Fordham Urban Law Journal

Volume 38
Number 2 *The Challenge of Urban Policing*

Article 6

2010

WHOSE FAULT?—DAUBERT, THE NAS REPORT, AND THE NOTION OF ERROR IN FORENSIC SCIENCE

D. Michael Risinger

WHOSE FAULT?—*DAUBERT*, THE NAS REPORT, AND THE NOTION OF ERROR IN FORENSIC SCIENCE

D. Michael Risinger[*]

"Handwriting is even more precise than DNA evidence for identification purposes."[1]

The notion of "error" and "error rates" is central both to the *Daubert* opinion[2] and to the recent NAS Report on the strengths and weaknesses of forensic science in the United States.[3] As might be expected, the NAS Re-

∗ John J. Gibbons Professor of Law, Seton Hall University School of Law. My thanks to Lesley C. Risinger for her usual indispensible aid, editorial and substantive.

1. Detective Chris White, testifying at trial as handwriting expert in *Florence v. Commonwealth*, 120 S.W.3d 699, 701 (Ky. 2003).

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

3. STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD (2009) [hereinafter NAS REPORT]. How to refer to this report has become something of a vexing problem. The report is a report of the Committee on Identifying the Needs of the Forensic Science Community, which is identified on the title page as a Committee of the National Research Council (NRC) (a joint endeavor of the National Academy of Sciences (NAS), the National Academy of Engineering (NAE) and the Institute of Medicine). *See id.* at iii. In addition, the title page suggests some formal conjunction with both the NRC Committee on Science, Technology and Law, Policy and Global Affairs, and the NRC Committee on Applied and Theoretical Statistics. Proper characterization of the report is complicated by the fact that the NRC and the NAE are merely administrative subdivisions of an organization whose name under its charter is simply the National Academy of Sciences. (The NRC was created as a subdivision of the NAS in 1916, and the NAE in 1964 and the Institute of Medicine in 1970. *See id*. at v.) Things are further complicated by the fact that when the NAS created the National Academy of Engineering, it retained the designation "National Academy of Sciences" for the non-engineering members of the academy, and after the most recent subdivisions, it rebranded itself with the umbrella appellation "The National Academies," without amending its charter, so that it is formally the National Academy of Sciences d/b/a "The National Academies." *See History of the National Academies*, http://www.nationalacademies.org/about/history.html (last visited Dec. 15, 2010).

Be all these things as they may, the body of the Report makes clear that the Committee was the result of a 2006 congressional charge to the National Academy of Sciences, that the Committee was formed in response to that charge to the NAS, and that the Report is the work primarily of that Committee. NAS REPORT, *supra*, at 1-2. It has become most common to refer to the Committee on Identifying the Needs of the Forensic Science Community port does a better job of explaining the kinds of error it is concerned with than did the opinion in *Daubert*. However, both to a greater or lesser degree fall short of a full consideration of the concept of error, and so doing, they invite confusion about how inaccurate results in criminal adjudication may occur, and who if anyone is to blame.

When I set out to write on this state of affairs, I was not particularly surprised by it. Courts at all levels are at times notoriously imprecise about important concepts, and the NAS Committee that generated the report was operating in a setting and in a tradition where the notion of error seemed to them, perhaps wrongly, to be intuitively obvious. What I was surprised about when I looked into the matter was that this unexamined approach to the concept of error prevailed in the very discipline where one would expect it to have been carefully taxonomized and theorized to a fare-thee-well in at least a dozen different ways, that is, philosophy in general and epistemology in particular. Nor is this condition any great secret within the philosophical literature.[4] It has been repeatedly noted as a glaring lacuna for

simply as the "NAS Committee" and the report as the "NAS Committee Report," or simply the "NAS Report," although in some circles the Committee is referred to as the "NRC Committee," and the report the "NRC Report." In two previous articles, which were essentially companion pieces, I referred to the report as the "NAS/NRC Report" to satisfy the editors of the first piece, who had adopted the convention "NRC Report" for that journal (JURIMETRICS). *See* D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Glass Nine-Tenths Full (This Is About the Other Tenth)*, 50 JURIMETRICS J. 21 (2009) [hereinafter Risinger, *A Glass Nine-Tenths Full*]; D. Michael Risinger, *The NAS/NRC Report on Forensic Science: A Path Forward Fraught with Pitfalls*, 2010 UTAH L. REV. 225 [hereinafter Risinger, *A Path Forward Fraught with Pitfalls*]. In the current article I have adopted what I believe has now become by far the most common convention, referring to the Committee simply as the "NAS Committee," and the report as the "NAS Report."

4. Lest the reader believe that I have failed to exercise due diligence in making so bold a claim, my research strategy on the issue was as follows. I first looked at every article with the word "error" in the title in the JSTOR philosophy database (126 articles over a period of 117 years). I read the thirty or so articles that seemed most promising, and the ones that were in fact germane are cited herein. I looked at every title in the Library of Congress with the word "error" in it (2226 books and other items). Surprisingly few looked promising at all (there were a lot of fiction books and videos, for instance). Less than half a dozen appeared to be relevant. A couple of relevant items had been reviewed in the *New York Times* book review and I had noted them at the time. The resulting list was not very different from that assembled by Kathryn Schulz in KATHRYN SCHULZ, BEING WRONG: ADVENTURES IN THE MARGINS OF ERROR (2010), and included DAVID W. BATES, ENLIGHTENMENT ABERRATIONS: ERROR & REVOLUTION IN FRANCE (2002); ROBERT A. BURTON, ON BEING IN CERTAIN: BELIEVING YOU ARE RIGHT WHEN YOU ARE NOT (2008); LEO W. KELLER, THE PROBLEM OF ERROR FROM PLATO TO KANT (1934); JAMES REASON, HUMAN ERROR (1990); and HAMARTIA: THE CONCEPT OF ERROR IN THE WESTERN TRADITION (Donald V. Stump et al. eds., 1983). I also familiarized myself with the classical sources referenced in these works. While this search strategy did not guarantee that I had discovered every source discussing the concept of error, I believe that it was sufficient for me to be reasonably confident that nothing notoriously ambitious or successful escaped, and that the quotations given in the following foot-

well over a century.[5] However, very little appears to have been done about it.[6]

---

note reflecting the evaluations by professional philosophers interested in the subject across more than a century are essentially reflective of a continuing consensus.

5.

> I wish now to consider, the elementary question whether an idea, which is one existence, can know an object, which is another; or, as I shall put it, what we mean by true opinion and false, by knowledge and error. The difficulty that waits to thwart us here is among the deepest difficulties in thought. I am not unaware of the literature existing on this subject, from the Greeks through Fichte, Hegel and their followers, to certain writers of the present day, though from some points of view it seems to me surprisingly slight.

Dickenson Sergeant Miller, *The Meaning of Truth and Error*, 2 PHIL. REV. 408, 409 (1893). "Our philosophic predecessors give little help. For though they have usually been eager to point out the errors of their predecessors they have reflected little upon the nature of Error in general, and the little they have contributed to the subject has often been of negative value." F. C. S. Schiller, *Error*, 11 PROCEEDINGS OF THE ARISTOTELIAN SOCIETY, NEW SERIES, 144, 145 (1910-1911).

> In ten successive chapters, Father Keeler searches for a doctrine of error in the works of the pre-Socratics and Plato, Aristotle, the Skeptic, Stoic, and Epicurean schools, Augustine, Aquinas, Scotus and the Spanish Scholastics, Descartes, Spinoza, the British philosophers, and Kant. He is convinced that a history of the problem of error is of crucial significance to epistemology, since 'a theory of knowledge, to be acceptable, must necessarily make room for a complementary theory of error' (p. ix). Yet he finds that the fact of error raised an important philosophic problem for only three of the great philosophers, Plato, St. Augustine, and Descartes, and even these three failed to formulate an epistemology adequate to account for error.

R. McK., reviewing LEO W. KEELER, THE PROBLEM OF ERROR FROM PLATO TO KANT, A HISTORICAL AND CRITICAL STUDY, 31 J. PHIL. 535, 535 (1934).

> Throughout the history of philosophy there has been a sustained interest in the concepts of knowledge, truth and meaning; interest in the concepts of error, falsity and nonsense, on the other hand, has been intermittent and spasmodic. Error, for example, has suffered at the expense of knowledge to such an extent that sometimes its very existence has been denied, or it has been explained away as being merely the absence of or privation of knowledge; many theories of truth are so constructed that no place can be found for falsity, and theories about what constitutes making sense pay, on the whole, little heed to what constitutes nonsense.

