the science and the standards that our examiners follow. Forensic science is an applied science, like medicine, computer science, engineering, or psychology. Ideally, expert testimony, usually offered in the form of an opinion, helps the jury place evidence in the proper context within a case.

And one of the basic premises in my comments is that there's no principled reason to carve out a special rule for forensic testimony. We don't require doctors to report an error rate when they give their opinion as to the cause of a plaintiff's lung cancer. We don't require environmental scientists to report an error rate when they offer an opinion as to what particular chlorinated aromatic solvent polluted a river. And we don't require engineers to supply an error rate when they testify about the cause of a bridge collapse. And creating a special rule for criminal forensic testimony strongly suggests that this is an effort to exclude relevant, probative, and admissible evidence. Rule 702 incorporates *Daubert* and other Supreme Court precedent and sets an appropriate flexible standard.⁹⁸

We agree, as I said before, that only reliable scientific evidence should be admissible and believe *Daubert*'s nonexclusive checklist for determining the validity of proposed testimony is the proper standard. Case law makes clear that the gatekeeping function should be flexible.

For example, in *Kumho Tire*, the Supreme Court stated, and I'm quoting, "*Daubert*'s list of specific factors neither necessarily nor exclusively applies to all experts or in every case. Rather, the law grants a district court the same broad latitude when it decides how to determine reliability as it enjoys in respect to its ultimate reliability determination."⁹⁹

Accordingly, a rigid test for admissibility relying on one element of *Daubert* is inconsistent with Supreme Court case law, the Federal Rules of Evidence, the intent of the Judicial Conference as embodied in the text and the Committee notes, as well as the intent of Congress, which reviewed and

98. FED. R. EVID. 702.

99. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141–42 (1999).

did not object to Rule 702 and its accompanying Committee notes. We believe in the adversarial system where both sides are adequately prepared, have received the discovery to which they're entitled, call their own experts, and cross-examine their adversary's witnesses, that this is the best way to determine the truth.

We, as prosecutors, are required to provide complete discovery to defense counsel. Defense counsel, in turn, must review the material that's provided, and both sides are required to be prepared to examine expert witnesses. A key part of my job in the Department has been to see that federal prosecutors go beyond what the law requires when it comes to discovery, and this is particularly the case in the last year with respect to forensics-related discovery.

As many people may know, in January 2017, then-Deputy Attorney General Sally Yates issued supplemental guidance for prosecutors regarding criminal discovery involving forensic evidence and experts.¹⁰⁰ This idea was based on considerations by the National Commission on Forensic Science, and the DOJ, myself primarily, worked with people like Judge Rakoff and others to develop this policy going forward.

It specifically describes four key steps that prosecutors need to take to meet their disclosure obligations for forensic evidence to meet their *Brady*,¹⁰¹ *Giglio*,¹⁰² *Jencks*,¹⁰³ and Federal Rule of Criminal Procedure Rule 16¹⁰⁴ requirements. And this year, as part of mandatory criminal discovery training for all 6000 federal prosecutors, both online distance education as well as in-person training, I am covering the supplemental guidance on forensics, and in this manner, all federal prosecutors will receive training on the guidance.

I would note that in 2014, building on a point that Dr. Dror made concerning cognitive bias, we trained prosecutors on cognitive bias. We recognize what the issues are. We make sure that prosecutors address them in a timely fashion. And based on the information provided in discovery for forensics, defense counsel can decide whether to seek *Daubert* hearings; they can consider putting on their own expert witnesses. I echo Professor Saltzburg's comments that where requested, defense counsel should be liberally granted expert witnesses so they can put forth an appropriate defense and can better prepare the defense in general.

For this reason alone, any contemplated rule changes, whether to the Federal Rules of Evidence or the Federal Rules of Criminal Procedure, we submit, are premature. Based in part on the materials they receive in discovery, through cross-examination, defense counsel influence how much weight the jury will give to the testimony. Most of us are familiar with the

100. Memorandum from Sally Q. Yates, *supra* note 18.

101. *See generally* *Brady v. Maryland*, 373 U.S. 83 (1963).

102. *See generally* *Giglio v. United States*, 405 U.S. 150 (1972).

103. *See generally* *Jencks v. United States*, 353 U.S. 657 (1957).

104. FED. R. CRIM. P. 16.

following Wigmore quote, “Cross-examination is beyond any doubt the greatest legal engine ever invented for the discovery of the truth.”¹⁰⁵

Our adversarial system is based on the principle that the truth emerges when opposing parties have the opportunity to cross-examine each other, parties, expert witnesses, and other witnesses. Each party is able to call its own witnesses and introduce conflicting evidence. Our legal system does not ask juries to find evidence indisputable; it asks juries to evaluate evidence which is often completely disputed.

PCAST and the changes predicated on PCAST’s suggestions ignore the basic nature of the criminal justice system. It ignores, as Justice Harry Blackmun wrote in *Daubert*, both the capabilities of the jury and of the adversary system generally. Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.

Through cross-examination, defense counsel can assist the jury in determining the weight to give the evidence, as opposed to barring it with respect to admissibility. What this comes down to is a fundamental philosophical difference. I and many other people in this room believe that the system isn’t broken, and while we should be striving to improve it, that doesn’t mean dramatic changes.

Other people, including some of my fellow panelists today, are calling for radical changes in the way forensic evidence is handled in court and urging that the Federal Rules of Evidence be revised to facilitate such changes. Interestingly enough, when I was on the opening panel with Judge Rakoff in the D.C. Circuit panel this year concerning forensics, the two people on the panel who said “let the juries do their job, let’s rely on cross-examination,” were the two practitioners, myself, which may come as no surprise, but also the veteran criminal defense attorney on the panel. This is clearly not a prosecution, antidefendant prospective.

What I’d like to leave you with is the following: The Department is committed to ensuring that our forensic experts testify only to statements that are supported by the available research, data, and science. You’ve heard about the ultras. You understand that we will be developing a system continually monitoring the accuracy of courtroom testimony.

The way to improve this area is not by revising Rule 702. The FJC is a tremendous resource. I work with them all the time. I urge this Committee to employ their assistance in redoubling efforts to train judges on Rule 702 and *Daubert*. And in 2000, when this Committee considered whether Rule 702 should require trial courts to follow specific procedural requirements in exercising their gatekeeping function over expert testimony, they reached a similar conclusion. Specifically, the Advisory Committee notes cited a 1998 law review article written by Professor Capra entitled “The *Daubert* Puzzle,” which stated, “Trial courts should be allowed substantial discretion in dealing with *Daubert* questions. Any attempt to codify procedures will likely give

105. JOHN HENRY WIGMORE, *WIGMORE ON EVIDENCE: EVIDENCE IN TRIALS AT COMMON LAW* § 1367 (4th ed. 2018 Supp.).

rise to unnecessary changes in practice and create difficult questions for appellate review.”¹⁰⁶

I respectfully request, as Judge Moore discussed, that we continue to allow trial judges similar discretion here. Thank you. [Applause]

PROFESSOR CAPRA: So I’ll say a couple quick things about that. The citation that Andrew refers to in the 2000 Committee note was to a different part of the article, not the portion just quoted by Andrew. And the second thing I’d say is, the problem that we’ve had since *Daubert* was written is that it’s a schizophrenic opinion, as everybody on the panel knows.

If you look at one part, it talks about how we’ve got to be really controlling on scientific experts because those opinions could be unreliable. The judge must act as a gatekeeper and the proponent must show reliability by a preponderance of the evidence. And then, in the second part, cross-examination is the greatest instrument for truth ever devised and everything should go to the jury.

If they were really going to be embracing allowing the jury to decide based on cross-examination, the Court would have employed a Rule 104(b) standard—the proponent just has to provide a prima facie case of reliability.

Our next speaker is Chris Fabricant from the Innocence Project.

