

Tulsa Law Review

Volume 45

Number 1 *Exhibiting Culture: Museums and
Indians*

Volume 45 | Number 1

Fall 2009

Context and Controversy: Why Questions of Validity and Reliability are Seldom Resolved in an Adversarial Setting

Mara L. Merlino

Victoria Springer

Follow this and additional works at: <https://digitalcommons.law.utulsa.edu/tlr>



Part of the [Law Commons](#)

Recommended Citation

Mara L. Merlino, & Victoria Springer, *Context and Controversy: Why Questions of Validity and Reliability are Seldom Resolved in an Adversarial Setting*, 45 Tulsa L. Rev. 133 (2013).

Available at: <https://digitalcommons.law.utulsa.edu/tlr/vol45/iss1/9>

This Article is brought to you for free and open access by TU Law Digital Commons. It has been accepted for inclusion in Tulsa Law Review by an authorized editor of TU Law Digital Commons. For more information, please contact megan-donald@utulsa.edu.

CONTEXT AND CONTROVERSY: WHY QUESTIONS OF VALIDITY AND RELIABILITY ARE SELDOM RESOLVED IN AN ADVERSARIAL SETTING

Mara L. Merlino* and Victoria Springer**

I. INTRODUCTION

Having read Dr. Cole's response to our article, *Meeting the Challenges of the Daubert Trilogy: Refining and Redefining the Reliability of Forensic Evidence*,¹ which was published in the Winter 2007 edition of the *Tulsa Law Review*, we find ourselves perplexed. The reason for this is not that he has written a response, but rather that he has interpreted so much of what we actually said quite incorrectly. It is difficult to say at exactly what point Dr. Cole's reading of the article became adversarial rather than objective. As a critic of latent print examination, he may have been predisposed from the start to disagree that forensic science practitioners have made any progress such as that described by our practitioner colleagues. Regardless, before we proceed, we would like to offer him a sincere apology for what we realize was a very poorly worded paragraph that could indeed be interpreted as a personal attack on his expertise, although it was not intended as such. Dr. Merlino accepts full responsibility for this language, which appears in the closing remarks of our article:

Cole sought to convince the court that latent print examination evidence should be excluded because it could not satisfy the requirements of *Daubert* However, Cole's own testimony did not meet these requirements. It is not enough to assert that the identification sciences are unreliable, biased, or invalid without any properly conducted research as a foundation for such assertions, and it is to Judge Hynes's [sic] credit that he held Cole's testimony to Cole's own high standards.²

In the following sections we will discuss in more detail the bases for this statement. At present, let us only say that we did not intend to imply that Dr. Cole's scholarship, while based on a literature review, was in any way lacking with respect to his testimony. The statement above might have been more empirically and appropriately worded:

[A]ccording to the judge, Cole's own testimony did not meet the requirements of Daubert.

* Assistant Professor of Psychology, Division of Behavioral and Social Sciences.

** Graduate Research Assistant, Grant Sawyer Center for Justice Studies, and Doctoral Student in the Interdisciplinary Ph.D. Program in Social Psychology.

1. Mara L. Merlino et al., *Meeting the Challenges of the Daubert Trilogy: Refining and Redefining the Reliability of Forensic Evidence*, 43 *Tulsa L. Rev.* 417 (2007).

2. *Id.* at 444.

For many judges it is not enough to assert that forensic testimony is unreliable, biased, or invalid without any properly conducted empirical research specifically addressing these issues. If, as the critics of forensic evidence argue, *Daubert* and its progeny are to be the touchstone for determining the admissibility of forensic expert testimony (which Dr. Cole stated was technical rather than scientific), then it is to the judge's credit that he also applied these high standards to Dr. Cole's testimony.³

This rewording is consistent with the stated purpose of our content analysis of published decisions, the actual context of our discussion of Dr. Cole's testimony, and our subsequent assertions, which were that *all* expert testimony should be held to these high standards, and that empirical research establishing the extent of the reliability of the forensic sciences is crucial.⁴ We believe that we are in complete accord with Dr. Cole and other critics of the forensic sciences on this point. However, Dr. Cole makes several assertions in his response with which we must disagree.

Below we will address some additional points Dr. Cole has attempted to make in his response. A substantial portion of Dr. Cole's response deals with things that we have not said and things that he inferred from arguments that we have not made.⁵ Some readers may view this as an attempt to create a "straw man" argument.⁶ However, we will take the less confrontational position that Dr. Cole's interpretation of our article may be due to his belief that we were attacking him as a scholar, and that Dr. Merlino may therefore be partially responsible for many of the misunderstandings we will discuss. By doing so we will attempt to remove our discussion from the adversarial context and return it to its original and intended context which, as we clearly stated in our Symposium article, was (1) a description of judicial decision-making in a subset of published latent print and forensic document examination cases; (2) a description of

3. Although New York is a *Frye* state, Judge Brennan, the attorneys, and Dr. Cole discussed the *Daubert* guidelines and the status of latent print examination as a science, and the the defense argued that the Court should reject *Frye* in favor of a more "expansive and liberal" (e.g., *Daubert*-type) review in making decisions about what ought to be the subject of expert testimony. See *People v. Hyatt*, No. 50115U (N.Y. Sup. Ct. Oct. 10, 2001).