J. L. Evans, *Error and the Will*, 38 PHILOSOPHY 144, 144 (1963).

> "The problem of error is one of philosophy's very serious and crucial problems." Alexander Koyre made this remark in a footnote. It epitomizes the state of the problem of error: The problem is "very serious and crucial," yet the treatments it has received have generally been scanty and peripheral, that is, metaphorically they might amount to a footnote.

Giora Hon, *Going Wrong: To Make a Mistake, To Fall Into an Error*, 49 REV. METAPHYSICS 3, 5 (1995).

6. On the other hand, the psychology literature dealing with error from a scientific perspective actually contains some robust taxonomic systems for the classification of error, at least error of the sort that will concern us in this paper. See generally REASON, *supra* note 4, at 1-61, for both the history of these efforts and their modern characteristics.

Needless to say, I will not be attempting a full-scale examination of the concept of error in this paper. However, I believe there are some observations that can be made that may be helpful in domesticating in helpful ways the notion of error as it might apply to forensic science expertise. Error in relation to forensic science presents fewer difficulties than a fully generalized treatment, because there are certain problems necessarily taken on by a full scale philosophic treatment that can be safely put aside. The issues of radical skepticism and the very possibilities of knowledge and error can be properly assumed away, for instance, because the givens of the law as a practical enterprise resolve those for the purposes of the law.[7] And the dif-

7. Professor Twining summarizes these foundational assumptions of system of legal proof thus:

> 1.Knowledge about particular past events is possible.
>
> 2.Establishing the truth about particular past events in issue in a case (the facts in issue) is a necessary condition for achieving justice in adjudication; incorrect results are one form of injustice.
>
> 3.The notions of evidence and proof in adjudication are concerned with rational methods of determining questions of fact, in this context operative distinctions have to be maintained between questions of fact and questions of law, questions of fact and questions of value, and questions of fact and questions of opinion.
>
> 4.The establishment of the truth of alleged facts in adjudication is typically a matter of probabilities, falling short of absolute certainty.
>
> 5.(a) Judgments about the probabilities of allegations about particular past events can and should be reached by reasoning from relevant evidence presented to the decision-maker. (b) The characteristic mode of reasoning about probabilities is induction.
>
> 6.Judgments about probabilities have, generally speaking, to be based on the available stock of knowledge about the common course of events; this is largely a matter of common sense supplemented by specialized scientific or expert knowledge when it is available.
>
> 7.The pursuit of truth (i.e. seeking to maximize the accuracy in fact determination) is to be given a high, but not necessarily overriding, priority in relation to other values, such as the security of the state, the protection of family relationships, or the curbing of coercive methods of interrogation.
>
> 8.One crucial basis for evaluating "fact finding" institutions, rules, procedures and techniques is how far they are estimated to maximize accuracy in fact-determination—but other criteria such as speed, cheapness, procedural fairness, humaneness, public confidence and the avoidance of vexation for participants are also to be taken into account.

ficult issue of normative error can also be sidestepped,[8] because forensic science explicitly deals only with conclusions about empirical facts.

However, limiting ourselves to the notion of factual error still leaves plenty to do. To begin at the beginning, such error can only exist as a function of the human mind. There can be no error except in regard to a belief or action that is judged at a later time by some human agent to have been wrong in some respect, by some invoked criterion. A mouse's stillbirth is not an error, merely an event (although it may be referred to in anthropomorphic metaphor, and somewhat misleadingly, as an "error of nature"). Independent of sentient belief, purpose or action, there is no error in the natural world. Whatever happens simply is. For our purposes, only humans can make errors, or be in error.[9]

There are in fact two different fundamental approaches to the concept of error which are important to consider, and which are significantly in tension with each other.[10] We may label them the normative idea of error and, for want of a better term, the objective idea of error. Failing to separate the two can lead to various confusions and troubles.

Although in the end it is not really important which is taken as primary, I regard the normative notion as foundational both in normal usage and in underlying psychology. In this most fundamental sense, an assertion of er-

---

> 9.The primary role of applied forensic psychology and forensic science is to provide guidance about the reliability of different kinds of evidence, and to develop methods and devices for increasing such reliability.

WILLIAM TWINING, RETHINKING EVIDENCE 73 (1990).

8. The most fully developed consideration of normative error (that is, error about a value judgment), what it might be, and whether it is even a coherent concept, is to be found in the substantial literature on "moral error." For current purposes, the best reference is simply the instruction: "insert the term 'moral error' into any search engine and see what results." Note that, as we shall see, bypassing examination of the concept of normative error in this paper does not prevent us from considering normative theories of factual error.

9. See the discussion in REASON, *supra* note 4, at 5-10. We need not pursue the question of the extent to which other animals, or beings yet unknown, are sufficiently sentient to be said to make errors in the strict sense we are using. For legal purposes we properly restrict ourselves to humans.

10. There is at least one other current attempt to identify inconsistent accounts of the phenomenon of error that are in tension with each other. In a recent popular book on the subject, journalist Kathryn Schulz suggests there are two ways of looking at error, the first of which considers error to be bad, and the second of which considers that error has positive aspects as a way station on the journey to more secure knowledge. *See* SCHULZ, *Two Models of Wrongness*, *in* BEING WRONG: ADVENTURES IN THE MARGINS OF ERROR, *supra* note 4, at 25-43. This book is well researched and has some interesting things to say, but it is undermined by the author's failure to police her assertions for consistency and her penchant for selecting the most dramatic way of asserting her positions over alternatives that might have been more nuanced.

ror entails a claim or charge of both mistake and fault, that is, a proposition that asserts that a belief, claim, or action is or was wrong in some respect, coupled with an argument invoking a ground to assert its wrongness which is made in such a way as also to assert that the person whose belief, claim, or action is under consideration was at fault in taking the action, or was at fault at the time in holding the belief or asserting the claim, or at least would now be at fault in continuing to do so. This in turn entails the proposition that there can be legitimate grounds to assert the wrongness of a belief or action, and to criticize another for so believing or acting. At base, an assertion of error in this sense is an attack on both the truth of and the justification for an explicit or implied (through action) knowledge claim, thus connecting the concept of error intimately with the most common traditional definition of knowledge as "justified true belief."[11]

On a general level, the territory of belief, claim, or action subject to a potential charge of error is coterminous with the extent of potential beliefs, claims, or actions. The grounds for charging error of belief or claim are at least equally extensive, while the grounds for charging error of action include attacks on any belief or claim that motivated the action, plus attacks on the propriety or rationality of the action even given the accuracy of those beliefs or claims.

Again on a general level, the range of grounds for charging error include such things as divine revelation, logical inconsistency, esthetic preference, and a multitude more. For our purposes we must very quickly narrow both the kinds of belief, claim, or action under consideration, and the grounds that will be considered in-bounds for making a charge of error. After all, our main focus is expertise proffered in courts of law, and forensic science as it is currently practiced, and the kinds of error and the kinds of argument properly under consideration in this context are considerably (and blessedly) much narrower than the entire possible field of the erroneous.

Within this more constrained field, the notion of error and the acceptable grounds for asserting error become much more limited. First, as previously noted, we are dealing with beliefs or claims concerning empirical facts in the world. Both the grounds for justifying such beliefs and the grounds for attacking them are limited, in general, by notions of empirical evidence, either of the critical common sense variety, or of the formal variety which is

11. On the centrality of "justified true belief" in the "standard analysis" of knowledge, see, e.g., SUSAN HAACK, EVIDENCE AND INQUIRY: A PRAGMATIST RECONSTRUCTION OF EPISTEMOLOGY 306 (expanded ed. 2009); MICHAEL WILLIAMS, PROBLEMS OF KNOWLEDGE 17-19 (2001).

the domain of science in the modern sense of the word.[12] So a charge of error in the normative sense is a charge that the person or group has beliefs that are unwarranted by empirical evidence, or undertakes various practices the results of which do not mean or deliver what is claimed for them.[13] In this sense, a claim of error may be a profoundly serious moral claim, especially if the alleged errors are taken as truth by various social actors in a way that injures a human or a group of humans.

However, there is another way the term "error" is commonly used, especially in the setting of science,[14] and most especially in the science of testing (the most relevant scientific discipline for our purposes, as we shall see), where there is no normative charge at all, at least in the primary setting in which the word "error" is invoked. This approach to "error" only applies to results, and it is a purely post hoc judgment. In this context, a decision based upon a belief that something is the case, or even that is it most likely to be the case, is an error if it turns out wrong, no matter how strong the warrant for the belief. On the other hand, a decision that turns out right because the hoped-for result obtains, is generally not an error.[15] A set of four hypotheticals will illustrate the contrast between the two notions of error.