MR. FABRICANT: First, I want to thank Professor Brandon Garrett and our research assistants on this project. We actually developed a data set for this particular symposium piece, and the reason that we did that is because the evidence at issue today, feature-comparison evidence, is overwhelmingly used in state criminal-court prosecutions. And while I imagine that in federal court you have a higher quality of resistance to unvalidated and unreliable techniques, the truth is that—and overwhelmingly in our wrongful conviction cases—there have been opposing experts and there have been vigorous cross-examinations of people, of experts who were just plain wrong and were unable to expose that truth through cross-examination. So I have to respectfully part ways on that particular point.

Forty-seven percent of the known 351 wrongful convictions that have been established through DNA involve the misapplication of forensic sciences. Of those, every single one of them was prosecuted in state court. Thirty-eight states have adopted the federal rules. So what we’re really talking about here today is how these rules are playing out in state court. And so what Professor Garrett and I did was examine how these rules are playing out.

In the twenty-two states that have adopted the 2000 version of the Rule, we looked for any reference to reliability language or citations to Rule 702(b) and (c),¹⁰⁷ what we called the reliability prongs. Since 2000, there were only 857 references, and of those, only about a third, 328, were involved in criminal cases. So, in all the criminal prosecutions in these states, there are only 328 cases that actually discussed the rule at all. Rulings in favor of the defense, three. Out of every case that we could identify in all states that have adopted this rule, three cases.

106. Daniel J. Capra, *The Daubert Puzzle*, 32 GA. L. REV. 699, 766 (1998).

107. FED. R. EVID. 702.

Exclusions, very, very rare as well. And in those cases, almost always by a defense having proffered a particular expert. And you get much more robust conversation about the rules in those exclusions of defense experts. I want to be clear about the limitations of our data—that this is a work in progress. I, as a practitioner, don't have the time that Professor Garrett has. It's limited to written opinions, to appellate opinions, so we don't have data about summary orders. We don't have data about something that may have been excluded prior to trial or how robust those particular opinions may have been.

So the data is what it is, but I think it presents a pretty good snapshot. And it also is very consistent with what we know about federal court practice and what I can tell you from twenty years of experience anecdotally, having to witness this same phenomena that was going to be discussed today over and over and over again.

So these are some of the summary findings. We find that courts in general are ignoring the rule. I think the paucity of the rulings themselves speaks to that, and I think that when we actually look at the cases that purport to apply this rule, it's essentially ignored anyway. Lip service is paid to citing the rule, but it isn't actually applied.

And when we actually look at the reasoning behind these decisions, what we see overwhelmingly is what an important role precedent plays in the admissibility of scientific evidence. And the problem, of course, and the scientists in the room will tell you about this, is that science is a process, and that we shouldn't be relying on a 200-year-old precedent because we've been admitting hair microscopy for 100 years to say that it's valid today. Not only do we know that that's not true, we also know that there have been seventy-five known wrongful convictions associated with hair microscopy.

So, when I hear talk about this is just a limitation of the science when they made an association between a hair, an exemplar hair and a crime scene hair and that it's not actually an error when somebody did thirty years on death row because somebody made that association—to me, that's where science and justice have to diverge and that we have to defer to science and we have to have accurate error rates.

What we were also seeing, and this is true again and again and again, is that you have all of these witnesses who are experts because they have very fancy credentials, very impressive resumes. That's why they're testifying in these cases. And what we see is this overwhelming deference to the expert credentials in lieu of any real reliability tests. I'm going to point to a couple decisions today, but I read them very frequently, and the rationale is that because so and so is qualified, that testimony is going to be helpful to the trier of fact and it rests on reliable principles and methods, period, end of story, it's admissible.

What we need is a very clear distinction between yes, the expert must be qualified, but that's just the first step of the inquiry. What we're also seeing in this data is the flexibility of *Daubert* being misused and misapplied. It's being used in techniques where error rates are ascertainable. They just don't exist, right? Here is a really good example, this *Coronado* case in Texas,

where the reasoning is that because biomarker analysis is based partly on experience and training, the hard science methods of validation, such as assessing the potential error, rate of error, are not always appropriate for its testing.¹⁰⁸ And so it's reliable. That's nonsense, right? This is pattern matching.

That kind of thinking is just shifting it over to calling it a soft science to jettison this requirement of error rates or to have any basis to make an accurate assessment of the probative value of the conclusions that are being offered. It's an effort to avoid the gatekeeping function, a decision like that, but it has a very pernicious effect because precedent is so valuable to courts when they want to avoid having a *Frye* or a *Daubert* hearing.

Another example is from the Delaware Supreme Court, presumably a learned court. This is a direct quote where it's saying, "It would have been more persuasive if the firearms expert could recall how he arrived at his conclusions that the markings were similar, but we require that the principles and methods, not the conclusions, be scientifically valid."¹⁰⁹

And so, if you read this opinion, what it'll show is that, well, the expert testified to just how reliable toolmark matching is, right? But he forgot all the data—well, he said he applied it reliably, and that's all we really require. So we see that as a substitute for reliable principles and methods.

I'm going to go through a quick, unreported decision from Florida, where there was a *Daubert* hearing held in a bite mark analysis case. [Referring to a PowerPoint presentation] We see these types of decisions a lot.

And the finding on sufficient facts and data are based on the—essentially, the collection of this data and that there it was actually collected, that's it. And that the facts and data exist at all is enough to satisfy that requirement. The product of reliable principles and methods, right, is the same thing. It's collecting the data, right? It isn't anything about what would make those reliable or valid principles for collecting this data.

And it conclusively states that whether a specific person could be the biter, even though we have no ability to understand how often they would get it right or wrong. And now, in error rates, this is an astonishing statement to me. This case cites to the Texas case,¹¹⁰ saying that we don't really need error rates because this is a feature-comparison evidence or it's a soft science and at the same time notes on a footnote that there is a study that establishes a 63 percent error rate. It just ignored it completely as if this was just a fact that you might find interesting.

There are twenty-nine, and as of yesterday, thirty known wrongful convictions and indictments based on bite-mark analysis, all since the year 2000. Twice I've litigated *Frye* and *Daubert* challenges to the admissibility of bite-mark evidence. Both times there's been at least two wrongful convictions overturned that were based on bite-mark evidence and still lost the admissibility challenge.

108. See generally *Coronado v. State*, 384 S.W.3d 919 (Tex. Crim. App. 2012).

109. *McNally v. State*, 980 A.2d 364, 370 (Del. 2009).

110. See generally *Coronado*, 384 S.W.3d 919.

Dr. Kafadar testified in one of those, and the argument was made that Dr. Kafadar was not a member of the relevant scientific community because she doesn't go down to the medical examiner's office and photograph the bite marks herself, so we have to ignore Dr. Kafadar. The scientific community in other jurisdictions are the bite-mark experts, right?

So you get the circular reasoning in that you ask the people who would generally accept the technique whether it's generally accepted and you're going to get the answer you can expect. What I would suggest to the Committee is that modification is needed to send a message to the courts that the rule is not working as intended. I think almost everybody in the room agrees with that, save a couple.

The more prescriptive language is necessary because it's being ignored and misapplied. I believe, based on our data and our personal experience, the rule precedent needs to be explicitly established in scientific evidence. I agree with Judge Rakoff entirely in that the tests applied need to be appropriate for the field of expertise.

So we have scientific evidence, and there is a scientific method for establishing reliability in those areas. There are things like crime scene reconstruction—behavioral sciences that have gold standards to talk about how we know those fields are reliable and admissible. That's different when you're talking about scientific evidence. So it has to be valid to the field of expertise and to the task at hand. For example, take forensic odontology. They do victim identification on one hand, right, and they have thirty-two teeth; they have X-rays. We know if somebody's missing, we can make that identification. That's the field of expertise. But if the task at hand is bite-mark analysis, the method is not validly applied to that task. And they should say that. And they should distinguish those different applications because it depends on the task at hand.

Another concern is data collection versus data interpretation. What you see over and over again are courts that are very impressed with the way that experts gather data, in other words, lift the latent prints, take photographs of bite marks, carefully take moldings of tire treads, and all the rest.