4. Merlino et al., *supra* n. 1, at 444.

5. Simon Cole, *Don't Shoot the Messenger by One of the Messengers: A Response to Merlino et al.*, 45 *Tulsa L. Rev.* 111 (2009). "[U]se of my testimony by Merlino et al. as empirical evidence supporting the reliability of latent print identification . . . "Nor can it serve, as Merlino et al. seem to claim, as an explanation for the admissibility of latent print testimony in most U.S. criminal cases." *Id.* at 117. "This analysis, it is claimed, explains why 'the majority of [admissibility] challenges to [fingerprint] evidence have been unsuccessful.'" *Id.* "Merlino et al., recall, promised an explanation as to why admissibility challenges to latent print identification almost always fail . . ." *Id.* at 121. "Merlino et al. purport to explain the admissibility of latent print evidence by reference to a case that was not about the admissibility of latent print evidence." Cole, *supra* n. 5, at 122. "In their contribution to the Symposium Issue, they invoke the one case to explain the outcome of the other thirty-eight." *Id.* "One possibility is that they believe there is some principle of law served by precluding rebuttal witnesses to certain forms of expert evidence under certain circumstance. But Merlino et al. articulate no such principle." *Id.* at 123. "[T]he argument appears to be that if rebuttal witnesses to a form of proffered evidence are lacking in reliability, then the evidence itself must be reliable." *Id.* at 124. "Merlino et al.'s hanging of the reliability of latent print identification on the supposed unreliability of its supposed critics . . ." *Id.* "[W]hereas Merlino et al., though they do not explicitly say so, appear to take the opposite view." Cole, *supra* n. 5, at 124. "In the interest of completeness I will discuss these items, even though Merlino et al. do not explicitly state that they support the reliability of latent print identification." *Id.* at 126. "Merlino et al. do not explicitly state how it is they think the reliability of latent print identification may be inferred from anatomical research. . . ." *Id.* at 128.

6. For a definition and examples of "straw man arguments," see Wikipedia, *Straw Man*, http://en.wikipedia.org/wiki/Straw_man (last accessed Sept. 14, 2009).

steps taken thus far by practitioners in the field in order to meet the challenges of *Daubert* and its progeny; and (3) a discussion of how the tenets of the sociology of scientific knowledge are demonstrated in the discourse surrounding the construction of the evidentiary reliability and the admissibility of forensic expert testimony. We believe that it is important at this point to stress to the reader that these were the stated purposes of our article, as Dr. Cole has criticized us for not addressing certain issues which not only were beyond the scope of these purposes, but also are not germane given his mistaken interpretation of the claims we make.

Internal Inconsistency in the Tulsa Law Review Symposium

Dr. Cole states that presenting conflicting positions destroys the “internal consistency” of the Tulsa Law Review Symposium in which both our articles appeared.⁷ We are unsure why he believes that it is inappropriate to present different points of view in such a symposium. Indeed, by definition, presenting multiple perspectives is the very purpose of a symposium—but in our view the *failure* to present different sides of the issue in such a forum would have indicated lack of objectivity on the part of the journal.

There can be no real discussion when only one side of an issue is presented, and doing so falsely implies that consensus exists in an area that is, in fact, hotly debated. The majority of contributors to the Symposium may believe that there are problems with the reliability and validity of the various forensic fields, or that there are biases inherent in the current practices of judges, attorneys, or practitioners. In reality, there are many academicians and practitioners who, based on the current lack of empirical reliability or biased data, remain unconvinced that these fields are so fatally flawed that gross miscarriages of justice frequently occur. We must stress that we are not discounting the studies that do exist. However, as Dr. Cole has testified on a number of occasions, studies of the reliability and validity of many areas of forensic practice have yet to be conducted, and these questions therefore remain largely unanswered from a scientist’s perspective. We commend the editors for giving the practitioners a voice in this Symposium so that their ongoing efforts to meet the challenges posed by *Daubert* and its progeny might be recognized.

II. REFINING AND REDEFINING RELIABILITY

Dr. Cole mistakenly states that we interpret *Daubert* as refining and redefining what is meant by “reliability.”⁸ We are unclear why he may have felt that this is the case,

7. Cole *supra* n. 5, at 114–115.

8. *Id.* at 115 n. 26.

