Forensic Science and Wrongful Convictions: From Exposer to Contributor to Corrector

SIMON A. COLE

INTRODUCTION

Brandon Garrett’s book, *Convicting the Innocent*, makes a number of important contributions to the scholarly and public discourse on miscarriages of justice. In this essay, I will focus on the contribution that is most related to my own research interests: its contribution to our understanding of the relationship between forensic science and miscarriages of justice. I will first endeavor to place Garrett’s contribution in historical context by briefly tracing the history of discussions about forensic science and wrongful convictions. I will then highlight in what way Garrett’s work has furthered our understanding. I will then discuss some of the criticisms of Garrett’s work by advocates of forensic science and try to explain how data limitations contribute to the difference of opinion between Garrett and his critics. I will conclude by suggesting a different, more theoretically grounded way of conceptualizing miscarriages of justice that might help us move beyond these differences of opinion. Ultimately, however, my suggestions will be highly speculative: data limitations, again, will make it difficult to make any strong empirical inferences about the relationship between forensic science and wrongful convictions.

* Associate Professor of Criminology, Law and Society, University of California, Irvine; Ph.D. (Science and Technology Studies), Cornell University; A.B., Princeton University. I am grateful to Joëlle Vuille and Norah Rudin for helpful comments on drafts of this Article. This material is partially based upon work supported by the National Science Foundation under grant No. SES-0115305. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.
I. Early Studies of Wrongful Conviction

Historically, forensic science and miscarriages of justice were rarely reflected upon in concert. Certainly, forensic science has been cited as a contributor to miscarriages of justice since as long ago as the Dreyfus case.1 But until recently, forensic science—compared to other issues like eyewitness identification, perjury, official misconduct, and interrogation practices—has tended to take a back seat in discussions of miscarriages of justice.2 Although the earliest U.S. study of miscarriages of justice mentioned “[t]he unreliability of so-called ‘expert’ evidence” as a contributor to wrongful convictions,3 most of the early studies that attempted to systematically identify causes of wrongful conviction discussed: eyewitness identification; false confessions; police and prosecutorial misconduct; bad lawyering; race; failures of the discovery process; and public pressure for a conviction, making scant mention of forensic science.4 Two Royal Commissions issued reports addressing problems with forensic science in Australia during the 1980s attracting little international attention.5 As Schiffer and Champod observed, “forensic science (to convict and to exonerate) is underrepresented and often wrongly understood in research concerning wrongful convictions.”6

---


This made some intuitive sense because the characteristics popularly associated with “science” seem antithetical to the characteristics of wrongful convictions. Wrongful convictions were thought to be caused by unclear, misguided, or fallacious reasoning, but science is supposed to embody clear, rational reasoning. Wrongful convictions were also thought to be caused by unjustified biases against people of certain races or classes, against persons with prior criminal records, or even simply against the police’s preferred suspect. By contrast, science is supposed to be objective and free of bias. Wrongful convictions were thought to be caused by deceitful and otherwise unreliable information given by witnesses, informants, co-conspirators, and the like. But science, goes the truism,

9 See Understand the Causes: Eyewitness Misidentification, THE INNOCENCE PROJECT,
“never lies.” Wrongful convictions were thought to be caused by evidence that was less reliable than it appeared—like eyewitness identification evidence—but science is by nature associated, in the popular imagination, with high reliability, indeed often with certainty.

II. Forensic Science as Exposer of Miscarriages of Justice

The development of forensic DNA profiling during the 1980s caused people to begin associating forensic science with miscarriages of justice.10 Beginning with Gary Dotson, and then David Vasquez in 1989, post-conviction DNA testing exposed a number of miscarriages of justice in the United States.11 Realizing the potential of post-conviction DNA testing to expose miscarriages of justice, in 1992, American attorneys Peter Neufeld and Barry Scheck founded the Innocence Project at Cardozo Law School as a legal clinic dedicated to such testing.12 Over the next two decades, the Innocence Project and other independent efforts exposed more than 250 wrongful convictions in the United States through post-conviction DNA testing.13 This set of wrongful convictions has taken on a degree of significance beyond the parties involved in the underlying cases themselves. These cases have acquired significance in drawing attention to the issue of miscarriages of justice, to flaws in the American justice system, and to capital punishment. In part, their significance derives from their ability to be packaged and conceptualized as a “data set” through media such as reports, books, and the Innocence Project’s own website. Additionally, their significance derives from their ability to achieve supposed “scientific certainty”14 or “epistemological closure.”15 Alleged miscarriages of justice exposed through post-conviction DNA testing were less vulnerable to disputes over whether they should be characterized as