That is important, meaningful, but it's useless if you don't have a database, a population-frequency database or a black-box study of some sort to test how often you make a correct association or how rare or how common those features are.

My last point is that there must be appropriate definitions of the relevant scientific community. If Dr. Kafadar can't testify in a forensic science case, then the courts are lost. Thank you.

PROFESSOR CAPRA: Thank you, Chris. [Applause]

MR. FABRICANT: I'm sorry. This happened today. Yesterday, Sherwood Brown, who spent thirty years on death row on the basis of shoe-print testimony, bite-mark evidence, and hair microscopy. He was exonerated. This week, George Perrot, in a hair microscopy case, the only one of the FBI hair-comparison cases that actually got overturned, Dr. Kafadar was admitted as an expert in that case and did provide very important

information. And Dr. Dror will be pleased to know that only last night did Dr. Kafadar learn of any of the facts of the case at all. [Laughter]

PROFESSOR CAPRA: So she was really blind. [Applause] Fantastic. Thank you, Chris. And our last speaker on the forensic panel today is Anne Goldbach.

MS. GOLDBACH: Thank you for inviting me. I am the forensics director for the state public defender agency here in Massachusetts, and I chose to look at Rule 702(d),¹¹¹ as it was dealt with in the First Circuit, the federal district court here in Boston, as well as our state courts. And I was really surprised to find how very few criminal cases there were since the amendments in 2000. Most of those cases involve feature-comparison evidence.

Of those cases that did directly or indirectly address the application of a methodology, most of them looked at the standards and methodology and whether they permitted a conclusion of a match or an individualization. A couple of them actually looked at whether or not the methodology had been properly used to reach an ultimate conclusion.

Because I only practice in state court, I wanted to look at the history of the amendment to Rule 702 back in 2000, and I saw that Bernstein and Lasker had written “Defending *Daubert*: It’s Time to Amend Federal Rule of Evidence 702” and indicated that this issue of the application of the methodology was one of the key areas in which they found the courts had gone astray.¹¹² And they said, in fact, this issue of the application of the methodology was one of the guiding principles back when the Advisory Committee was considering the amendment of the Rules in 1997.

Bernstein and Lasker’s review of the decisions, although mostly civil, found that a number of the courts held that the application of the methodology was beyond a judge’s role as a gatekeeper. Instead, these courts ruled that it was for the jury and not the judge to decide whether the application had been done properly.

Turning to the federal court, the First Circuit, there was a decision just two weeks after the Supreme Court approved the amendments in 2000, *United States v. Shea*.¹¹³ It was a DNA case where the defendant argued that the government expert had failed to take into account one faint allele dot that would have skewed his conclusion of a match.

The court decided that this assertion was not a problem, that the methodology was fine, the application was fine, and it was simply a matter of the weight of the evidence for the jury to assess.¹¹⁴ Two years later, in a handwriting analysis case,¹¹⁵ the defense asserted that while it was alright for an expert to talk about the similarities and differences between the letters

111. FED. R. EVID. 702(d).

112. See generally David E. Bernstein & Eric G. Lasker, *Defending Daubert: It’s Time to Amend Federal Rule of Evidence 702*, 57 WM. & MARY L. REV. 1 (2015).

113. 211 F.3d 658 (1st Cir. 2000).

114. See *id.* at 667–68.

115. See generally *United States v. Mooney*, 315 F.3d 54 (1st Cir. 2002).

that were examined and an exemplar of a defendant's handwriting, the science, the methodology, and what was known at that point about handwriting did not permit an expert to identify a person as the author of the letters.

The First Circuit disagreed with that and said the examiner's testimony was reliable; it was based on valid technical specialized knowledge.¹¹⁶ He had used the same methodology other examiners had used. He was generally peer reviewed in terms of handwriting. There was an alleged error rate of 6.5 percent.¹¹⁷ He was certified, and he'd done proficiency tests. The defendant had said, "well, wait a minute, there's a lack of standards here about how many points of similarity there needs to be before you can say somebody's the author of a letter, and the studies that are cited, there's problems with those."

The court didn't have any problem with that and said the defense had simply misunderstood the meaning of *Daubert*—it does not require unassailable expert testimony, that the opinion doesn't have to be correct, it just has to be supportable.¹¹⁸

There were a couple of more interesting cases at the federal district-court level and one of those was Judge Saris's. That was *United States v. Monteiro*,¹¹⁹ a ballistics case where Judge Saris found that the methodology was reliable, the examiner was qualified after a pretrial hearing, but that there was a problem in allowing the expert to testify because the expert had failed to properly document his work and the so-called peer review in the case simply involved having a colleague confirm the work.¹²⁰

And, of course, Judge Saris can correct me if I'm wrong, but I believe that she said that the government did have leave to supplement their work. They could present more documentation if it existed, and if there was more information on peer review, that could be presented.

In another case, Judge Gertner also addressed the question of ballistics testimony and really fretted over the fact that ever since *Daubert* had been handed down, all jurisdictions had admitted ballistics information.¹²¹ That was a case where she noted that the examiner was not certified, had only been shown casings from one weapon, had not relied on FBI reference materials regarding class characteristics, had no record of past experience, no documentation whatsoever regarding the comparisons, and no data on error rates.¹²²

And I agree with her warning: "The more courts admit this type of evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure."¹²³ Here, in

116. *See id.* at 61–63.

117. *Id.* at 62.

118. *See id.* at 61–63.

119. 407 F. Supp. 2d 351 (D. Mass. 2006).

120. *See id.* at 355.

121. *See generally* *United States v. Green*, 405 F. Supp. 2d 104 (D. Mass. 2005).

122. *See id.* at 109–117.

123. *Id.* at 109.

Massachusetts, we don't have an adopted rule of evidence, only a proposed rule of evidence. And, unfortunately, the inquiry ends quite often if it has been demonstrated that the forensic discipline has been generally accepted in the relevant community. So the inquiry under what would be Federal Rule 702(b) and (c) often stops at that point and the evidence is admitted.

I will say that both in Judge Saris's case and in the Massachusetts cases, when it comes to ballistic evidence as well as other types of evidence, it is still okay to say under the case law that the expert has a reasonable degree of scientific certainty. And so I urge the Committee members to try very hard to fix this very difficult yet problematic aspect of testimony. Thank you.

PROFESSOR CAPRA: Thank you. [Applause] That concludes the forensic panel, but I wanted to give Dr. Lander the floor for a closing remark.

DR. LANDER: Thanks so much. I don't want to say anything substantive, but I do want to give my extraordinary thanks. The President tasked us about two and a half years ago with considering this area and making a set of recommendations to improve the reliability of forensic evidence in the courts. That was our task. We worked hard and made a set of recommendations to various parts of the federal government. One of them was to the Judicial Conference of the United States, to this Committee.

I am just delighted that you have taken up the charge. We concluded that Rule 702 is broken in some way and something had to be done. And I am enormously grateful that you have attended to the problem. How you're going to solve the problem, I don't know. You've heard many different considerations about how you might approach it, but I mostly want to say thank you for engaging with the hard work.

I also want to say thank you to the FBI, who I have enormous respect for and with whom I think I have very small disagreements, not major ones. And I want to also thank my colleagues/adversaries in the Department of Justice, whose engagement, passion, and vigor I appreciate. And I understand their challenges. And I'm grateful to them for their engagement. And so mostly I am just thrilled having spent a lot of work with twenty-one other members of this council on this report that it has gotten some attention from you, and I wish you good luck in finding the solution.

PROFESSOR CAPRA: I wish the Committee good luck on finding the solution as well. [Laughter] Thank you, Dr. Lander, I really appreciate those sentiments.