I must admit that I find Merlino et al.’s interpretation of *Daubert* as “refining and redefining . . . reliability” puzzling. “Reliability”—or, as the *Daubert* Court surely meant, “accuracy” or “validity”—are scientific concepts that cannot necessarily be “redefined” by a court. As I understand it, *Daubert* changed the admissibility threshold to require evidence of reliability, rather than mere relevance [as under the Federal Rules of Evidence] or “general acceptance” in a field [as under *Frye v. U.S.*]. *Daubert* did not “redefine” the notion of reliability; it simply required that reliability—defined in the same way it always has been defined—be shown for evidence to be admissible. The notion that *Daubert* “redefine[ed]” reliability suggests that latent print identification may have satisfied some prior notion of “reliability” only to have the Court “redefine” reliability. It suggests that the absence of evidence of reliability was caused by shifting legal definitions, rather

since our article contains no language to this effect. Part I of our article states:

Part III presents a discussion of the impact of *Daubert* on the field of forensic document examination from the point of view of two experts, and the steps taken by forensic document examiners to meet the requirements of the *Daubert* trilogy. Part IV presents a discussion of the impact of *Daubert* on the field of latent fingerprint examination and the steps taken by latent fingerprint examiners in response to the *Daubert* trilogy, again from the perspective of two experts.⁹

In subsequent sections of the article, the practitioners describe trends within their fields toward standardization of training and testing, proficiency testing, certification, better understanding of procedures and methodology, and willingness of members of their professions to participate in empirical research conducted by unbiased researchers. The final three paragraphs of the article quite clearly summarize our position on the current state of knowledge, the need for continued research, and the importance of critical thinking to moving forward as disciplines. Given what we actually do say, it would seem evident that “refining and redefining the reliability of forensic evidence”¹⁰ refers to the efforts of the practitioners to improve their fields, and not to efforts to redefine the meaning of a scientific concept.

III. ENGAGING THE SCHOLARLY LITERATURE

Dr. Cole additionally argues that we have not sufficiently engaged the scholarly literature in the area of latent print examination and provides a list of references for what he believes should have been included. As we stated in the forensic document examination section of our article, an extensive discussion of the arguments of the critics was beyond the scope of this article. In this section, we did in fact provide a list of references so that interested readers could engage in those discussions if they chose. We did not engage in such a discussion in the latent print examination section for the same reason, but failed to provide the same list of references. We thank Dr. Cole for making this list available to interested readers.

He correctly points out that we did describe the findings of several empirical studies of latent print examination. However, citing and rebutting the arguments contained in the scholarly works he lists does not provide the kind of evidence needed to answer questions of reliability. In *People v. Hyatt*, Dr. Cole states that

the general agreement [on the definition of science] seems to be around the idea that science consists of testing of statements that have been tested in some way through experiment or some other way rather than simply asserted and assumed or believed because somebody, an authority, says that they are true, that your statements must be tested.¹¹

than by the simple omission of empirical data gathering. In fact, establishing the reliability of latent print identification has always simply required empirical evidence concerning the extent to which latent print examiners produce correct results, regardless of what the courts say.

Id.

9. Merlino et al., *supra* n. 1, at 418–419.

10. *Id.*

11. Tr. Transcr. at 9–10, *Hyatt*, No. 50115U (available at http://www.onin.com/fp/ny_v_hyatt_simon_cole_testimony_4oct01.pdf).

However, with due respect to the scholars whose work he cites, in his response Dr. Cole seems to assert the opposite—that scholarly works that do not offer empirical, systematically collected data of the kind that would either support or refute the reliability of latent print analysis, constitute the truth. Scholars on *both sides* of the issue offer thoughtful and meaningful statements of their positions. Unfortunately, these opposing arguments have been made in adversarial circumstances that are not conducive to achieving the environment of trust and cooperation needed to discover where the truth really lies.¹²

Dispute is important and raises many valid questions, but dispute alone does not constitute empirical data. For instance, Dr. Cole states that the argument that observations made during case work can be used to test the reliability of latent print examinations is untenable “because the ground truth is not known in casework.”¹³ We can take the opposite position and argue that while the broad premise of this argument may be true, Dr. Cole’s specific conclusion is founded on an oversimplification of the kinds of data that may be rendered through observational processes in the sciences.