---

12. *See* D. Michael Risinger, *The Irrelevance, and Central Relevance, of the Boundary Between Science and Non-Science in the Evaluation of Expert Witness Reliability*, 52 VILLANOVA L. REV. 679, 705-06 (2007).

13. Normative theories of error have the most difficulty with inadvertent actions that result in unintended bad consequences. Theories of negligence are essentially attempts to bring a normative theory of error to bear on such actions.

14. The entire frame of reference for the error taxonomy set out in REASON, *supra* note 4, is in general a version of the non-normative frame of science, although, since Reason is concerned not only with defining and understanding types of error, but also with designing systems with an eye to taking steps to prevent errors, normative judgments must be reimported at the system design level to determine the proper prescribed responses to errors necessary to minimize them. *See id*. at 194-95 (distinguishing between "errors" simplicitur and "violations"). In its modern cradle period, scientific thought did not always separate the objective and normative approaches to error. See, e.g., Francis Bacon's famous discussion of "idols" of the mind as sources of error in 1 FRANCIS BACON, NOVUM ORGANUM, at pts. 1-68 (1620), *reprinted in* 30 GREAT BOOKS OF THE WESTERN WORLD 107-16 (Robert M. Hutchins ed., 1952). The primary initial locus for the objective concept of error appears to have been precipitated by contemplation of measurement error. *See* EDWARD G. BORING, A HISTORY OF EXPERIMENTAL PSYCHOLOGY 134-35 (1929). From there it expanded to cover statistical error in all sorts of settings. See, e.g., YARDLEY BEERS, INTRODUCTION TO THE THEORY OF ERROR (1957) (observational and statistical errors in physics), including the standard empirical testing literature represented by Chapter 4 of the NAS Report. *See, e.g.*, Brad J. Bickerstaff, *Comparing Diagnostic Tests: A Simple Graphic Using Likelihood Ratios*, 19 STAT. MED. 649 (2000).

15. Theories of probability and counterfactual prediction can be brought to bear to characterize both intentional and negligent acts that turn out well as errors in an objective sense, but we need not pursue this for the purposes of this paper.

Assume that, in a basic five-card draw poker game, a player draws to an inside straight believing this to be a good strategy.[16] He does not fill the straight. Was the decision to draw an error? Clearly from the objective point of view the answer must be yes. The result was not the one intended or desired or hoped for, and it redounded to the player's detriment in the player's own terms. But what about the normative perspective? Here the answer is likely to be yes also. One of the first general rules one learns in regard to playing poker is that it is a bad gamble to draw to an inside straight. The chances of filling it are remote, and the hand without the fill is nearly worthless. So by the standards of intelligent poker play, assuming the object is to win, the belief that drawing to an inside straight is an intelligent strategy, whether based on ignorance or mistaken belief that could have been avoided by reflection or education, or on a belief in the special instinctive hunch powers of the player in spite of knowing the odds, is an error, and the act based on that belief, is an error. And this would be so whether the player filled the straight or not (which will happen a little more than once every twelve attempts).[17] A belief that this is a good strategy is still an error, and a lucky shot even knowing the odds doesn't change the arguments concerning the nature of the action (decision), since this is always analyzed ex ante. The "no harm, no foul" principle does not apply to the determination of normative error.

But in the objective framework, the filled straights would not count as "errors" at all, since this is always analyzed ex post merely by reference to desired results.

Now consider this further homely hypothetical. A person notices that the (donated) prize at a church lottery is worth more than the total cost of the one thousand lottery tickets to be sold. He has arrived after only one ticket has been sold. He buys the other 999 tickets. The only ticket he has not bought is the winner. Was it an error to buy the 999 tickets?

From the objective perspective, the ticket purchase would be an error, since the decision came out wrong, albeit against long odds. Again, the

---

16. An "inside straight" hand is any hand with mixed suits having four out of five numbers for a straight (2, 3, 5, 6, for example), but missing a number, not on the end of the series (which could be completed by any card bearing the number on either end), but a number in the middle, which can thus be completed by drawing only a card with one specific number on it. As noted in the text, one of the first heuristic rules that a beginning poker player learns is "never draw to an inside straight," and in fact this was a part of the tagline exchange upon parting between gambling brothers Bret and Bart Maverick in the original *Maverick* TV series. ("Never hold a kicker"; "Never draw to an inside straight.") (original memory on file with author).

17. Four cards out of the remaining forty-seven unknown cards will fill, so drawing one of the four will happen about every 11.75 attempts.

classification as error rests on comparing actual outcomes to desired outcomes ex post. But from the normative perspective the decision was not erroneous, since the belief that motivated the action was fully justified, and provided an adequate ground for the action based upon it.[18]

As a general proposition, the objective approach to error is the approach most commonly adopted in the sciences, and particularly in regard to approaches to hypothesis testing and the design and evaluation of tests in scientific contexts, whereas the normative framework is the approach most commonly embraced in most other areas of endeavor.[19]

At this point you may start to have an inkling of how a failure to account for these two common and competing notions of the essence of error might result in misunderstanding and hurt feelings when claims about error are made from the objective perspective of science, but heard as normative attacks alleging negligence or worse by the members of the group under discussion. I believe this has contributed to the defensiveness of various parts of the forensic science community when the question of error and error rates in forensic science has been raised. And to a substantial extent, I think this has been the fault of Sherlock Holmes, a point to which I will return below.

First, however, the perceptive reader will have noticed that I have just used the terms "error" and "error rates" together in such a way as to echo their use in the majority opinion in *Daubert*.[20] What did the court intend by those words in that opinion?

One of the most troubling aspects of the *Daubert* opinion is the unthinking way in which the Court tossed off the so-called "*Daubert* factors." As I have pointed out elsewhere, while it has become conventional (though not universal) to speak of "the four *Daubert* factors," it is not even clear how many "*Daubert* factors" there really are.[21] And beyond the issue of proper

18. We are, of course, bypassing any discussion of the deeper normative question, the moral question about whether it is proper to take advantage of one's superior knowledge of the item donated (and thus not subject to the direct knowledge of the church authorities regarding value) in order to deprive the charity of the excess marginal value of the prize above the total realized by selling all the tickets.

19. In this way the concept of error tracks usage pattern similar to that of the notion of "bias," one factor that can account for objective errors in the results of various processes. For a discussion of the normatively neutral concept of bias in psychology, see D. Michael Risinger, Michael J. Saks, William C. Thompson & Robert Rosenthal, *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CALIF. L. REV. 1, 10-12 (2002).

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

21. Referring to "four" factors has become standard, though the real number of factors is subject to debate. The *Daubert* opinion spake thus, without numbering factors: "a key question [in regard to a theory or technique] . . . will be whether it can be (and has been) tested."

enumeration, the Court's language has created many more problems in the lower courts than it has solved. Here I will concentrate on the issues raised by the Court's invocation of "known and potential error rate" as something to be considered in determining whether proposed expert testimony is sufficiently reliable for the purposes of the law under Federal Rule of Evidence 702.

I have reflected on that puzzle for nearly two decades, and as near as I can figure, neither Justice Blackmun, nor the other Justices, nor whatever law clerk may have supplied the locution, had a clear idea of what was meant, or how it should be or would be interpreted. Here is the pertinent language from the opinion: "[I]n the case of a particular scientific technique, the court should consider the known or potential rate of error."[22] The Court then cites *United States v. Smith*, a 1989 Seventh Circuit case which it says surveyed "studies of the error rate of the spectrographic voice identification technique."[23] So the drafter of this language in *Daubert* ap-

509 U.S. at 593. "Another pertinent consideration is whether the theory or technique has been subjected to peer review and publication. Publication (which is but one element of peer review) is not a sine qua non of admissibility . . . ." *Id.* "Additionally, in the case of a particular scientific technique, the court should consider the known or potential rate of error . . . and the existence and maintenance of standards controlling the technique's operation." *Id.* at 594. "Finally, 'general acceptance' can yet have a bearing on the inquiry." *Id.* These were summarized in *Kumho Tire* as "several factors" without numbering, but with four bullet points. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 149-50 (1999). However, it is easy to separate whether a claim "can be tested" (its empirical nature or theoretical falsifiability) and the degree to which it has been subjected to actual testing, into two separable but nested factors. In addition, the potential rate of error is arguably always 100% in the absence of some kind of testing (though not necessarily the kind of formal testing that would lead to more specific and quantifiable knowledge of an error rate). Knowledge of error rates is thus a product of testing. In addition, can "standards of control" for a technique's operation be a relevant factor if there is no reason to believe such "standards" enhance reliability? This too would seem to be a question of testing, at least in some contexts. Finally, a fortiori "general acceptance" is the product of peer review, so one can argue that there are really eight explicitly referenced "*Daubert* factors" (falsifiability, testing, peer review, publication, potential error rate, known error rate, standards of practice, general acceptance) or only three (falsifiability, testing which reveals error rate, peer review). In addition, the *Daubert* Court invokes the relevance-based concept of "fit," 509 U.S. at 591, which is perhaps best seen as an analogue to "external validity," and which can easily be asserted as a fifth (or ninth, or fourth) "*Daubert* factor." *See* Mark P. Denbeaux & D. Michael Risinger, Kumho Tire *and Expert Reliability, How the Question You Ask Gives the Answer You Get*, 34 SETON HALL L. REV. 15, 32 n.64 (2003). Courts have not always referred to four *Daubert* factors, either. *See, e.g.*, United States v. Crisp, 324 F.3d 261, 266-67 (4th Cir. 2003) (five factors); United States v. Prime, 220 F. Supp. 2d 1203, 1204 (W.D. Wash. 2002) (five factors); United States v. Griffin, 50 M.J. 278, 284 (A.F. Ct. Crim. App. 1999) (six factors). The Advisory Committee's note for the 2000 revision of Rule 702 lists five factors, but then adds five more that were derived from intervening case authority.