PANEL TWO: RULE 702 AND *DAUBERT*

PROFESSOR CAPRA: There is an interesting segue from the last panel to this one. Anne Goldbach was talking about the Bernstein article.¹²⁴ The Bernstein article came up with some cases which indicated that some courts are ignoring two requirements of Rule 702—the requirements of sufficient facts and data and of reliable application. Those two requirements are not found in *Daubert*, which is concerned solely with the reliability of the

124. See generally Bernstein & Lasker, *supra* note 112.

methodology. Those are factors that were added by the Committee to *Daubert*, essentially, in the 2000 amendment to Rule 702. And some courts are ignoring them because they are relying only on *Daubert*.

I prepared a report for the Advisory Committee on the Bernstein article, which the Committee considered under Judge Sessions's time as Chair. Judge Sessions made the suggestion that broader questions about Rule 702 might be included in the forensics conference that was already being planned. And so here we are, our patiently waiting panel, thank you so much for waiting. We're talking more about *Daubert* issues now in general but somewhat outside the forensics area. It's an open forum. Judge Saris?

JUDGE SARIS: Okay. First of all, Professor Capra can manage my trials anytime. [Laughter] He's keeping us on such a good track. But I want to thank the Committee for inviting me to speak about the role of *Daubert* motions in federal criminal and civil litigation. I've got several basic points that I want to make, and then I'll circle back.

The first is that *Daubert* motions play a significant and beneficial role in complex civil litigation. But they play practically no role in federal criminal trials. So there's a very big difference between the way it's treated on the civil side of the ledger and the criminal side of the ledger.

Second, while *Daubert* proceedings have improved the quality of expert testimony in court, in my view, they also, we have to recognize, slow down the litigation and add considerable expense to federal litigation; in my view, it is worth it. But we have to acknowledge the fact of expense and delay.

Third, more flexible and effective case management in civil and criminal cases is necessary to reach *Daubert* motions earlier in the litigation so that you can think about it more slowly. It is complicated. Some research has said that it takes two to three months to get through a *Daubert* hearing. We have to write it up, so sometimes it gets published. It has to be available for appellate review. You have to think about it. It's hard.

And finally, we need innovative techniques to ensure that not just judges but, someone pointed out here, juries understand the science. I didn't understand a lot of those formulas that were being flipped up here on PowerPoints earlier. It takes a while to understand it.

I was lucky enough to be asked to serve on the Advisory Committee for PCAST. I was one of the judges. I don't actually remember anyone calling out for a new rule. What they were calling out for is for judges to use the rule. And I found that that report¹²⁵ was extremely helpful at least in flagging some of the big issues that happened in the forensic area.

Just to circle back, I've been a judge now on the trial court for thirty years, it's hard to believe, at both the state and federal level. Actually, I started before the issuance of *Daubert*. And on the state courts, when I was on the superior court, all expert evidence came in basically without challenge. I rarely received any challenges at all to expert testimony, either of the civil or the criminal side. I am one of those judges who have, and I cringe now, let

125. See generally PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 2.

in bite-mark identification, ballistic identification, issues on other areas like shaken baby syndrome, but I have to say it was never challenged. The bar didn't challenge it. It was never presented as an issue before the court.

So this has all changed dramatically in the last twenty years. There is virtually now no substantial civil case where there is not a challenge, and not just scientific. Economic experts come up a lot. In almost all damages cases, there's a challenge to an economics expert.

They necessitate extensive briefing. And often I like evidentiary hearings so someone can explain it to me. They arise often in motions in limine, which drive me nuts. Just on the verge of trial, someone will raise it. Oh, you think you've gotten through the motion for summary judgment, then bang, you're about to hit trial and on the eve of trial, you have a choice between extending the trial date or holding the kind of hearing that's appropriate.

And initially, as we all know, *Daubert* was intended to liberalize the admissibility of scientific evidence. It was an opinion written by Justice Blackmun in a plaintiff's case involving a decision by the California Supreme Court that birth defects had been caused by a mother's ingestion of a certain drug.¹²⁶ The evidence had been excluded and the Supreme Court said no, there's a liberal thrust to the rules and set forth all the factors. This was not viewed as a rule of exclusion. And I think that's important to realize when you're trying to—when you cast a new rule. You can't deem it that way.

It altered the nature in state courts as well, and the question was how are judges supposed to do this. And I think that we have to be thankful for this kind of proceeding today. Someone mentioned the FJC, and what's maybe one way of actually trying to address some of these factors that came up in PCAST is the reference manual on scientific evidence as a go-to treatise that is extremely helpful and maybe needs to be updated with some of the information that came in here.

Tom Sobol was the lawyer in the case, so I'll just mention how complicated this is. I had a case involving the drug Neurontin and whether it generally—whether it caused suicides. Well, the state had a bunch of cases; I had cases in multidistrict litigation. It involved weeks of hearing a huge amount of experts, and I ended up writing a 159-page opinion.¹²⁷ This is not the kind of attention that usually happens in criminal cases, I think partly because of a lack of resources.

We have learned as judges how to address scientific evidence in the context of patent cases. It's actually way ahead of other kinds of cases, and I'll just go quickly through the kinds of things we've learned to use and we need to use, tutorials from experts. Sometimes you have to hire an independent expert and get the parties to do it. Each side needs to have its own experts. Sometimes we've hot-tubbed experts by having them dispute with one another so that you can see the one-on-one. I've often required

126. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 582 (1993).

127. *See generally In re Neurontin Mktg. & Sales Practices Litig.*, 748 F. Supp. 2d 34 (D. Mass. 2010), *amended by* No. 04-cv-10739-PBS, 2011 WL 3852254 (D. Mass. Aug. 10, 2016).

textbooks to be provided to me so I could learn—like everybody’s agreed that this is the textbook.

And so we should learn from the patent litigation in how we do this on the criminal side as well as in general product liability. I appreciate very much bodies like the National Academy of Sciences, the PCAST, and the reference manual because judges—I know you’ve all been real super critical of judges. But let me tell you, when you have two book-ended experts saying the opposite thing and they both are credentialed, I’m sorry if we sometimes use proxies. The reality is that you have two very highly credentialed people disagreeing.

You know, I majored in political science, so, at some level, you have to empower the judge to be able to understand it. I’m quite certain *Daubert* has not “liberalized” the admission of evidence. The literature says that, but that’s not necessarily the be all and end all. I think it’s an important role for gatekeepers to play in order to make sure it’s good evidence that comes before the jury.

And the last thing I’ll say is I’m not sure whether you need a new rule of evidence or not. I certainly think we need new ways of handling these as a matter of civil case management. They need to be addressed early in both federal criminal proceedings and in civil proceedings. We need to have better discovery beforehand and we need to schedule it in so it’s considered in a thoughtful way.

And I guess at this point, I just want to add one thing. Everyone talks about science and the truth, and that is important. But Justice Breyer said in the introduction of the manual of complex litigation that what we’re trying to do is to get to a jury two opinions that are within the range of scientific reliability. We’re doing the best we can. We have to decide these cases.

PROFESSOR CAPRA: Thank you, Judge. [Applause] I want to turn to Judge Rakoff for a few words on how judges avoid *Daubert* motions.

JUDGE RAKOFF: Yes, Professor Capra just came over to me and said keep it to five minutes.

PROFESSOR CAPRA: That’s what I said.

JUDGE RAKOFF: And I was very grateful because that’s four minutes more than my wife normally gives me. Also, I’m fortunate because 90 percent of what I was going to say was just said by Judge Saris, and I totally 100 percent endorse everything she just said. Her experience has been my experience as well. The *Daubert* decision, in my view, is a great decision. And Rule 702, whether or not it needs to be further perfected, is a great rule.

We’re concerned here ultimately with finding the truth, and it’s not so easy for juries, even armed with Wigmore’s great cross-examination, to determine the truth when we’re dealing with matters that are so esoteric as experts sometimes present. And, therefore, a gatekeeping function for a district court seems to be totally consistent with the role of courts historically, and *Daubert* and Rule 702 give judges that role to, frankly, keep out the junk and let the good stuff in. And I think that’s all to the good.