Empirical data is derived from systematic observation, and such observations can be made using a variety of research methods. Data collected by latent print examiners (and forensic document examiners) are not experimental and should not be characterized as such. Indeed, it is not possible to both understand the inner workings of the forensic examinations as they are applied in the field and at the same time impose experimental controls on those examinations, because experimental controls reduce the ecological validity of the actual circumstances in which examiners function.¹⁴ Experimental procedures might reasonably be used to study specific features of the process of forensic examination by isolating them and controlling for the impact of other factors that may influence the outcome of the examination, but ultimately the isolated factors must be understood within the whole context of the examination environment. Thus, experimental research to identify possible issues followed by empirically rigorous and scientifically accepted naturalistic observation methods holds the greatest potential for evaluating the performance of examiners without altering the very process under observation.¹⁵

The designation of research as “naturalistic” or “non-experimental” in no way compromises the strength of the systematic scientific design. Systematic observation means that the features of the phenomenon to be observed are specified in advance so that multiple observers are able to use these methods and arrive at similar and unbiased conclusions. Methods analogous to those used by latent print examiners are used in many different fields of expertise to systematically gather and interpret data from images, published works, bodies of literature, and video and audio recordings. Linguists can use content analysis to reliably attribute the works of unknown writers to known writers, art historians can attribute the work of unknown artists to known artists, and indeed, content

12. See Tamara F. Lawson, *Can Fingerprints Lie? Re-weighing Fingerprint Evidence in Criminal Jury Trials*, 31 Am. J. Crim. L. 1, 37 (2003).

13. Cole, *supra* n. 5, at 127.

14. See Earl Babbie, *The Basics of Social Research* (4th ed., Thomson Wadsworth 2007). This alteration of behavior is referred to as “reactivity,” and is a source of bias in research.

15. *Id.*

analysis of data sources such as video and audio recordings and published print media is one of the primary methods used in venue research.¹⁶

The conscientious application of these observational methods has led to the publication of lauded empirical works and the discovery of historically verified truths, such as the recent identification of the woman who was the subject of Leonardo da Vinci's enigmatic "Mona Lisa." The identity of the woman could not be "known" before the investigation, and yet the methods used to investigate her identity have yielded a verifiable result. A similar story unfolds from the scientific investigation of the mysterious death of King Tutankhamun, who is now believed to have died from an infection rather than murder. Again, his cause of death cannot be "known" *a priori* to the generations of Egyptologists and other scientists who have dedicated decades to the study of the boy pharaoh. Despite Dr. Cole's arguments, we believe the pursuit of scientific answers in all fields of inquiry will continue to produce tenable results in a world replete with unknowns.

Here we *do not* argue that the use of non-experimental methods (e.g., observational designs such as content analysis and historical analysis) in other disciplines is proof of reliability in latent print examination, but we do suggest that comparing known sources to unknown sources without knowing the ground truth does not necessarily invalidate the technique. These may be opposing arguments, but neither truly resolves the issue. This can only be accomplished by actually conducting the research.

IV. IMPROPER USE OF *PEOPLE V. HYATT*

Many of the issues Dr. Cole raises center on our discussion of the transcript of the *Frye* hearing in *People v. Hyatt*. Dr. Cole states in Part II of his response:

I will discuss use of my testimony by Merlino et al. as empirical evidence supporting the reliability of latent print identification. I will show that the preclusion of my testimony cannot serve as evidence of the reliability of latent print identification. Nor can it serve, as Merlino et al. seem to claim, as an explanation for the admissibility of latent print testimony in most U.S. criminal cases.¹⁷

Dr. Cole's interpretation of our discussion of the tenets of the sociology of science, which are demonstrated in this transcript, is both inaccurate and perplexing. First, our use of Dr. Cole's testimony was neither intended nor used as empirical support for the reliability of latent print identification. Neither did we claim that it was an "explanation" for why challenges to such evidence fail. In the paragraphs preceding our discussion of *Hyatt*, we touched on the tenets of the sociology of scientific knowledge as described by Sheila Jasanoff,¹⁸ and argued that the reciprocal influences of judge, attorney, and expert

16. *Id.* (Among other purposes, content analysis has been used to establish authorship for purposes of securing political and military intelligence, to provide legal and evaluative evidence, and to relate known characteristics of sources to the messages they produce.); see also Ole R. Holsti, *Content Analysis*, in *The Handbook of Social Psychology* 596 (Gardner Lindzey & Elliot Aronson eds., 2d ed., Addison-Wesley Publ. Co. 1969).