11 See id. at 63, 74 & n.71 (stating that Gary Dotson was the first person in the United States to be exonerated by post-conviction DNA testing).
13 Facts on Post-Conviction DNA Exonerations, INNOCENCE PROJECT, http://www.innocenceproject.org/Content/Facts_on_PostConviction_DNA_Exonerations.php (last visited June 1, 2012) (stating that in the United States, 289 people have been exonerated by post-conviction DNA testing).
miscarriages of justice at all. While some post-conviction DNA exonations may be challenged, even the most determined “innocence skeptics” concede that the vast majority of post-conviction exonations were miscarriages of justice.16

III. Forensic Science as Contributor to Miscarriages of Justice

Thus, forensic science was initially perceived as a powerful tool for exposing wrongful convictions.17 However, the earliest analyses of post-conviction DNA exonations as a data set revealed a paradox. Forensic science was not merely the engine for exposing miscarriages of justice; it also appeared to be a contributor to miscarriages of justice.18 An analysis of the first twenty-eight cases of post-conviction DNA exonations noted:

A majority of the cases involved non-DNA-tested forensic evidence that was introduced at trial. Although not pinpointing the defendants, that evidence substantially narrowed the field of possibilities to include them. Typically, those cases involved comparisons of nonvictim specimens of blood, semen, or hair at the crime scene to that of the defendants. Testimony of prosecution experts also was used to explain the reliability and scientific strength of non-DNA evidence to the jury.19

Thus, post-conviction DNA exoneration introduced forensic science into the discourse on miscarriages of justice in two ways: (1) as a tool for exposing miscarriages of justice in a way that allowed for bypassing debates over whether the alleged miscarriages of justice were, in fact, miscarriages of justice; and (2) as a potentially important cause of miscarriages of justice. Paradoxically, forensic science was little discussed as a cause of miscarriages of justice until its role was exposed—by forensic science.

19 CONNORS ET AL., supra note 17, at 15.
During this same period, a miscarriage of justice “crisis” arose in the United Kingdom. Among the most prominent alleged miscarriages of justice were three 1974 Irish Republic Army (“IRA”) bombing cases. These cases resulted in the convictions of the so-called “Guildford Four,” “Birmingham Six,” and “Maguire Seven.”20 All of the cases involved explosive residue evidence.21 These cases prompted two official inquiries, which highlighted the role of forensic science in miscarriages of justice. The Royal Commission on Criminal Justice’s 1993 “Runciman Report” discussed a number of issues concerning forensic science including: failure to adhere to objectivity and impartiality; problems with interpretation of evidence; failure to communicate findings clearly; inequalities between defense and prosecution resources; defense access to samples; prosecution bias; expert shopping; and the low accuracy of the residue detection techniques themselves.22 In 1994, the “May Inquiry” discussed the role of forensic science in the Guildford Four and Maguire Seven cases. The May Inquiry primarily blamed individual forensic scientists for the

21 Id. at 189. Clive Walker & Carol McCartney, Criminal Justice and Miscarriages of Justice in England and Wales, in Wrongful Conviction: International Perspectives on Miscarriages of Justice, supra note 6, at 183, 188.
failings of forensic science in cases involving miscarriages of justice.\textsuperscript{23}