The problem with *Daubert* and Rule 702 is that these hearings, if they’re done right, take a lot of time. And judges are very busy and they always have

more on their docket than they can handle in an ideal fashion. I had a two-week *Daubert* hearing in a case. It was fascinating. I loved every minute of it, but who can afford two weeks on a single *Daubert* hearing? Two weeks—three weeks ago, I got a new case where there are six *Daubert* challenges. So, if I spent two weeks on each one of them, I wouldn't be able to handle the rest of my cases.

So the biggest problem I think judges face is the amount of time it takes to do this stuff right. And the reason it takes so much time is, first of all, you're dealing with something that's highly technical. Most judges are political science majors or, in my case, an English major. What do we know about this stuff? So we've got to learn it. Then the test, whether *Daubert* or Rule 702, involve multiple elements. It's not just a single issue you've got to decide. It's many issues. And then it's become increasingly popular as a tactic in civil cases.

It hasn't raised itself very much in criminal cases, I think because, again, as Judge Saris said, the defense bar is ignorant. They don't have their own experts in many cases. And they don't have adequate discovery in many cases. But for whatever reason, they have not done as much as one would have expected. On the civil side now, though, you see it all the time. And so it's a big time-consuming matter.

Unfortunately, too many judges have reacted to that by finding ways to evade Rule 702. And so, as we saw earlier, many judges just ignore entire elements of the Rule. The one that I see most often in some opinions I've read are the applicability. They look at the methodology and they look at the data, but they don't look at the application of one to the other. Then there are cases like we saw earlier where the courts will say, well, we always—we treated this in 1892 in the following way and that's our binding decision, as if the law had not changed. Or they find other ways to avoid the issue. The report was too late or whatever. Or the fallback always is, well, cross-examination will work it out.

So what's the solution? I'm not sure the solution is an amendment to the Rule. I do favor a science rule and a nonscience rule, but putting that aside, the real solution I think for judges is better management techniques for dealing with these. The tutorial that was mentioned is one increasingly popular technique. You bring in people to educate you very quickly on the basics of what the parties agree to, and then you know what the real differences are.

Hot-tubbing, I'm dying to do a hot-tubbing in my court where you get the two experts—

PROFESSOR CAPRA: They're clothed. They're clothed, though.

JUDGE RAKOFF: Yes, yes. They are—well, that's what they say. I mean, I haven't tried it yet. You get the two experts—

JUDGE SARIS: Some experts say it's more reliable. [Laughter]

JUDGE RAKOFF: They're—in civil cases, sometimes you can meld the *Daubert* decision with a trial on the merits since you're going to be making your rulings later with specific findings of fact and so forth. So there are a

number of techniques that can be tried, but those, I think, need to be paid greater attention to by the judiciary, but probably not for this Committee.

PROFESSOR CAPRA: Thanks, Judge. We're going to move on to Judge Grimm who's going to talk about discovery in criminal cases and its effect on the *Daubert* rulings.

JUDGE GRIMM: So this panel is to more broadly look at the applicability of the *Daubert* standard as it has been advanced in the changes to Rule 702 from 2000.¹²⁸ And in the last year, the following experts offered testimony in criminal cases in the District of Maryland: psychologists dealing with varied aspects of mental illness or competency or sanity, medical doctors, gang experts, experts regarding coded drug language, sex trafficking and the effect that it has on victims as they are vulnerable to sex trafficking, eyewitness identification, terrorism, ethnography in terms of language and language use, Bitcoin and digital currencies, computer forensics, possession versus distribution quantities of drugs, counterfeit money, currency, explosives, and firearms.

Keep in mind that in 2000, two significant changes occurred to the 700 rules. Number one, a number of these experts who previously had been allowed to testify as lay witnesses were no longer able to do it after the change to Rule 701, which now says that lay opinion testimony has to be rationally based upon perception helpful to the fact finder and cannot involve scientific, technical, or specialized information within the scope of Rule 702;¹²⁹ 702 deals with scientific, technical and specialized.

And while I think that Judge Rakoff and others who have said this have a point, which has some merit to consider about the difference between science and everything else, where do you draw the line between science and technology? When an engineer is testifying based upon scientific principles, is that technology or is that science? And once you start trying to draw that line, it becomes a little bit difficult.

In order for a trial judge to do what is required by Rule 702—sufficient facts and data, methods and principles reliably applied to the facts of the case,¹³⁰ and then to consider the *Daubert* factors of testing, error rate, peer review, general acceptance, whether or not there's a standard testing protocol as well as the other examples in the Advisory Committee note—the judge needs two things: facts and time. You have to have both in order to be able to do your job right.

There are structural challenges to applying *Daubert* and Rule 702 to experts in criminal cases. Number one, the speed with which criminal cases take place. You have the speedy trial rule. You have the requirement that these cases go from indictment and arraignment to final dispositions in a very short period of time.

Now everyone knows that the number of cases that are tried within seventy days from the time of arraignment are very few because there are many things

128. See FED. R. EVID. 702 advisory committee's notes to 2000 amendments.

129. *Id.* r. 701.

130. *Id.* r. 702.

that cause the speedy trial clock to be stopped. But it is still true that there is a pace associated with criminal cases that makes it a challenge to do the types of things that you need to do to make a considered ruling on expert testimony.

Secondly, there are a whole host, at least in the federal court, of issues that potentially interfere with the development of the case in a criminal case that have an impact on how you can do your job on *Daubert*. Most importantly, there is the pressure to plead quickly and the notion that if there is vigorous challenge by the defense before a certain period of time, any willingness of the government to afford that third point for promptly accepting responsibility and pleading before the government has to be put to the burden of its case under section 3E1.1(b) of the Sentencing Guidelines¹³¹ and will be withdrawn.

And I have seen my share of letters saying if you do not plead by such and such a date or if you file any discovery motions, then not only will we not give you your *Jencks* until after the witness has testified at trial but also, we will not be willing to give you the third point on 3E1.1(b).

The next thing is that there are significant deficiencies in the pretrial disclosure of expert information in Rule 16¹³² as compared to the civil rule.¹³³

Now I am very pleased to hear Andrew Goldsmith say that there's guidance going out to the troops to go beyond what the rule requires, and I commend that. In fact, I think that is fantastic. But right now, there is a difference between Criminal Rule 16 and Civil Rule 26 regarding what has to be disclosed. In the criminal context, Rule 16(g) says, "A summary provided by this subparagraph must describe the witness's opinions, bases, and reasons for those opinions with qualifications."¹³⁴ This is always drafted, in my experience, by the prosecutor.

Whereas the civil case, it says, there must be a written report signed by the expert with a "[c]omplete statement of all opinions, the basis and reasons for them, the facts or data presented, the exhibits that will be used to summarize or support them, the witness's qualification, a list of all other cases in which the witness has testified before and the compensation paid."¹³⁵

Now maybe the compensation in other cases is not so significant, but it does give you a source for getting testimony before by that expert so that you can use it for effective cross-examination.

Two other things—the first is the funds available for the use. In order to have the vigorous cross-examination that *Daubert* itself made reference to, as Andrew said in his citation to the *Daubert* opinion, and in order for you to be able to have that kind of clash that will allow the judge to be able to determine admissibility, and thereafter, the jury to be able to determine

131. See U.S. SENTENCING GUIDELINES MANUAL § 3E1.1(b) (U.S. SENTENCING COMM'N 2016).

132. FED. R. CRIM. P. 16.

133. FED. R. CIV. P. 26.

134. FED. R. CRIM. P. 16(g).

135. FED. R. CIV. P. 26(a)(2).

weight when it is admissible, you've got to have enough time and information. And you've got to be able to have the defense have their access to experts to be able to challenge the prosecution.