17. Cole, *supra* n. 5, at 117.

18. Sheila Jasanoff, *What Judges Should Know about the Sociology of Science*, 77 *Judicature* 77 (1993). Dr. Cole also takes issue with our use of this article, which he believes is inappropriate because the article was written to inform judges about the social construction of scientific facts. Dr. Cole's explanation of the concepts of experimenters' regress and boundary work in his reply to our Symposium article is useful and informative.

must all be considered when trying to understand how and why admissibility decisions are made. We prefaced our discussion of *Hyatt* by stating:

[A]lthough the falsifiability, error rate, existence or maintenance of standards controlling the technique's operation, peer review and publication, and general acceptance of forensic expert testimony have been successfully challenged on some occasions, the majority of challenges to this type of evidence have been unsuccessful.¹⁹ Critical examination of published decisions and transcripts of these cases offers some *insights* into why this is the case . . . [W]e offer as *one example* the exclusion of the defense's proffer of the testimony of Dr. Simon Cole in *People v. Hyatt* . . . Both boundary work and experimenters' regress are clearly evident in the transcript of this pre-trial hearing.²⁰

We then very briefly described the *prosecution's* use of boundary work (describing Dr. Cole as "simply a historian" who lacked qualifications as a latent print examiner and had minimal knowledge of latent print examination), and experimenters' regress (questioning Dr. Cole about his sources of information to demonstrate that Dr. Cole's data were not obtained using the scientific method), and quoted those portions of Judge Brennan's decision that outlined his reasons for the exclusion of Dr. Cole's testimony.

Given the context of our discussion of this case (i.e., the use of boundary work and experimenters' regress by Dr. Cole, the attorneys, and Judge Brennan), we can only conclude that Dr. Cole's perception that he was under attack was the basis for his assertions that we were inappropriately attempting to use *Hyatt* to demonstrate the reliability of latent print examination, or that we were using this case to explain why latent print testimony is admissible in most U.S. cases. We assure him that we had no such agenda when we selected *Hyatt* as an example of the use of boundary work and experimenters' regress. Our reason for choosing the *Hyatt* transcript was its extended discussions in direct and cross examination of both *Frye* and *Daubert* issues. We might just as easily have selected any of the several forensic document examination or latent print case transcripts, or any of the published judicial decisions we had on file, and still accomplished our intended and stated purpose.²¹

V. RESEARCH METHODS AND INTERPRETING DESCRIPTIVE AND EXPLANATORY DATA

Dr. Cole's critique of the data presented in Part II of our Symposium paper (hereafter the Symposium Data Analysis) is misguided for several reasons. Dr. Cole engages in a great deal of speculation about our intended use of these data, most of which is simply incorrect given the purpose of the content analysis, our criteria for

However, we disagree with Dr. Cole's argument that the article's intended audience invalidates the application of Jasanoff's discussion as to what occurs during the process of an evidentiary hearing. The processes Jasanoff describes can be used to generate testable hypotheses, and Jasanoff's work has been demonstrated to be a viable perspective for studying the social construction of the admissibility of expert testimony. See Mara L. Merlino et al., *Judicial Gatekeeping and the Social Construction of the Admissibility of Expert Testimony*, 26 *Behav. Sci. & L.* 187 (2008).

19. Merlino et al., *supra* n. 1, at 443. This assertion was based on the empirical analysis of codeable published decisions presented in Part II of our paper.

20. *Id.* (citations omitted).

21. Our introduction states, "Part V examines the empirical data and the discussion by forensic professionals in the context of the sociology of science, and discusses how the tenets of this sociological perspective are demonstrated in the discourse surrounding the social construction of evidentiary reliability and the admissibility of forensic expert testimony." See *id.* at 419.

selecting codeable cases, the operational definitions of our variables, and the manner in which the results are reported. Space constraints in our Symposium article precluded not only a comprehensive discussion of the case law and extensive discussions of the claims made in the scholarly literature, but also a lengthy description of the extant data on judicial use of boundary work and experimenters' regress and the theoretical underpinnings and methodology of our content analysis. We can best address Dr. Cole's critique by providing a very brief theoretical underpinning for our study and explaining our research methodology and the bases for our data analyses in greater detail.²²

According to Solomon and Hackett, studying the intersection of science and law, where scientists and legal professionals engage in interdisciplinary discourse, is important when attempting to understand the construction of science in law.²³ They state that work there "is generally visible, formal, recorded, and highly consequential for substantive outcomes, procedural and substantive precedents, and even the public understanding of science."²⁴ By examining published opinions in which judges describe their decision-making processes, it is possible to begin to understand how they use boundary work and experimenters' regress (the processes by which scientific facts are constructed) in the context of litigation to arrive at their admissibility decisions.

Jasanoff briefly articulates the mechanisms for creating scientific "facts" in her 1993 *Judicature* article, pointing out that a phenomenon "out there in nature" passes through a series of steps before it can be considered a scientific fact.²⁵ Scientific facts depend upon certain background features that are necessary for their production. Scientific claims are "contingent" on experimental or interpretive conventions that have been agreed upon within relevant scientific communities (e.g., if $\alpha < .05$, then the findings are statistically significant). These contingencies become visible when controversy erupts. Jasanoff wrote that the social dimension of science (e.g., practices, conventions, institutions, and interests that sustain scientific progress and legitimate particular scientific facts) is precisely what is at issue in most legal proceedings involving proffers of scientific evidence.²⁶ Indeed, Jasanoff points out that the deconstruction of evidence is limited only by the extent to which attorneys can find issues to raise either for or against its admissibility.²⁷ Consequently, when challenging the admissibility of the opposing party's expert testimony, attorneys may deconstruct testimony using a "contingent" discourse repertoire that stresses the indeterminacy of the opposing expert's facts, while adopting an objective, "empiricist" discourse repertoire when making their own proffers of expert testimony. *Hyatt* is an excellent example of how this process occurs.