22. *Daubert*, 509 U.S. at 594.

23. *Id.* at 594 (quoting United States v. Smith, 869 F.2d 348, 352-54 (7th Cir. 1989)).

peared to have in mind how well a "particular scientific technique" operates to do the job humans use it for. I have phrased this carefully to lead up to the next point. Assuming standard protocols are defined and observed, techniques, tests, processes, and the like, scientific or not, do not themselves commit "errors" per se. They merely have results. It is the use to which humans put those results that constitutes or gives rise to error in the objective sense when they turn out to generate wrong answers to the questions addressed. However, in many testing fields it is common (although not universal) to speak somewhat loosely and ascribe resulting errors to the test in a kind of short hand. Thus, a test may be said virtually interchangeably to have a "false positive rate" (which definitely is a characterization of the test results per se) and a "false positive *error* rate" (which is actually a characterization of what occurs when humans make use of the test by applying it at face value as a heuristic in a particular way to make decisions, take actions, etc.). Note that in either case, the "error" that is referred to is simply an inaccurate correspondence between test result and true conditions in particular cases, independent of any necessary criticism or assertion of fault in regard to anyone's actions, beliefs, or decisions.

So it would appear that the Court in *Daubert* seemed to have had in mind the objective notion of error applied conventionally to tests and techniques in the area of scientific hypothesis testing in general and the testing of various asserted diagnostic tests in particular. This would make sense given its reference to a forensic science case as illustrative, since, taken on a general level, most of forensic science deals with various kinds of processes and tests claimed to be diagnostic of one thing or another—the fact that a test result indicates that a certain crime scene residue can be attributed to a particular source, for example.[24]

24. The security of this conclusion about what the Court had in mind is undermined a bit by the invocation of the mysterious phrase "potential error rate." If the phrase "potential error rate" is a term of art with a determinate meaning in some technical area, I haven't discovered it yet. On its face the concept verges on nonsensical, especially when apparently contrasted with "known error rate." If a diagnostic test is put forward with no information on how often apparently positive results indicating the condition of interest are false, and how often apparently negative results are true, then the "potential error rate" that could result from adopting the test is 100% (which is much worse than random). This is because it is theoretically possible to have a process that is reliably always wrong (and remember, in the hypothetical, we have no information indicating that this is untrue of the test in question). Of course, if this is known, and if the issue is a binary issue of yes/no or right/wrong (as it is in most forensic identification contexts, at least), then such a process could be used as a perfectly accurate process by always accepting the opposite of its output as true. But that assumes that this circumstance is known. If it is not known and the process is taken at face value, then the actual error rate resulting from relying on the process would be 100%, which, again, would be much worse than random. It is unlikely that any process actually fulfills the 100% potential error rate possibility, but some processes when tested do some-

If, as it appears reasonably likely, the Supreme Court in *Daubert* had this notion of "error" in mind, it is clear that the NAS Committee explicitly adopted this as its model of error. Indeed, more than half of Chapter 4 of the NAS Report, "The Principles of Science and Interpreting Scientific Data" is devoted to explaining this notion of error in the context of testing, with worked examples.[25] Whether this twenty-four-page chapter captures all that would be suggested by its title is debatable, but it certainly describes the main characteristics of science's idea of error in regard to validation testing well enough. It sets out the standard procedure for determining the diagnosticity of a binary test for a condition such as a disease—creating a test population made up both of subjects known to have the condition of interest and known not to have the condition, running the test under evaluation on that population, and entering the results in the familiar 2 x 2 four-cell matrix for false positives, true positives, false negatives and true negatives, then deriving from those results figures for sensitivity, specificity, positive predictive value, and negative predictive value. The Report then sets out the results of a hypothetical validation study of this nature that might be performed on the visual hair comparison technique, and the light its results might shed on the common origin of two hairs, one from a crime scene and one from a person of interest in regard to that crime. This clearly represents the fundamental structure for validation of heretofore unvalidated forensic techniques envisioned as necessary by the NAS Committee.

---

times score worse than random. Bite mark identification is one notorious example. See the results of proficiency tests taken by members of the American Board of Forensic Odontologists set out in Michael Bowers, *Identification from Bitemarks: Scientific Issues*, 4 MOD. SCI. EVIDENCE 680-87 (David L. Faigman, David H. Kaye, Michael J. Saks, Joseph Sanders, Edward K. Cheng eds. 2009). To make a long story short, the "potential error rate" is in this view the same as the "known error rate" whenever there is any knowledge warranting the conclusion that the potential error rate is not 100%.

Or perhaps the Court meant the notion of known error rate to be reserved for circumstances where there is good empirical evidence that indicates a fairly mathematically determinate error rate, and the notion of potential error rate was to cover limits of error rates known by reference to more qualitative and less quantitative sources of information, which establish broad ranges of potential error which are indeterminate at their specific boundaries. Or perhaps it was a caution meant to remind people that false positive and false negative rates are not the same as error rates if they are being applied to test subjects drawn from a candidate universe with base rates of occurrence of the condition in question different from the "known-condition" population that was used to generate the rates for true and false positives and true and false negatives to begin with (although that seems unlikely). Perhaps the best thing to do with the concept of "potential error rate" would be to ignore it, as I have largely done in the text. Unfortunately, since it was uttered by the Supreme Court, litigants and lower courts find themselves obliged to deal with it in some way which is generally destined to make very little sense.

25. NAS REPORT, *supra* note 3, at 111-25.

It also represents the Committee's adoption of science's standard objective notion of error when techniques are used as tests of common origin—how often a human using the test as determinative would be wrong (or in error, if you will). But the Report is careful to note that the accuracy of the test is dependent on the question it is being used to answer. They designed their hypothetical visual hair comparison test to be directed at determining only whether such comparison would be accurate in assigning hairs to predesignated classes of hair. They emphasized that "this accuracy evaluation does not apply to other tasks that are beyond the goal of the particular analysis, such as pinpointing the individual from whom the specimen was obtained."[26] This is of course right as far as it goes, and anyone who has ever read anything I have written in this area knows that I will applaud the realization that validation is a task-specific enterprise.[27]

Earlier in the Report, however, the Committee made an observation, that perhaps, if developed properly, has the potential to ameliorate at least to some degree the feeling among forensic scientists of being wrongly charged willy nilly with error in a normative sense:

> Assertions of a "100 percent match" contradict the findings of proficiency tests that find substantial rates of erroneous results in some disciplines (i.e., voice identification, bite mark analysis).
>
> As an example, in a [sic] FBI publication on the correlation of microscopic and mitochondrial DNA hair comparisons, the authors found that even competent hair examiners can make significant errors. In this study, the authors found that in 11 percent of the cases in which the hair examiners declared two hairs to be "similar," subsequent DNA testing revealed that the hairs did not match, which refers either to the competency or *the relative ability of the two divergent techniques to identify differences in hair samples*, as well as to the probative value of each test.[28]

In the referenced article by Houck and Budowle, the authors are at pains to say that the results of their study do not indicate that the visual hair comparisons were "erroneous," but merely indicate the difference in resolving

26. NAS REPORT, *supra* note 3, at 120.

27. *See, e.g.*, D. Michael Risinger, *Defining the "Task at Hand": Non-Science Forensic Science After* Kumho Tire v. Carmichael, 57 WASH. & LEE L. REV. 767 (2000); D. Michael Risinger, *Goodbye to all That or a Fool's Errand, by One of the Fools: How I Stopped Worrying About Court Responses to Handwriting Identification (and Forensic Science in General) and Learned to Love Misinterpretations of* Kumho Tire v. Carmichael, 43 TULSA L. REV. 447 (2007).