And while the public defenders may have access to it, if it's a CJA panel counsel, they have to file a motion and get approval by the court. So, when you have all this happening in that compressed time line with the absence of information and the variety of areas you have to apply it, there are very significant challenges in terms of applying *Daubert* the way the rule was intended in criminal cases. [Applause]

PROFESSOR CAPRA: Thank you, Judge. I would just add that the Criminal Rules Committee is taking up a potential amendment that would beef up the disclosure requirements at the suggestion of Judge Rakoff. So wow, we're all here. This is such a small world of rulemaking, but that's what's happening. And so we'll see what happens with that. That's right Judge Rakoff?

JUDGE RAKOFF: Yeah, it's been raised. We don't know where it will go, but yeah, it's definitely an issue.

PROFESSOR CAPRA: Our next speaker is Zach Hafer.

MR. HAFER: Thanks. Good afternoon. My name's Zach Hafer. I'm an Assistant United States Attorney here in Boston. I've been an AUSA for the last ten years. I agree with Judge Saris. It's fairly rare that you engage extensively with *Daubert* in criminal cases. I've had a little bit more experience than others just by the nature of some of the cases.

There's been people who have said I think we all agree that Rule 702 is broken. That has not been my experience, and I'll get to why that hasn't been my experience in just a minute. I want to first just address Judge Grimm's point and what Andrew Goldsmith said earlier, which is that the Department totally is in favor of early, complete, thorough expert disclosure.

I think it's very important to the adversarial system that both sides have access to full, complete information in a timely manner with a reasonable amount of time to challenge it in court. And to that end, the memo by Deputy Attorney General Sally Yates that issued on January 6, 2017, sets forth four specific categories of information that prosecutors should turn over as reasonable as is practical. This includes not just the written summary and not just the expert's CV but also access to any forensic expert's case file and reports and more comprehensive summary.¹³⁶

So that's certainly been our practice in Massachusetts to follow in criminal cases where *Daubert* issues arise, and I think it will allow for, to the extent possible, the type of time that's necessary to resolve complicated issues.

I want to just start quickly by talking about how as an AUSA we typically confront *Daubert* pretrial. And as an AUSA, typically, you're playing defense on *Daubert* pretrial. You're essentially arguing that this expert who we want to call, who's been challenged by the defense, that their method and principles and everything else passes *Daubert* muster.

136. Memorandum from Sally Q. Yates, *supra* note 18.

We typically don't move to exclude defense-proffered experts en masse prior to trial, and I think there's three reasons for that, which dovetail with my own view as to the three primary reasons why the current regime is working. And the reasons typically that we won't move to exclude an expert en masse first is just general concerns that we have as prosecutors that a defendant who is on trial for his liberty wants to call someone in his own or her own defense. They should be allowed to call the person. And, obviously, more selfishly, we worry about an appellate record where an entire swath of evidence has been excluded.

Second, we're generally confident in our own abilities to expose shakier methodologies and leaps in logic and problems like that through cross-examination, and we think that's a better tool to address issues with methodology.

And third, it's my own experience that juries with more relevant probative information make better decisions than juries with less relevant probative information.

So, for those reasons, we don't typically play offense pretrial on *Daubert*. That's not to suggest that we don't think there's a gatekeeping role. There is. There are times the Department will challenge an expert. A recent high-profile example is the *Libby* case.¹³⁷ Scooter Libby was in the Bush White House and—

PROFESSOR CAPRA: The memory expert.

MR. HAFER: The memory expert, right. And the Department, Patrick Fitzgerald and the District of Columbia, challenged the memory expert proffered by Defendant Libby as unreliable and also as testifying to a matter that was not beyond the ken of the average juror. Essentially they wanted an expert to testify that as time passes, you don't remember things as well and, you know, we took the position that, one, there's not enough science on that, and two, jurors can figure that out without an expert.

So I just want to wrap up here the last couple minutes with the three primary reasons I think the current regime is working pretty well. This isn't to minimize the issues that have been raised with wrongful convictions. That's obviously something we care very deeply about, but the first is—and, actually, Anne, you raised this—I think that the gatekeeping function at least in my own practice is working. And that opinion that Anne cited by Judge Saris, I think, is a good example. The deficiencies were identified in the peer-review component of the ballistics expert that was proffered, and the judge said no, that's not coming in; you can try to fix it, but it's not coming in. So the first reason I think it's working is I think judges by and large are performing their gatekeeper function as intended.

The second reason I think it's working is I think the adversary system is working well. I know there are examples, and they're typically, I know, more in state court of overstuffed, overburdened, underfunded defense lawyers. But in my own experience and certainly in our district, we have a very

137. See generally *United States v. Libby*, 429 F. Supp. 2d 1 (D.D.C. 2006).

talented defense bar. I echo what Andrew said earlier and what Judge Grimm said that to the extent CJA-appointed attorneys in criminal cases aren't given access to their own experts, they should be. That's something that is I think better for the administration of justice.

And third and finally is I think juries generally get this stuff right when they're educated by advocates who have done their homework and have access to the resources they need to either present the sound methodology or cross the unsound methodology.

I just want to tell a quick war story from a recent case I did which I think makes that point. It's rare in a criminal case that you have a special-verdict form. This trial I worked on, we did have a special-verdict form. It was an expert-intensive case and so you were able to see step-by-step certain conclusions the jury made. They had heard from sixteen or seventeen experts on topics as esoteric as neuroimaging and forensic neuropsychology, and you were able to see based on the verdict form that the jury found essentially that this particular defendant did have a traumatic brain injury but that that traumatic brain injury did not explain his behavior.

And it was a nuanced, sophisticated finding. The testimony that the jury heard from the neuroradiologist and the forensic neuropsychiatrist was based on a brain scan that had been done of the defendant using a technique called "diffusion tensor imaging." It's a very novel technique. It's not used clinically. It's only used for research purposes. A neuroradiologist had scanned the defendant's brain, identified what he said was a hemosiderin deposit in the right frontal lobe. Then the neuropsychiatrist came on the stand and said that deposit accounted for the defendant's ability not to control his own impulses and other executive-functioning impairments.

And, in my view, that testimony as the proposed working draft of an amendment on forensic is drafted would not be subject to these heightened requirements. And I don't see, as other speakers have said before me, including Professor Saltzburg, a principled basis to distinguish testimony like that linking something in an image to behavior from the forensic type of testimony we've heard today. Thank you very much, Professor Capra.

PROFESSOR CAPRA: Thank you. [Applause] Now we have three lawyers to finish up from the civil side, *Daubert* from the civil side, and I turn first to Lori Lightfoot.

MS. LIGHTFOOT: Thank you. Thank you for being here, and it's been a great day, an interesting experience today. I'm going to pare down my remarks because most of what I was prepared to say, I think you've already heard a lot of it, but I will highlight some of the points that I think are important.

I want to talk to you a little bit from my perspective as a civil practitioner, but, frankly, given my former role as a federal prosecutor, some of that will clearly bleed into the conclusions that I draw about *Daubert* as well.

One is I want to give you context from the perspective of big law, which is what I am now part of. It's become understood that the opposing side's experts are absolutely going to be challenged under *Daubert* and hopefully discredited in such a way that they can never take the stand, or if they do, that

they're so wounded that their views will not be credible. That, unfortunately, is just the reality of the practice of law today. And really, the notion of going after and taking down the other side's expert begins at the start of the litigation.

Frankly, I think our clients would think it's malpractice if we didn't take on this mantra. I'm not just thinking about how I can make sure that in complicated issues, whether it's on damages, whether it's on environmental cases, or the like, an expert that can help explain complicated issues in an effort to get at the truth. I'm also focused on what can I do to make sure that the other side's expert is discredited. And, as you all know, there's a whole industry that's now built up. There's not a day goes by that I don't get an inquiry from some consulting group saying hire my experts. I could literally dine out for probably thirty days straight if I allowed that to happen because of this huge industry that's now been built up around expert witnesses.

And I have to believe that the exercise of finding the expert, analyzing the other side's experts, going through the depositions, building all the supporting materials that underlie and undergird the expert's conclusions in the summary reports, and then plopping all of that into the lap of the judges and the clerks has to be an enormous burden. But if the role of Rule 702 is supposed to be gatekeeping, the concern that I have is that what I'm seeing at least from my experience is that there's a huge breach in the gate, and what do I mean by that?