When scientific knowledge crosses the boundary from science to legal science, an

22. We note here that this information was available to readers upon request, but Dr. Cole did not request this information. Dr. Cole requested by email the list of cases we used in our analysis, and in email exchanges dated Sept. 12 and Sept. 16, 2008, we also offered to forward our coding materials and protocol as well, and to answer any additional questions. He declined these offers.

23. M. Solomon & Edward J. Hackett, *Setting Boundaries Between Science and Law: Lessons Learned from Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 21 *Sci., Tech. & Human Values* 131 (1996).

24. *Id.* at 144.

25. Jasanoff, *supra* n. 18, at 77–78.

26. *Id.* at 77.

27. *Id.* at 78.

additional set of contingencies is introduced through the process of evaluating its admissibility. These contingencies include the legal authorities (e.g., statutes and precedent) governing the admissibility of expert testimony. Thus, admissible expert testimony must in every case reach a threshold that includes the relevance, materiality, probativeness, and reliability before the evidence can be heard by triers of fact (e.g., judges or juries). It is up to trial judges to determine whether or not proffers of expert testimony meet these additional contingencies.²⁸ However, Gatowski and colleagues found that judges had difficulty operationalizing such scientific concepts as falsifiability and error rate, and a number of judges in their survey responded to survey items by addressing these concepts from a legal context rather than a scientific one.²⁹ The degree and nature of judicial deconstruction can be used to investigate the relationship between expert testimony in its proffered form and its mediated form. The mediated form follows the boundary work and experimenters' regress that occurs and the legal contingencies that are applied during its movement from scientific or technical fact to legal scientific or technical fact.

The Symposium Data Analysis was an extension of research that has been published elsewhere. The purpose of the earlier research was to understand how judges use expert characteristics (boundary work) and evidence characteristics (experimenters' regress) to arrive at and explain their own admissibility decisions. Among the questions we sought to answer were whether judges would more closely scrutinize the qualifications of the experts following *Daubert*; whether judges' evaluations of the merits of the evidence would employ constructs and processes similar to those used by scientists; whether the qualitative or quantitative basis of the evidence was related to the extent of judges' evaluations of the merits of the evidence; whether or not judges

28. See Lawson, *supra* n. 11, at 37 (for a cogent discussion of judicial gatekeeping, the distinctions between legal and scientific reliability, and the admissibility of latent print examination).

29. Sophia I. Gatowski et al., *Asking the Gatekeepers: Results of a National Survey of Judges on Judging Expert Evidence in a Post-Daubert World*, 25 L. & Human Beh. 433. Dr. Merlino notes that the findings of this survey are often somewhat misunderstood. We reported:

From the answers that were provided, the researchers could only infer a true understanding of the scientific meaning of falsifiability in 6% (n = 23 of 400) of the judges' responses. In fact, for the 352 judges who indicated that falsifiability was a useful criterion, the coders could only infer a true understanding of the concept in 4% of the responses elicited. Responses such as, "I would want to know to what extent the theory has been properly and sufficiently tested and whether or not there has been research that has attempted to prove the theory to be wrong" and "if it is not possible to test the evidence then it would weigh heavily with me in my decision" are illustrative of responses coded as 'judge understands concept'. An example of a response coded as "questionable" would be "I would want to know if the theory has been tested" without further articulation of what this means.

Id. at 445. Our coding guidelines were quite specific concerning what language demonstrated a clear understanding of the *scientific meaning* of this term. However, the language used by many of the 59% of judges whose understanding was designated "questionable" suggests that the judges actually may have a better understanding of falsifiability than we were able to indicate. A better measure of judges' understanding may actually be the 35% (n = 140) whose responses indicated that they clearly did *not* understand the *scientific meaning* of falsifiability. Interestingly, judges' understanding of the term "reliability" indicated that many judges evaluate the proffered evidence for its trustworthiness or credibility rather than for its reliability in the scientific sense. Many of these judges had what could be considered a more "legal" understanding of falsifiability (e.g., falsify is synonymous with counterfeit). Similarly, when judges were asked about error rate, very few of their responses demonstrated a clear understanding of Type I or Type II error. Although we have taken note of them, we have not systematically evaluated the impact of these subtle differences in the paralogues of the two disciplines, although we have discussed the importance of clear communication among different disciplines elsewhere. See Merlino et al., *supra* n. 1, at 443.