28. NAS REPORT, *supra* note 3, at 47 (citing Max M. Houck and Bruce Budowle, *Correlation of Microscopic and Mitochondrial DNA Hair Comparisons*, 47 J. FORENSIC SCI. 964 (2002) (other internal citations omitted)) (emphasis supplied).

power between the visual technique and mtDNA analysis.[29] They are, in essence, saying that in classifying the pairs of hairs in each case as visually indistinguishable according to the standards of their discipline, the visual hair analysts were not mistaken. This brings into play not only the objective versus normative versions of the concept of error, but another distinction which is one of the few helpful taxonomic plays in the philosophic literature on "error," the Israeli philosopher of science Giora Hon's proposed distinction between error proper and mistake.[30] For Hon, a mistake is something which is wrong by reference to determinate criteria conceded to be applicable to a belief, act, or decision within the normal practice of an area of knowledge.[31] A calculation error is a mistake. Flipping the wrong switch in a cockpit is a mistake. In visual hair comparison, mixing up samples would be a mistake, and mischaracterizing color would be a mistake. The point here is that people who make mistakes do so through kinds of inattention or sloppy practice for which they are responsible, and it is proper to criticize them for such mistakes. Mistakes and various procedures for their elimination, or at least reducing them to a minimum, are the stuff of human factors research and rumination. They are also the stuff of accreditation, individual certification, and so forth. I am convinced that most forensic science practitioners think that all the error talk they hear is charging them with making large numbers of mistakes which they feel sure is not the case, and they feel put upon as a result. But the most serious sources of error in the objective sense are not mistakes, but inherent limitations on the precision of tests utilized—their diagnosticity, if you will, determined by analogy to specificity and sensitivity in other contexts, their false and true positive and negative results as applied to the tasks they are used as tools to resolve, just as in the case of other diagnostic tests. Such bad outcomes are not due to practitioner mistake, but to limitations in the technique even when perfectly applied. This, in Hon's terms, is the stuff of error.[32]

Hon's fairly rudimentary taxonomic move is quite helpful and potentially fruitful, but of course it does not accommodate or distinguish between many kinds or sources of objective error. To accommodate those types, we must move to the much more sophisticated error taxonomies to be found in the psychological "human factors" and "error studies" literature, perhaps

29. Houck & Budowle, *supra* note 28, at 966.
30. *See* Hon, *supra* note 5.
31. *Id*. at 6.
32. *Id*. at 6-10.

best represented by that outlined by James Reason in 1990.[33] For instance, while it is hopefully rare, errors that are hard to classify as "mistakes" can resort from affirmative malfeasance such as "dry labbing,"[34] and from levels of disregard of good practice that cross a line between mistake due to inadvertence or simple negligence, to gross negligence approaching affirmative malfeasance. These types of error source are easily accommodated under Reason's rubric of "violations."[35] On the other end, wrong results stemming from observer effects that induce unconscious bias which changes decision thresholds are hard to classify even as negligent mistake on the part of the individual examiner, given the unconscious nature of the phenomenon, and the failure of those above their pay grade to adopt any protocols to control the precursors of such observer effects. But Hon's simple distinction still makes an important foundational point: There are wrong results that are properly regarded as somebody's fault, and errors that are no-fault—that is, they are an inevitable result of the imperfections

33. *See* REASON, *supra* note 4. Reason's book is marked both by its historical account of the development of such contemplations of human error and how to control them, and by its taxonomic sophistication as well as its substantive conclusions, and is clearly the classic of this literature. However, it must be noted that Reason's work grew from a rich ferment that began with Professor of Mechanical Engineering John W. Senders' organization of two important conferences to study "error as a behavioral phenomenon in its own right rather than simply as an index of performance." JOHN W. SENDERS & NEVILLE P. MORAY, HUMAN ERROR: CAUSE, PREDICTION AND REDUCTION 9 (1991). These were the 1980 Human Error Conference in Columbia Falls, Maine in 1980, and (in collaboration with Neville P. Moray) the Conference on the Nature and Sources of Human Error in Bellagio, Italy in 1983. Professor Reason was a participant in both, and acknowledged their seminal value. REASON, *supra* note 4. Senders and Moray worked for a number of years on a volume to summarize the various questions posed and positions taken by the participants in those conferences both during the conferences and in later publications and follow-up responses to Senders' inquiries. This volume, SENDERS & MORAY, *supra*, was not published until the year after Reason's book, but it contains much of interest in its own right on the general nature of the questions then in play among the "error studies" community, and should actually be read not only in conjunction with Reason's book, but ideally before it.

34. "Dry labbing" occurs when a forensic technician simply fills in the results of a test report without actually doing the tests. *See* Paul C. Giannelli, *Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs*, 86 N.C. L. REV. 163, 184 (2007).

35. "Violations" are intentional decisions not to follow standard procedures, sometimes because risks appear slight, and sometimes for more questionable reasons. REASON, *supra* note 4, at 194-95. Reason classifies error according to the cognitive processing level upon which they occur. "Slips and lapses," occur when recurrent skilled routines miscarry, most commonly because of attention disturbances. "Rule-based errors" occur when a decision maker selects the wrong pre-packaged routine or rule for a problem that is presented, often when a skill miscarries; "knowledge-based errors" occur when one has no proper rule or heuristic to resort to and must attempt to solve a problem presented by conscious analysis from general principles. *See id.* at 94 fig.3.1. While each of these distinctions is important, I actually prefer Hon's term "mistakes" as an umbrella term to cover all of these together, and to distinguish between such "mistakes" and error flowing from weak diagnosticity on the one hand, and "violations" on the other, and I have so used it in the text.

of our knowledge at any given time. And there are errors that are not the fault of the bench analyst, who is doing everything right according to his training and his role in the system, but may properly be regarded as the fault of the persons who are in positions of power over the process, because those in power now have available to them new information that makes it at least a mistake, and perhaps worse in its willful disregard, not to undertake the changes in the standards and practices that would reduce error to its unavoidable minimum.[36]

No one likes to be wrong.[37] And everybody likes to be right. This truism has been the subject of a considerable amount of published reflection recently,[38] the point of which is to argue roughly that: (1) The double phenomenon is motivated by feelings of rightness (or "knowing" or belief) and wrongness that are inherently pleasant and unpleasant respectively, being analogous to positive and negative emotions in their deployment of neurochemical rewards; (2) The evolutionary benefits of such a reward system are substantial, motivating and sustaining all sorts of behaviors such as hunting, exploring, etc.; and (3) that the buzz we get from feeling right can lead to all sorts of gambling behaviors and over-investments of belief on thin evidence, so long as the likelihood of being shown to be wrong by evidence that cannot be ignored is low (which it usually is, given our ability to

36. In a recent article, James M. Doyle has proposed dealing with the whole problem of inaccurate results in the criminal justice system by the adoption of an error studies approach modeled on that adopted in recent medical quality control efforts, particularly those championed by Dr. David Berwick, and reflected in the landmark National Institute of Medicine publication TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM (Linda T. Cohen, Janet M. Corrigan & Molla S. Donaldson eds., National Academies Press, 2000). *See* James M. Doyle, *Learning from Error in American Criminal Justice*, 100 J. CRIM. L. & CRIMINOLOGY. 109, 119 et seq. (2010). The vast implications and potential objections to such an approach are of course beyond the scope of this article.

In a narrower context, a group (of which I was a member) led by Jennifer Mnookin has very recently recommended that similar mechanisms be adopted in order to allow such an approach to error, and to the lessons to be learned from clearly identified instances of error, in forensic science practice. *See* Jennifer L. Mnookin, Simon A. Cole, Itiel E. Dror, Barry A. Fisher, Max Houck, Keith Inman, David H. Kaye, Jonathan J. Koehler, Glenn Langenburg, D. Michael Risinger, Norah Rudin, Jay Siegel & David A. Stoney, *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA. L. REV. (forthcoming Feb. 2011) (manuscript 53-57) (on file with authors).

37. On the powerful unpleasantness of the feeling of having been wrong, see SCHULZ, *supra* note 4, at 25-27.

38. *See, e.g.*, BURTON, *supra* note 4, at 1-40. Burton is a neurologist. The main object of his book seems to be to elevate certain "feelings," most particularly the feeling of "knowing" to phenomena on a par with emotions in their ability act as neurochemical motivating rewards (or punishments). ". . . [T]he *feeling of knowing* and its kindred feelings should be considered as primary as the state of fear and anger." *Id.* at 40.

ignore counterevidence once we have obtained the feeling of rightness, in order to avoid the feeling of wrongness).