In Canada, the 1998 “Kaufman Report” discussed the role that microscopic hair comparison played in the wrongful conviction of Guy Paul Morin for murder.\textsuperscript{24} Morin had been exonerated by post-conviction DNA testing. Among other things, the “Kaufman Report” noted that both the overstatement of the probative value of forensics analysts’ findings and the failure to disclose contamination problems contributed to Morin’s wrongful conviction.\textsuperscript{25} In 2000, Barry Scheck, Peter Neufeld, and Jim Dwyer published \textit{Actual Innocence}, which analyzed the first sixty-two post-conviction DNA exonerations.\textsuperscript{26} The book devoted two chapters to forensic science as a contributor to miscarriages of justice, splitting the issue into scientific misconduct (“White Coat Fraud”) and the unreliability of forensic science (“Junk Science”).\textsuperscript{27} The former discussed notorious forensic vigilantes, such as Fred Zain.\textsuperscript{28} The latter discussed the unreliability of microscopic hair analysis and bite mark comparison.\textsuperscript{29} It also addressed the need for regulation of forensic laboratories, proficiency testing, clear reporting of error rates in order for fact-finders to assign weight to forensic evidence, transparency, and independence from law enforcement.\textsuperscript{30} In addition, the book revealed that the Innocence Project’s founders became familiar with forensic DNA profiling by working on the Coakley case, which involved serology evidence that was misleadingly interpreted.\textsuperscript{31} The authors cited “serology inclusion” as the second leading contributor, after “mistaken identification,” to the wrongful convictions exposed by post-conviction DNA testing—contributing to thirty-two of the sixty-two wrongful convictions.\textsuperscript{32} “Defective or Fraudulent Science” was listed as the fifth leading cause with twenty-one cases, “microscopic hair comparison” was sixth with eighteen cases, “other forensic inclusions” was the eleventh leading cause with five cases, and “DNA inclusions” was the twelfth leading cause with one case.\textsuperscript{33}
Michael Saks and Jonathan Koehler cited the role of forensic science as a contributor to miscarriages of justice to support their claim that forensic science was not as reliable as it was often claimed to be and reform of forensic science was urgently needed. Saks and Koehler suspected that Scheck et al.’s splitting of forensic science into multiple categories might obscure the significant role that forensic science plays as a contributor to miscarriages of justice. As a result, they published a slight reanalysis of the Innocence Project data—which at that time represented the first eighty-six post-conviction DNA exonerations—by aggregating all forensic contributors into just two categories. Errors in forensic science testing were found to be present in sixty-three percent of cases, second only to eyewitnesses misidentifications. The giving of false or misleading testimony by forensic scientists was found to be present in twenty-seven percent of cases, the fifth most common contributor. The article’s placement in the prestigious journal *Science* helped ignite renewed efforts to reform forensic science in the United States. Some forensic scientists, upset by the article’s portrayal of their field, questioned “the data sampling techniques, methods, [and] criteria” that went into Saks and Koehler’s representation of the forensic sciences as a leading contributor to miscarriages of justice. Saks and Koehler responded that “[r]esearch on DNA exonerations is obviously in its infancy, and we support calls for a more complete and scientific review of these cases.”

Of course, post-conviction DNA testing is only one method of exposing miscarriages of justice. In a comprehensive study of U.S. exonerations from 1989, the beginning of the post-conviction DNA exoneration era, to 2003, Gross and colleagues found a total of 340 exonerations, slightly less than half of which were exposed by post-conviction DNA testing. Gross and his colleagues devised their own system for categorizing the causes of these wrongful convictions; forensic science was lumped into their “perjury” category. The study found that twenty-four of the wrongful convictions involved perjury by a forensic scientist. This analytic

---

35 Id. at 892-93.
36 Id. at 892 fig.1.
40 Id. at 543.
approach narrowed the apparent contribution of forensic science to miscarriages of justice.

In 2008, Garrett published an extensive analysis of the first 200 post-conviction DNA exonerations. He reported that “[f]orensic evidence was the second leading type of evidence supporting these erroneous convictions,” appearing in fifty-seven percent of the cases.\(^{41}\) A more extensive analysis of the role of forensic science in wrongful convictions reported that “[t]wo hundred thirty-two innocent persons have now been exonerated by post-conviction DNA testing[\(]\)” and offered an extensive analysis of these cases.\(^{42}\)