differentially applied the *Daubert* guidelines to different kinds of evidence; and whether the increased scrutiny of the evidence was related to its exclusion.³⁰

To obtain the data for the Symposium Data Analysis, we analyzed cases published between July 1, 1993 and October 31, 2007 in which the admissibility of testimony concerning forensic document examination and latent print expert testimony was at issue.³¹ Codeable cases were identified by searching the Lexis database of state and federal cases using the search strings “expert testimony w/p latent print and admit! or admit!” and “expert testimony w/p forensic document and admit! or admit!” The resulting cases were each evaluated for codeability using guidelines adapted from those developed by Dixon and Gill.³² Codeable cases contained a substantive discussion of the admissibility of proffered expert testimony that included the rule(s) of evidence relevant to the analysis, and a discussion of how the evidence met or failed to meet the criteria for admissibility. Challenges to admissibility were substantive (e.g. related to the characteristics of the experts or the evidence), rather than procedural challenges in which the attorneys objected to the timeliness of the expert’s report or other statutory issues.³³ Cases in which no proffer of evidence was made (e.g. a party claims that a decision should be overturned because an attorney failed to proffer expert testimony) were excluded because there was no evaluation of the evidence.³⁴ This evaluation of cases identified 30 forensic document examination cases containing 37 codeable proffers of evidence, and 35 latent print cases containing 39 codeable proffers of evidence.

Several categories of information were abstracted from the cases: (1) case characteristics, or information about the location and stage of the proceeding, the charges involved, and the characteristics of the judge deciding the case; (2) review characteristics, or information about the Rules of Evidence, the application of precedent, and the justifications for conducting the review; and (3) reliability characteristics, or information about the judges’ evaluations of the characteristics of the expert, and the characteristics of the testimony (e.g. the actual factors judges evaluated during their evidentiary review). Variables included: (1) the number of expert characteristics (boundary work factors) evaluated by the judge; (2) the number of evidence characteristics (experimenters’ regress factors) evaluated by the judge; and (3) the admissibility of the evidence.

Expert characteristic (boundary work) and evidence characteristic (experimenters’ regress) factors were empirically identified from the content of the cases. Judges

30. A discussion of that research is beyond the scope of this response, but we invite interested readers to refer to this research for the theoretical background, a discussion of the research on judicial decision making post-*Daubert*, and findings concerning the social construction of the admissibility of psychological/psychiatric, toxicology, and damages evidence. See Merlino et al., *supra* n. 1, at 422.

31. Codeable cases for the earlier study were identified by searching the Lexis database of U.S. District Court cases using the search string “expert testimony w/p admit! or admit!”, specifying two-year increments (e.g., 07/01/1993 to 06/30/1995, 07/31/1995 to 06/30/1997). This search identified 3,260 U.S. district court civil (3060) and criminal cases (200) in which some discussion of admissibility appeared within the same paragraph as the words “expert testimony.”

32. Lloyd Dixon & Brian Gill, *Changes in the Standards for Admitting Expert Evidence in Federal Civil Cases since the Daubert Decision*, 8 *Psychol., Pub. Policy & L.* 251, 256–257, 263 (2002).

33. *Id.*; see also Jennifer Groscup et al., *The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases*, 8 *Psychol., Pub. Policy & L.* 339, 343 (2002).

34. *Id.*

discussed the experts' education, training outside academia, experience, skill or subject matter knowledge, publication record, and reputation when considering expert characteristics. These factors were therefore designated boundary work factors. When considering the characteristics of the evidence, judges discussed the *Daubert* guidelines, the basis of the testimony, statistical significance, any court-appointed neutral expert's evaluation of the testimony, the purpose for which the research was conducted, and the clarity and coherence of the expert's explanation. These factors are considered experimenters' regress factors. The admissibility of the evidence was a two category (admissible/inadmissible) nominal-level variable. These categories were created by collapsing several levels of this variable (e.g., the admissible category included the levels admissible/properly admitted/improperly excluded; the inadmissible category included the levels inadmissible/properly excluded/improperly admitted).

Our unit of analysis was the opinion, and the unit of observation was a proffer of evidence within the opinion. The coding scheme was based on that developed by Dixon and Gill, whose code book also contained some variables that were common to studies by Krafska et al., Merlino et al., and Richardson et al. Additional codes for criminal cases were developed as the coding proceeded. The codebook was revised throughout the coding process as the need for additional codes arose.³⁵ Inter-coder reliability was increased through the use of a check-coding and check-code verification protocol. This sampling methodology has been used by researchers in a variety of post-*Daubert* studies and the research design and methodology are commonly used in to quantify qualitative data.³⁶

We will not discuss our findings, which are available to readers in our Symposium article. However, we must clarify for readers who may be unfamiliar with research methodology how researchers are able to draw conclusions from data, depending on the nature of the research design. There are several different purposes for conducting empirical research: (1) exploration; (2) description; and (3) explanation. These three purposes are related, but distinct with respect to their goals and what can be asserted on the basis of the data resulting from each kind of study. Exploratory studies are generally conducted when a problem has not been clearly defined. In this type of study, the goal is to gain a greater understanding of the phenomenon. These research designs are small-scale and usually involve small numbers of participants or observations, and the research results are normally not appropriate for generalizing to the larger population.