Now of course all of this is currently highly conjectural and in the beginning stages of investigation. But the general insight that there is a neural reward system underlying feelings of rightness, knowing and belief independent of content, which resists revision that would trigger the unpleasant feeling of being wrong and promotes various rationalizing behaviors in the face of new evidence, seems to me likely to prove right in some form. Assuming this to be the case, people seeking to improve forensic science as it is currently practiced in order to eliminate as many sources of error as possible must take into account these sources of resistance, and take such reasonable steps as are available to ameliorate them. It is perhaps a small contribution to that end to point out that the vast majority of forensic practitioners in accredited laboratories probably make relatively few mistakes in their performance of the standard skill-based practices that they have been taught. Furthermore, it could hardly be expected of them to reject the prevailing group view of both the powers of their discipline and the reasons for believing in those powers when they were taught those things as part of their training and practice. They inevitably came to regard those positions as right, with all that that entails in regard to the rewards of belief and the difficulty of revision. These beliefs were promulgated in reliance on apparently well-grounded authority, and seemed reasonable at the time. While errors in the objective sense may be attributed to the deficiencies of that vision exposed by new information and fresh reflection, it would be wrong to attribute those errors normatively to those that practice what they have been taught with few mistakes according to the standards, prevailing beliefs and practices of the enterprise in which they are embedded. Whether a similar absolution can be given to the leaders of the forensic science enterprise is another question, to which I will return below.

I have said earlier, somewhat cryptically perhaps, that I blame Sherlock Holmes. In fact, I have written somewhat extensively in other places about why I think it is reasonable to blame Holmes, and the strong version of "heroic positivism" that he represented and helped to popularize (especially in the budding forensic science fields of the late nineteenth and early twentieth centuries) for some of the failure of forensic science to adopt a more modern vision of science and its requirements than it has done.

It has always been striking to me the extent to which practitioners of forensic science in general and the traditional pattern matching disciplines in particular, took for their patron and the instantiation of their self image, their ideals, and their mission, not a flesh and blood human pioneer who had advanced the science of their enterprise in some empirically reliable way, but an imaginary person: Sherlock Holmes. As I have written previously,

> [E]mbracing Holmes in this way is perhaps understandable. As I just indicated, many of these areas of forensic practice and claimed expertise grew up in the same late Victorian period in which Holmes held forth from the imaginary address of 221B Baker Street, and some credibility was bestowed upon them directly by Holmes himself by references in the stories. So it is easy to see how those interested and involved in developing these emerging forensic specialties might view Holmes as a kind of avatar of their enterprise, and an ideal toward which to direct their efforts. This identification had some good effects perhaps, but it also had significant bad effects, trapping these "forensic sciences" in a late Victorian model of thought which both stunted their growth and limited their reliability, while at the same time providing an explanatory account of their claims to expertise which would be embraced by courts, and (given the conservative nature of the legal system and the judges within it) be retained by courts as their world view long after its expiration date. . . .
>
> The Holmes presented in the Canon is the superman of 19th century positivism. He is embedded in a world in which certain knowledge of an event is in principle always attainable from later circumstances, if only a person knows enough and can process what is known correctly. This is the fundamental position first given voice in the early 19th century by, ironically, the probability theorist Henri LaPlace (and still embraced by some). It was so congenial to the 19th century materialist determinism (positivism) that it became the dominant lens of the scientifically minded for all phenomena, including human motivation. As Stephen Kern explains in the introduction to his recent, brilliant book A CULTURAL HISTORY OF CAUSALITY:
>
>> A materialist determinism applied to mental life peaked with the "mental physiologists", such as Henry Maudsley, who, in 1874, argued that "lunatics and criminals are as much manufactured articles as are steam-engines and calico printing machines." The French essayist and fictionist Paul Bourget elaborated such thinking in his novel The Disciple (1889) which ridiculed the extreme positivism of one arrogant character, who updated Pierre LaPlace's famous determinist hypothesis of 1814 in speculating "if we could know correctly the relative position of all the phenomena which constitute the actual universe, we could, from the present, calculate with certainty equal to that of the astronomers the day, the hour and the minute when England will evacuate India . . . or when a criminal, still unborn, will murder his father."
>
> What was intended as caricature by Bourget was not so intended by Conan Doyle, however, in one of the early Holmes stories, "The Case of the Five Orange Pips":

> "The ideal reasoner" [Holmes] remarked "would, when he had once been shown a single fact in all its bearings, deduce from it not only all the chain of events which led up to it but also all the results which would follow from it. As Cuvier could correctly describe a whole animal by the contemplation of a single bone, so the observer who has thoroughly understood one link in a series of incidents should be able to accurately state all the other ones, both before and after. We have not yet grasped the results which the reason alone can attain to. Problems may be solved in the study which have baffled all those who have sought a solution by the aid of their senses. To carry the art, however, to its highest pitch, it is necessary that the reasoner should be able to utilize all the facts which have come to his knowledge, and this implies, as you will readily see, the possession of all knowledge, which, even in these days of free education and encyclopedias, is a somewhat rare accomplishment; it is not so impossible, however, that a man should possess all knowledge which is likely to be useful to him in his work, and this I have endeavored to do."

> There are two ways this LaPlacian theoretical statement can be taken. One is to accept it as some near-metaphysical statement of a practically unattainable state of knowledge, then operationally disregard it when going about the pragmatic job of assembling and evaluating very imperfect "best available" information. The other is to accept the account, *à la* Holmes, as something surprisingly close at hand, at least for some kinds of knowledge, if we but attend. The first construction leaves actual knowledge to approach the unattainable ideal as near as claimants may establish by empirical evidence, accepting that it will often be not very close at all, given the limitations of the human condition. The second construction both accepts as likely the attainability of such knowledge, and accepts as strongly plausible those who make claims to it based on their personal magic over such observational data as they gather. The latter attitude we may refer to as pathological positivism. And it is clear from the quotation and from many others in the Sherlockian Canon, that it is this form of pathological positivism that is represented by both Holmes's attitudes and Holmes's performances. Holmes is the Wizard of inference. He does not practice science, he practices magic.[39]

The extreme positivist embrace of the potential availability of certain knowledge to the individual observer undertaking a sanctioned and disciplined method of observation appeals to many humans (vide astrology), although science was busily moving itself past this view into a frame of reference in which knowledge was viewed as fundamentally probabilistic and

39. D. Michael Risinger, *Boxes in Boxes: Julian Barnes, Conan Doyle, Sherlock Holmes and the Edalji Case*, INT'L COMMENT. ON EVIDENCE, Dec. 2006, at 1, 6-9, *available at* http://www.bepress.com/ice/vol4/iss2/art3 (internal footnotes omitted).

the enterprise of science fundamentally communal even as the popular embrace of heroic positivism reached its zenith in the early twentieth century. Heroic positivism continues to flourish today in the popular mind as a tenable account of the science. More to the point, it is clearly the dominant and romanticized lens through which forensic science is popularly viewed, as anyone familiar with Patricia Cornwell novels or various CSI programs can attest. And even more to the point, as I have previously indicated, it is at the core of the account forensic scientists in the pattern matching disciplines give about themselves. It is this heroic positivist paradigm that accounts for why fingerprint examiners believe that they can be justified in asserting that they can perfectly determine that a latent print found at a crime scene can be attributed to the ridged skin of a single individual to the exclusion of all other potential sources in the universe.

In a broadly cited article in *Science*,[40] Michael Saks and Jonathan Koehler spoke of "the coming paradigm shift" in forensic science. What they were envisioning was essentially the replacement of the outdated heroic positivist foundational views of forensic science with a set of views about the enterprise of science drawn from more modern approaches, which had displaced the heroic positivist paradigm in virtually every other area claiming to practice science in the modern sense. The NAS Committee Report is a very significant event fostering this shift. I retain enough of the old positivist faith in progress and historical directionality to believe that this is virtually inevitable in the long run. How long the run is going to be is unclear, however. As Thomas Kuhn, the originator of the notion of paradigm shift in the history of science, famously observed, it often takes a generation or more for a change to occur.[41] This is because the full shift must await the passing of the last generation personally invested in the rightness of the traditional paradigm, who, by weight of seniority, are in control positions within the structure of the enterprise and are thus situated to resist the new paradigm's full acceptance until they are no longer influential.[42] The transition period may see a lot of attempts at compromise and middle grounds, with people in control giving only as much as is made absolutely necessary by external circumstance or internal politics. It seems clear that the enterprise of forensic science is in this condition now. How long this

40. Michael J. Saks & Jonathan J. Koehler, *The Coming Paradigm Shift in Forensic Identification Science*, 309 SCI. 892 (2005).

41. THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970).

42. *Id.* at 122-30. Kuhn quotes Max Plank: "[A] new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it." *Id.* at 124.

fundamentally chaotic and unsatisfactory condition can last only time will tell.

One thing that is clear is that the fact that forensic science is lodged in the legal system, and serves the criminal justice branch of the legal system as its primary consumer, has not helped to foster the change. Having accepted forensic science on its old paradigmatic terms decades ago, the judiciary has been as resistant as the leaders of the forensic science community to embracing anything that would result in significant rejection of the current forensic science product, or call into question the wisdom of having accepted it for so long at face value. In addition, the intensely partisan nature of trials has often driven the testifying forensic scientist into extreme positions either at the behest of the prosecution or in response to what was regarded as unfairly critical cross examination (and anyone being cross examined is likely to regard the experience as unfairly critical). The result is that the legal system has reinforced the embrace of the traditional paradigm, both by appearing to bless it and by making its invocation of certain knowledge grounded on experience a welcome port in the storm of cross examination. As a long time observer of and participant in the legal process, I think I am qualified to offer the opinion, which appears to have been shared by the NAS Committee, that this has been a significant error.

Which returns us to the main focus of this paper—the notion of error and the issue of fault.

Some decades ago I got in trouble with the document examiner community over an article in the *University of Pennsylvania Law Review*.[43] The article began with a consideration of the classic early Renaissance treatise *Malleus Malificarum*, which dealt with how to determine whether someone is a witch.[44] The document examiners seemed to think that my co-authors and I were accusing them of witchcraft. But the point actually set out in the paper was that even very intelligent and rational people (which the authors of the treatise clearly were) could convince themselves of the reality of things based on weak evidence, things which better evidence might reveal to be inaccurate.[45] This was a general point about human tendencies, and was not particularly and specifically directed at document examination, or even forensic science, but at the necessity of being careful to require good evidence in formulating beliefs about the accuracy of expert practices

43. D. Michael Risinger, Mark P. Denbeaux & Michael J. Saks, *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise,"* 137 U. PA. L. REV. 731 (1989).

44. *Id.* (citing HEINRICH KRAMER & JAMES SPRENGER, MALLEUS MALLIFICARUM (Montague Summers trans., Dover Editions 1971) (1486)).

45. Risinger, Denbeaux & Saks, *supra* note 43, at 731-32.

in general. The real metaphor we invoked for document examination, and for traditional forensic identification disciplines in general, was in fact folk medicine, that is, areas of asserted expertise that testing would probably reveal to be accurate in regard to some claims but not others.[46] What we lamented in the first instance was the failure to generate proper and reliable information to provide validation, and to distinguish between the accurate practices and the inaccurate ones.

By "proper and reliable information" I do not necessarily mean the generation of formal data sets that yield proper quantified statistical information bearing on validity. Not every belief about the reliability of a process must be based on such data, and indeed, if we absolutely required such data, we should have to abandon many bases of inference that are well warranted by critical common sense. In this regard, I am not an acolyte of the great nineteenth century physicist Lord Kelvin, who claimed that only such formal quantified measurement could count as knowledge worth anything.[47] To Lord Kelvin, I say, behold this broken tea bowl. The perfect puzzle fit of the fracture, corroborated by the continuity of the design, would warrant beyond-reasonable-doubt confidence that the two were once part of a single bowl, without formal data on random match probabilities. It seems clear, at least in vitreous materials like glass and porcelain, that fracture involves random forces that generate so many points of unpredictable and difficult- or impossible-to-reproduce correspondence that the inference of common source approaches certainty very closely.

On the other hand, one must be extremely careful in making such claims. For instance, correspondences in regard to fracture or tear in other materials may be much less indicative of common source. And I can construct (in fact have published) a version of the vitreous fracture argument that works well in regard to the accuracy of common source inferences using rolled fingerprints, but much less well or not at all as one begins to deal

46. *Id.* at 734. This analogy was itself perhaps more inflammatory than necessary. We could just as well have invoked the journey of normal medicine from unvalidated to more validated techniques in relatively recent times. For a classic reflection on how much of this trip was only just begun in the twentieth century, see LEWIS THOMAS, THE YOUNGEST SCIENCE: NOTES OF A MEDICINE WATCHER (1983).

47.

> In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

1 BARON WILLIAM THOMSON KELVIN, *Electrical Units of Measurement*, POPULAR LECTURES AND ADDRESSES 73 (1883).

with the problems of small or smudged latent prints.[48] The burden of persuasive explanation should always be on those relying on such non-quantified claims to justify the reliability of expertise absent formal data. This is because humans can be subject to a lot of wishful thinking and exaggerated self-belief for which formal data is a great corrective, besides being valuable in itself. For this reason, reliability assessments based on formal data from well designed and appropriate studies are to be preferred, pursued, and cherished, even though it is often, perhaps usually, very difficult, expensive, and time-consuming to generate such data, requiring great care in study design and sensitivity to the fact that each sub-task in an area of claimed expertise must be identified and subjected to separate study. And as Jennifer Mnookin, Michael Saks, and I have pointed out repeatedly in various places,[49] such data does not necessarily have to be pointed toward the development of random match probability models, at least in the first instance. Properly designed "black box" studies of the success or failure of practitioners under different test conditions can yield useable data bearing on the reliability of expert results, which can fill the gap between no formal reliability data and the much more difficult task of generating DNA-like statistical systems.

Today I think it can be rightly said that the need for formal validation for many forensic science applications is well established and fairly generally recognized. After the issuance of the NAS Report, most of the leaders of the forensic science establishment have conceded the necessity of such validation, in order to quiet criticism if nothing else. Even while admitting this, however, there remain those who proclaim that such validation is a mere technical detail which will show that everything done in the name of forensic science heretofore has always yielded accurate results, at least when the results were derived by properly trained individuals adhering strictly to the procedures set out by the best authorities in the area, and all that is really required to eliminate error is laboratory accreditation and examiner certification. This has been a recurrent theme in the response of the

48. *See* Denbeaux & Risinger, *supra* note 21, at 66-74.

49. *See, e.g.*, Jennifer L. Mnookin, *Of Black Boxes, Machines and Experts: Problems in the Assessment of Legal and Scientific Validity*, 5 EPISTEME 343 (2008); Michael J. Saks, *Remediating Forensic Science*, 48 JURIMETRICS J. 119, 124 (2007); Michael J. Saks, *The Aftermath of* Daubert*: An Evolving Jurisprudence of Expert Evidence*, 40 JURIMETRICS J. 229, 239 (2000); D. Michael Risinger, *Preliminary Thoughts on a Functional Taxonomy of Expertise for the Post-*Kumho *World*, 31 SETON HALL L. REV. 508, 522 (2000); D. Michael Risinger & Michael J. Saks, *Science and Nonscience in the Courts:* Daubert *Meets Handwriting Identification Expertise*, 82 IOWA L. REV. 21, 40-41 (1996).

forensic community to the NAS Report.[50] In my opinion this is almost certainly the product either of a public relations instinct or wishful thinking or both. In the traditional forensic identification disciplines, as in folk medicine, careful testing of the claims is virtually certain to find areas of unreliable performance, just as there are likely to be many areas of reliable performance. The path forward from here to reliable knowledge of the accuracy of the various such disciplines in all their many claims will involve a lot more research than people have been willing to admit or contemplate up to now. The necessary studies will run at least into the hundreds over the course of many years, and will require the full and committed cooperation of the forensic science community. I believe that full and committed cooperation is an ethical obligation.

When a forensic examiner gives an opinion that a piece of evidence collected during an investigation "matches" or is "indistinguishable" from a known source, and a jury relies on that testimony, in whole or in part, in its decision to convict, and the conviction is later shown to have been factually wrong, the criminal justice system has definitely made an error. And the jury returned a verdict that was wrong. But has the forensic examiner made an error, or is such result merely attributable to the imperfect diagnosticity of the forensic discipline involved? This was the question that I noted earlier had been raised in passing in the NAS Report itself.[51] It sounds almost like a metaphysical question, but it has important real-world implications when it comes to determining the responsibility for errors resulting from forensic science testimony. To a great extent, whether to count this as an error of the forensic discipline, or merely an unfortunate result of imperfect diagnosticity, depends in large measure on what is known about the diagnosticity of the discipline, and what is claimed for it explicitly or by implication in front of the jury.[52]

50. *See, e.g.*, Joseph P. Bono, *President's Message*, ACAD. NEWS (American Academy of Forensic Sciences, Colorado Springs, C.O.), May 2010, at 1. I have heard Mr. Bono and others express themselves even more explicitly in this regard in presentations to various conferences.

51. *See supra* notes 27-28 and accompanying text.