I feel that increasingly the courts are reluctant to bar in whole or in part any expert testimony that has the appearance of credibility. Hopefully, in civil cases, we're past the point where somebody that has completely no basis to be testifying is actually showing up in federal court. It still happens from time to time, but really, that's, I think, more and more the exception and not the rule.

You're going to get somebody who on paper looks incredibly credentialed, but what I'm seeing is the failure to actually evaluate with rigor the four specific independent criteria under Rule 702 and hold the experts to that. If it's close, what it feels like the courts are doing more and more as the path of least resistance is to let it all in and, you know, you've heard it here today, rigorous cross-examination will point out the defects and juries are smart and they'll figure it out.

I don't think that that's a reality. And, again, that's worked to my benefit in many instances, but as somebody who cares about the administration of the courts, I don't think that that's the right place for courts to be. I do believe that there's got to be a lot more rigor around the analysis of whether or not these experts actually meet the criteria under Rule 702—that their testimony is something that will actually aid the trier of fact, and that it is reliable in that whatever methodology it is, depending on the nature of the expert, that they've applied, and tested, scientific standards to the particular facts of this case in a way that is relevant.¹³⁸

138. See FED. R. EVID. 702.

And, again, as Judge Rakoff and others have said, in order to get to that point of a rigorous inquiry, that takes a huge amount of time, and we all recognize that. But I don't see any shortcut if we're actually going to apply the rule and the rule's going to actually have any meaning. I don't have strong feelings about whether or not the rule should be changed, but I have very strong feelings that the existing rule absolutely must be applied. And I think it was Professor Lander who said let's send out some advisory or some kind of guidance to give trial courts the kind of instruction that they need and would be helpful in helping them analyze what are increasingly complex problems.

And a couple more final points. I believe that Professor Allen got it right when he said that, unfortunately, what we're seeing increasingly is that suspect conclusions are presented to finders of fact as the truth and that they should defer to that. I think Professor Moriarty got it right when she said we ought to encourage judges to think that their gatekeeping function is essential but also complicated and that we ought to slow down and take the time it needs to examine these issues on a case-by-case basis. And I think, again, it was Professor Saltzburg who said send out some kind of advisory to really get at the need for nuanced interpretation and enforcement of Rule 702.

You know, all the detritus that gets collected during the course of discovery in a complex case ought not make its way into the trial, and if judges and courts are not rigorously evaluating evidence on the Rule 702 standards, I can tell you, in my experience, it has a significant effect on the outcome of cases in terms of whether or not a case settles. Some cases are going to settle at a much higher price because the hard work isn't done and the client isn't willing to roll the dice and take the chance at trial on hoping that the jury is able to screen out the nonsense that might exist in opposing counsel's expert witness. Thank you. [Applause]

PROFESSOR CAPRA: Laura Shamp. This one from—I can say the plaintiff's side, can I not?

MS. SHAMP: Absolutely the plaintiff's side.

PROFESSOR CAPRA: Okay. Go for it.

MS. SHAMP: Thank you so much for inviting me to speak today. It's been a fascinating conversation. Not very many people have talked about *Daubert* from the civil side, and that's where I practice. I am a plaintiff's lawyer, and I'm a trial lawyer. I'm in court three or four times a year trying cases. Sometimes in federal court; sometimes in state court.

Before I came today, I spent some time talking to some of my colleagues in the plaintiffs' bar to try and get some of their ideas about the difficulties that we face with respect to *Daubert* and how to deal with it.

So I want to do two things in my time today. I want to shine a little light on the practical reality that plaintiffs face when we're facing *Daubert* challenges, and Ms. Lightfoot just mentioned some of those. I have a 180-degree different view of them, and I think Judge Saris always said it is an enormous burden that is being placed on litigants. And I want to talk about that. And then I want to touch on a few ideas and suggestions, some mine, some my colleagues', that the Advisory Committee might consider with

respect to how we might address some of these concerns with practical applications of Rule 702.

First, I agree that almost every complex case is going to face a *Daubert* challenge, even on opinions that are well respected, that are right down the mainstream, that have no question at all. The defense is going to challenge the plaintiff's expert no matter what. And what the defense said today is what is my experience as well, that their attempt is to simply cripple our experts, even if they are well qualified, even if their opinions are right down the middle. *Daubert* is being used as a technique to try and cripple the plaintiff's case. And I think that it's extremely important because legions of lawyers on behalf of the defendants will marshal case law to discuss scientific methodology.

And I think that the premise of that is problematic because case law is being utilized in order to attack scientific methodology. And then judges are being put in the positions of writing these opinions about scientific issues, and I don't think that *Daubert* was intended to create that kind of jurisprudence, but it is.

Moreover, legal precedents are being wielded to persuade a judge who's essentially making a factual determination. It is a factual determination, and in my opinion the use of precedent in this way is fundamentally a flawed idea because I cannot emphasize enough that the Seventh Amendment provides a constitutional right to a trial by jury. And *Daubert* is being used as a device to take away a plaintiff's right to a trial by jury of the facts.

And I don't think that was what it intended to do. I think that the Advisory Committee ought to see from the civil side that that is what it is being used for. And the reality of the flood of *Daubert* motions is it does not fall equally upon all parties to the litigation. Because the plaintiffs bear the burden of proof at trial, it is borne more heavily by the plaintiffs, and it is the plaintiffs who then suffer when their experts are facing these enormous challenges. And I think it's fair for the people in this room to think about what the costs are of that and how those costs create a barrier to access to the civil justice system. It is extraordinarily important to our citizens to have the right to trial by jury.

And if procedural rules are putting road blocks in front of individuals who are simply trying to have a jury decide their claim, then that right to trial by jury that's guaranteed by the Seventh Amendment is in jeopardy. I don't think it was what *Daubert* was intended to do. I think that we all have to recognize the social importance of juries deciding these issues. We have to look no further than the recent GM switch case.¹³⁹ It was a friend of mine who was a plaintiff's lawyer in Georgia who took on General Motors and saved lives because of that. For individuals who are trying to face up against big law and big money, if we put procedural obstacles in front of them, they will simply go away. If *Daubert* becomes this tool that they're able to use, it will simply eliminate an individual's ability to challenge corporate conduct.

139. See generally *Melton v. Gen. Motors LLC*, No. 14-cv-1815-TWT, 2014 WL 3565682 (N.D. Ga. July 18, 2014).

And I don't think that's what *Daubert* was intended to do. I don't think that's what it should be doing. These frivolous attacks on really qualified experts can cost hundreds of thousands of dollars.

I've heard the judges today say it is two weeks of the judge's time to be engaged in these *Daubert* hearings. Imagine the cost on individual litigants who are simply trying to bring a case in front of the court. That cost essentially closes the door to courts and meritorious claims.

It's incredibly important that juries be allowed to make these factual determinations, and a lot of people have already mentioned that with the idea of cross-examination and that juries are really good at it from my experience. When you have twelve people sitting in there hearing the evidence and with experts being cross-examined by civil trial attorneys who are skilled, juries most of the time get it right and they're able to do that better than a random judge poll there.

Another implication that ought to be considered is the chilling effect on good scientists who are being challenged by *Daubert* and the way that they are crippled and discredited. And real scientists, who are having their reputations attacked through these Rule 702 hearings, it simply chills their desire to be a plaintiff's expert and additionally increases the cost. If a scientist whose job is to do the good work that we as a society need is being pulled away from his or her job every day to spend two weeks being crippled by big law, I don't think that that's appropriate.

One suggestion that we have is to shift the burden. That is, that the defendants have to put up a prima facie case that the plaintiff's experts are not qualified and actually lay out that prima facie case, and shift the burden then. And that way, you limit the frivolous challenges.

PROFESSOR CAPRA: To be entitled to a *Daubert* hearing, the proponent would have to—

MS. SHAMP: They have to put forth a prima facie that says that the expert is not qualified or the expert's opinion is not reliable.