Descriptive studies are those that provide information about *what* is occurring. The U.S. Census, national employment statistics, and FBI crime statistics are all examples of descriptive research. These studies tell us *how many* people live in the U.S., *what kinds* of jobs exist within our economy, or *who* occupies cells on death row in our prisons. Descriptive studies provide evidence that a phenomenon does or does not occur, but are

35. Merlino et al., *supra* n. 17; see also Jennifer Groscup et al., *supra* n. 27; Shirley Dobbin, *The Social Production of Rape Trauma Syndrome: The Making and Management of a Syndrome Claim* (unpublished doctoral dissertation, U. Nev., Reno, 1998) (copy on file with the U. Nev., Reno); James Richardson et al., *A Case Law Survey of Social and Behavioral Science Evidence after Daubert* (Redondo Beach, California, June 1998) (copy on file with Author); Carol Krafska et al., *Judge and Attorney Experiences, Practices, and Concerns Regarding Expert Testimony in Federal Civil Trials*, 8 *Psychol., Pub. Policy & L.* 309 (2002).

36. See Babbie, *supra* n. 14.

unable to address *why* this is the case. This is the purpose of explanatory studies, which are conducted to test and understand causal relationships among variables. This research involves theoretical explanations, experimental design, and the generation of testable hypotheses.

It is clear, given the presentation of the results of the Symposium Data Analysis, that our data are descriptive, but Dr. Cole incorrectly characterizes our study as explanatory.³⁷ We neither posited any causal relationships, nor tested any hypotheses. Further, we did not offer any causal explanations for the outcomes of these judicial decisions. Our data are reported purely descriptively, and we specifically stated several times that our findings pertain *only* to our sample of cases.³⁸ We did not claim, as Dr. Cole states, that our data *explain* why “the majority of [admissibility] challenges to [fingerprint] evidence have been unsuccessful.”³⁹ Nor did we indicate in any way that these data were intended to support the reliability of latent print or forensic document examination. We simply described statistically the characteristics of the cases in our sample.

VI. CONCLUSION

Dr. Cole has taken great care to craft a response to our Symposium article that clearly articulates his position concerning the reliability of latent print examination. Unfortunately, he based his response, either intentionally or unintentionally, on claims that were never made, positions that were never stated, and assumptions that are clearly inconsistent with the stated purposes of our article. In our response to Dr. Cole, we have attempted to clarify for the reader exactly what was stated in our Symposium article and the theoretical and methodological underpinnings of the data we presented. We also point out a number of inconsistencies between what we actually wrote and Dr. Cole’s interpretation of this information. In doing so, we not only accept some responsibility for Dr. Cole’s misunderstanding, but also take great care to try to avoid perpetuating a pejorative exchange that serves neither side of this debate.

It remains our hope that judges, attorneys, academic researchers, and forensic practitioners can find a way to join together in an effort to set aside current biases and search for a true understanding of the forensic sciences using the principles of science. As we stated in the closing remarks of our Symposium article, it is not enough for practitioners to assert that current practice is the best practice without empirical support. Nor should the critics who have raised questions about the reliability of the forensic fields assert that they are fatally flawed without well-conducted empirical research that demonstrates whether, and to what extent, this is the case. This was the intended message

37. Cole, *supra* n. 5, at 121. Dr. Cole states, “Merlino et al. . . . promised an explanation as to why admissibility challenges to latent print identification almost always fail, an explanation derived from an empirical analysis of “case law.” *Id.*

38. See Merlino et al., *supra* n. 1, at 420; *Id.* at 421–422. (“Judges in *this sample* mentioned expert qualifications Judges in *this sample* mentioned only eight unfavorable characteristics The data available in *this sample* suggest that judges differentially focus” (emphasis added)).

39. *Id.* at 443. Our original article made no reference to either causal relationships or fingerprint analysis, and only suggested that reading the discourse about admissibility in the published materials offers some insights into why challenges to various types of evidence have been unsuccessful.

of our Symposium article. It is our sincere hope that proponents of both sides of the issue are able to step forward from the adversarial context of legal debate and into a new context of discovery, where collaboration, cooperation, communication, and critical thinking reside, and where true knowledge might really be achieved.