52. Technically, test diagnosticity is made up of two components. In the presence of X, how often X is indicated (sensitivity), and when X is indicated, how often is X true (specificity). Since we are only considering situations in which X has been called (not cases of X where there has been no call), we are technically dealing only with the half of diagnosticity properly called specificity, but the word diagnosticity is more intuitively intelligible, so it has been used here for that reason. This is not really inaccurate, since we are dealing with an aspect of diagnosticity, just not the whole of it.

The word "diagnosticity" has become a common label for the predictive or discriminating power of tests or other signs. *See, e.g.*, Geeta Menon, Priya Raghubir & Norbert Schwarz, *Behavioral Frequency Judgments: An Accessibility-Diagnosticity Framework*, 22 J. CONSUMER RESEARCH 212 (1995). The word has been around a while, but apparently not

Forensic scientists at all levels should embrace the notion that testimony of imperfect but well-validated diagnosticity is much more helpful to the proper ends of the criminal justice system than testimony making unvalidated claims of certainty. The normal way in which ABO blood grouping testimony was presented to the jury in a typical case from a couple of decades ago illustrates clearly how expert testimony of low diagnosticity can be presented so as to make clear such relatively low diagnosticity of the information to the jury. The jury would have been told that the blood group of the defendant indeed "matched" that of the blood found at the scene, but the jury would also have been told the approximate statistical incidence of that group in the population. (For instance, if the defendant and the crime scene blood were both type O, rh factor negative, then the jury would be told that more than a third of the population of the United States [37.4%] shares that blood grouping.[53]) In such a case, if the jury overvalues the evidence, it is difficult to fault the expert information given, the expert who testified to it, or the party adducing it, or to call any resulting erroneous jury verdict an "error" of the expertise.

However, if testimony concerning a comparison process (say, bitemarks) asserts that the meaning of a match is that the defendant's teeth were the source of the bitemarks on the victim's skin (thus asserting directly or by clear implication perfect individuation, or a zero error rate), when in fact the process is much less diagnostic than a reasonable juror would conclude from such testimony, then any error resulting from juror reliance on the tes-

long enough to garner an entry in the Oxford English Dictionary, or Merriam-Webster's, or any other dictionary that I was able to track down, on-line or off. One would probably suspect that the word originated in medicine, since physicians are most associated with diagnosis, but that turns out to be almost certainly wrong. A search through the PubMed and JSTOR databases indicates that word seems to have first appeared (without explanation) in a 1959 article on North American prehistory. *See* Richard S. MacNeish, *A Speculative Framework of North American Prehistory as of April 1959*, 1 ANTHROPOLOGICA 7 (1959). Throughout the 1960s it appears mostly in articles on educational psychology. *See, e.g.*, Willard E. North & John Schmid, *A Comparison of Three Ways of Phrasing Likert Type Attitude Items*, 29 J. EXPERIMENTAL EDUC. 95 (1960); Roger W. Shuy, *The Relevance of Sociolinguistics for Language Teaching*, 3 TESOL Q. 13 (1969). By 1970 it was being used in Bayesian decision theory of the sort pioneered by Howard Raiffa. *See, e.g.*, Charles F. Gettys, David W. Martin, Leon H. Nawrocki & William C. Howell, *Human Evaluation of the Diagnosticity of Potential Experiments*, 83 J. EXPERIMENTAL PSYCHOLOGY 25 (1970). In its original decision theory meaning, diagnosticity was a measure of the extent to which a piece of information supported one hypothesis over a competing hypothesis. From there, the term was taken up more broadly by the measurement and testing literature.

53. Of course, there is an "error rate" for the test for blood groups that determines what counts as a match. (Here we would be concerned with false positives). However, those tests are well validated, and the error rates are small. And of course a false positive can be conceptualized not as an "error" but simply a result of the limit of the test's diagnosticity (specificity).

timony is properly counted as an error of the forensic discipline (and the practitioner).[54]

We can see from this that any comparison process can be purged of charges of error when testimony of inclusion is given if [and only if] the approximate statistical incidence of coincidental matches in some appropriate reference population is known, and communicated clearly to the jury. At that point, the diagnosticity of the process is known, and any resulting error is on the jury, or on God. So research into this question (which can be done as much through black box testing as through statistical modeling) is fundamental to knowing the risk of error in verdicts which is being imposed on the system in the absence of such knowledge through explicit or implied overclaiming. And cooperation with such research would seem to be both a professional and an ethical obligation of forensic practitioners at all levels.

We would all really like to have processes of diagnosticity approaching perfection. When the process said X was true, not X was never the case (specificity), and the process declared X in the presence of every X (sensitivity). We should work to improve diagnosticity for old processes, or to

54. *See* Bowers, *supra* note 24. For a particularly egregious example, see the performance of Dr. L. Thomas Johnson in the case of Robert Lee Stinson, documented in *Guilty, Said Bite Expert; Bogus, Says DNA*, CHI. TRIB., July 10, 2008, http://www.chicagotribune.com/news/chi-bite-mark-exonerationjul10,0,2835607.story. Stinson was finally released in January of 2009, and charges were finally dropped in mid-2010. *See Charges Dropped in Wisconsin Case*, INNOCENCE PROJECT BLOG (July 28, 2009, 5:17 PM), http://www.innocenceproject.org/Content/Charges_Dropped_in_Wisconsin_Case.php.

For results from a recent and ongoing program of research showing that the preconditions of secure identification from bitemarks are rarely present in the real world, see Mary A. Bush, Raymond G. Miller, Peter J. Bush & Robert B. J. Dorion, *Biomechanical Factors in Human Dermal Bitemarks in a Cadaver Model*, 54 J. FORENSIC SCI. 167 (2009); Raymond G. Miller, Peter J. Bush, Robert B. J. Dorion & Mary A. Bush, *Uniqueness of the Dentition as Impressed in Human Skin: A Cadaver Model*, 54 J. FORENSIC SCI. 909 (2009); Mary A. Bush, Kyle Thorsrud, Raymond G. Miller, Robert B. J. Dorion & Peter J. Bush, *The Response of Skin to Applied Stress: Investigation of Bitemark Distortion in a Cadaver Model*, 55 J. FORENSIC SCI. 71 (2010); Mary A. Bush, Howard I. Cooper & Robert B. J. Dorion, *Bitemark Profiling and Arbitrary Distortion Compensation Examined: Inquiry into Scientific Basis*, 55 J. FORENSIC SCI. 976 (2010); Mary A. Bush, Peter J. Bush & H. David Sheets, *Statistical Evidence for the Similarity of the Human Dentition*, 56 J. FORENSIC SCI. (forthcoming 2011) (on file with author); H. David Sheets, Peter J. Bush, Cynthia Brzozowski, Lillian A. Nawrocki, Phyllis Ho & Mary J. Bush, *Dental Shape Match Rates in Selected and Orthodontically Treated Populations in New York State: A Two Dimensional Study*, 56 J. FORENSIC SCI. (forthcoming 2011) (on file with author); Mary A. Bush, Peter J. Bush & H. David Sheets, *Similarity and Match Rates of the Human Dentition in Three Dimensions: Relevance to Bitemark Analysis*, INT'L J. LEGAL MED. (forthcoming 2011) (on file with author); H. David Sheets & Mary A. Bush, *Mathematical Matching of a Dentition to Bitemarks: Use and Evaluation of Affine Methods*, FORENSIC SCI. INT'L (forthcoming 2011) (on file with author).

invent or adopt new ones with improved diagnosticity. But failing that, we must do research on (and cooperate with research on) the actual diagnosticity of current process, most of whose diagnosticity in general, and even more importantly, as to each subtask, and under the conditions of real practice is not currently known. There are potential roadblocks for such research, including failure of forensic practitioner cooperation, failure to undertake a program of research on a sufficiently task-specific level, and the possibility that research designed to protect the status quo, faux research, if you will, will be allocated the bulk of research money and drown properly designed research according to a kind of Gresham's law.[55] A program of good research will take a long time to design and conduct, but it must be done. When we have developed the data on the actual diagnosticity of the forensic processes we use, the information given to juries about the results of those processes may be less apparently probative, but it will then be valid and, in the most important normative senses, error free.

55. *See* D. Michael Risinger & Michael J. Saks, *Rationality, Research and Leviathan: Law Enforcement Sponsored Research and the Criminal Process*, 2003 MICH. ST. L. REV. 1023, 1037-50; Risinger, *A Glass Nine-Tenths Full*, *supra* note 3, at 21 n.67; Risinger, *A Path Forward Fraught with Pitfalls*, *supra* note 3, at 239-40. There is reason to hope that the NAS Report has made funding agencies such as the National Institute of Justice more careful to select well-designed research for funding.