PROFESSOR CAPRA: But the burden would be on the proponent to convince the judge.

MS. SHAMP: The burden would then shift once the prima facie case has been made, then the burden would shift back.

PROFESSOR CAPRA: Good. Thanks.

MS. SHAMP: Another idea would be fee shifting. I think that if a *Daubert* challenge is made and it's not successful that the cost of doing that should be shifted to the person who made the nonmeritorious motion. I think that's a very good idea that would cause big law to be less likely to want to raise those challenges.

PROFESSOR CAPRA: Wouldn't it cause plaintiffs to be less likely to as well?

MS. SHAMP: Absolutely. Well, I mean, I think that, in actuality, it's not plaintiffs who are doing these challenges, but I do think that it would decrease the burden on the courts. Also important is the opportunity to cure, and I think this comes with what Judge Saris said, and that is have these hearings

early in the case and not allow them to be coupled with motions for summary judgment.

What we see is that the *Daubert* challenge comes coupled with a motion for summary judgment and then it's like a second bite at the motion-for-summary-judgment apple. And if the expert is rejected under *Daubert*, then the motion for summary judgment's granted and a meritorious claim is thrown out of court. And so I think there ought to be an opportunity to cure so that if there is some technical or other problem with the plaintiff's expert, that they have the opportunity to cure that defect. What happens now is that the plaintiffs have to come up with six experts in case one of them is excluded.

PROFESSOR CAPRA: Right.

MS. SHAMP: And that just costs everybody more time and money, whereas if there's the schedule that sets that up and then, as Judge Saris suggested, maybe a scheduling so that the plaintiffs and the defense expert reports are done at the same time, because what's happening now is that the plaintiff goes first and then the defense gets to cherry pick off of it. If they both have to show their cards at the same time, then I think it would be a more honest process. So those are my suggestions.

PROFESSOR CAPRA: Thanks. Some of those were procedural rather than evidentiary, but they're pretty interesting. Thanks, Laura. [Applause] And now we go to, other than the crowd, the most patient man on earth—Tom Sobol, who is going to give our last presentation.

MR. SOBOL: Thank you. And let me start by saying that given the fact that I do mostly civil litigation, my experience is a little bit different than some of the other discussions earlier today. In all of my antitrust cases, I've never seen the need for bite-mark evidence. [Laughter] I'm hoping that there could be one case when I can, you know, either try to put it in or knock it out somehow.

So preparing here today, my office did a little bit of a homework assignment for us. We looked at how many cases just in the District of Massachusetts there have been that have decided *Daubert* cases. The district court has decided a *Daubert* issue in 190 cases. It's gone up to the First Circuit fifteen times, and the First Circuit has flipped three of those fifteen *Daubert* rulings.

That wasn't really terribly interesting, but what really was interesting was that the statistics backed the gut feeling that *Daubert* challenges have become more and more frequent. So, when we looked at how those had happened over time, if you broke out into three periods since *Daubert* first came out, there were fourteen opinions in the first third; there were seventy-four opinions in the second third; and there were 102 opinions in the last of the three time periods. So, obviously, it's a burgeoning issue.

But, in addition to that, not only is it happening more frequently, but also, in terms of the depth of what happens in any particular case, it's also becoming more and more complicated. So it's not uncommon in antitrust or consumer protection cases that you're going to have at least a half-dozen experts on each side. You probably are going to have more. It's not

uncommon in some complex antitrust cases to have a dozen experts on each side.

If you then imagine that the report that gets filed is between thirty and seventy pages long, not including the attachments or the exhibits, and there's always a rebuttal report, which is going to be about half that length without its attachments, right? And then, of course, there's a deposition, and there are deposition exhibits. So before you even get to the briefs that are going to be filed, there's just a monumental amount of paper that accompanies any particular expert.

Now, from my perspective, I think that *Daubert* presents major case-management challenges, and the reason for that is as a plaintiff's lawyer, I believe that the biggest failing of the judicial system—and I'm glad that this has nothing to do with any of the present company—I found it remarkable that there was no jurist that was going to be here that falls prey to any of my criticisms. [Laughter]

PROFESSOR CAPRA: Or he left, whichever.

MR. SOBOL: Right. Yeah, is that between the time that the case is filed and getting the jury impaneled, the civil justice system is challenged trying to get to that point where you're impaneling the jury. And I'm a plaintiff's lawyer that embraces *Daubert* because I like knocking out the defendant's experts or portions of their reports, which I'll turn to in a moment. So I embrace it. I don't walk away from it like some of my colleagues at times may want to.

But when I hear things like it's going to take a lot of time and effort, what that rings to me is that what we're now creating is yet another hurdle to getting a case to trial, right? And so, in my experience, what I have seen is that the judiciary has reacted in two very different ways to this deluge. On the one hand, the courts at times will use *Daubert* as a reason to get rid of a case and the court will engage in a fact-finding mission finding that the facts that the expert has relied upon in some way are inadequate or whatever and, therefore, the case will go away.

Alternatively, I've seen situations—this is quite common—where the judge will do a drive-by. I don't mean that in too pejorative a way, but will do a drive-by through *Daubert*. So I had, for instance, a major antitrust case here in Massachusetts where there were about twenty-four experts cumulatively for all sides. The *Daubert* hearing took about twenty minutes for all of them. Not a single expert was disqualified or found inadequate because there wasn't the time or the attention or the ability to do so—so that's another way that the courts have defaulted.

Yesterday, in a matter that I was not handling but another antitrust case here in Massachusetts, the court handled seven *Daubert* challenges in seventy minutes by having each side have five minutes per side for each *Daubert* motion. And that's obviously not giving it the kind of attention that I think that Judge Saris would want.

So what I conclude is that *Daubert* poses major case-management challenges, and I have five suggestions for improvement, which I can rattle off very quickly. The first is if a jurist finds him or herself mucking around

with the facts a lot in dealing with *Daubert*, it's probably time to put down the pen and just deny the motion. On the other hand, if what the jurist is doing is saying hey, wait a second, this expert's opinion does not match the law, then you're right in your sweet spot because you are the law and you should be making sure that the expert's opinion matches the law.

The third one I would have is ruthless page limits, ruthless page limits that must be enforced.

JUDGE SARIS: I'm writing that down. [Laughter]

MR. SOBOL: My fourth suggestion is to encourage the litigants to identify portions of the reports that should be disqualified and have them actually give you page and paragraph numbers so the court actually knows what is being struck or not and why. It can really focus it. And the final one is don't move your trial date.

PROFESSOR CAPRA: Thanks. That's a great end. [Applause]

So before I turn it over to Judge Livingston for final remarks, I would like to make some final remarks.

First of all, thanks to the crowd for being so patient and to the Advisory Committee for being so patient. I wish we could have had more time for interchange with the panelists—but, you know, I overstocked the panel. It was great, but we could do this for two days really. And so thank you for your patience. Thanks to the Advisory Committee. Thanks to all the panelists. It was just wonderful. And I don't know why I'm more relieved that we got done in time than for anything that happened today. [Laughter]

I'll have to think about that on the plane home. That's all I can say. And thanks, Judge Sessions, for thinking this up and allowing me to execute it. I appreciate it greatly. And now I turn it over to Judge Livingston.

JUDGE LIVINGSTON: And I'm just going to conclude with thanks, again, to the Boston College School of Law. This has been a great venue for us, and a special thanks to the people—the scientists and the law professors and the practitioners and the judges in this room. You didn't get to hear from us very much today, but let me just tell you you've done important work for us because you give us the food for thought, and now we get to go home and have some time to reflect and to look back at the rule and to prepare for our next meeting, where we will begin together. We will be thinking about it before then, but begin together to work on the problems that you've raised and illuminated today.

PROFESSOR CAPRA: Again thanks to all. And a special thanks to Dan Coquillette, who went above and beyond for us. [Applause]