A. The Contribution of Convicting the Innocent

Professor Garrett’s book, *Convicting the Innocent*, which extended the analysis to the first 250 post-conviction DNA exonerations, now stands as the definitive analysis of the data set of wrongful convictions exposed through post-conviction DNA testing.\(^{43}\) In this larger study, Garrett found forensic evidence was used in 185 of the 250 post-conviction DNA exonerations studied—or seventy-four percent of the time.\(^{44}\) The key methodological advance in Garrett’s study was his examination of the trial and hearing transcripts and case files of all of the post-conviction DNA exonerations, which he successfully obtained in 220 of the 250 cases.\(^{45}\) Thus, his study was based on the trial or hearing transcripts and case files, when available, rather than on the Innocence Project’s capsule case summaries. This is important because it gives us a sense not just of what type of evidence was present in the case, but what words “the forensic analysts actually said on the witness stand.”\(^{46}\) It may be argued that ensuring the accuracy of the words expert witnesses use to convey the probative value of their findings is at least as important as ensuring the accuracy of their forensic techniques.\(^{47}\) In addition, Garrett’s method allowed him to drill further into the nature of the forensic testimony. Rather than merely noting that forensic evidence was “involved” or “implicated” in a wrongful conviction, Garrett was, in many cases, able to report whether the evidence was inculpatory, exculpatory, or neither; what weight was assigned to it by

\(^{41}\) See Garrett *supra* note 10, at 64, 81.


\(^{43}\) See GARRETT, *supra* note 7, at 108-14.

\(^{44}\) Id. at 89.

\(^{45}\) See id. at 7.

\(^{46}\) See id. at 89, 90.

the government’s expert witnesses; whether it was challenged by a defense witness; and whether the government’s expert witness characterized the evidence in a scientifically proper manner. He was not, however, able to measure what weight the fact-finder assigned to the forensic evidence.

Garrett concluded that “[m]ost . . . forensic analysis at these trials offered invalid and flawed conclusions,” and that he “was confronted with a parade of invalid forensics.” Garrett sorted the “problems with forensics” into “two recurring types,” which he labeled “reliability” and “validity.” The first problem was the use of “unreliable” forensic evidence, which Garrett defines as “[a] method . . . [that] does not produce consistent or accurate results.” Garrett does not report the proportion of cases that included “unreliable” forensic evidence. More than one type of “unreliable” forensic evidence may have occurred in the same case. The second problem is the presenting of “invalid” conclusions based on a forensic analysis where the analysis itself might be “reliable” or “unreliable.”

1. Reservations About Categorizing the Causes of Forensic Problems

Although it is not central to my argument in this paper, I should express some reservations about this scheme of categorizing forensic problems. First, the terminology may be confusing. While “reliability” and “validity” are often used interchangeably in law, in science the two concepts are sometimes distinguished. When “reliability” is distinguished from “validity,” it is usually in order to make the following distinction: “reliability” refers to consistency of outcomes (i.e., the process yields the same results when performed at different times or by different analysts or laboratories), whereas “validity” refers to whether the process measures what it is intended to measure. But, this is not the distinction being drawn by Garrett. Instead, he conflates the two concepts under the notion of “reliability”: “A method is unreliable if it does not produce consistent or

---

48 GARRETT, supra note 7, at 108-14.
49 Id. at 7.
50 Id. at 9.
51 Id. at 89.
52 Id. at 86, 90.
53 Id. at 89-90.
54 GARRETT, supra note 7, at 89-90.
55 Id. at 90.
accurate results.” The best term for this is probably “validity”—the term Garrett has reserved for the other category of forensic problems—since “validity” is generally understood to include “reliability” but not vice versa. Specifically, a process cannot be more valid than it is reliable—it cannot produce accurate results at a greater rate than it produces consistent results. However, a process can be more reliable than it is valid—it can produce consistent results at a greater rate than it produces accurate results.

Meanwhile, what Garrett calls “validity,” appears to refer primarily to making correct inferences from the evidence. As Garrett himself notes: “[T]he term validity refers to whether claims or inferences are supported by the evidence.” However, this is the meaning of the term “validity” as used in logic rather than the meaning used in science. Thus, it seems to me, the distinction that Garrett is trying to make would have been clearer had he chosen different labels for the “two recurring types of problems with . . . forensics.” For example, he might have labeled them: (1) “problems of validity” and (2) “problems of inference.”

This terminological confusion is exacerbated by some problems with the integrity of the categories themselves. Garrett views only two forensic techniques as “reliable”: serology and DNA profiling. This appears to be because the techniques are “not based on a subjective opinion” and because expert witnesses are able to defensibly convey the probative value of serology or DNA profiling evidence by estimating the rarity of the biological features found consistent with the suspect. However, Garrett’s argument is slightly misleading. The issue is not the subjectivity of the opinion. After all, even opinions about DNA profiles are in some sense subjective. Rather the issue is whether the opinion is based on an instrumental measurement or a visual analysis. A conclusion that serology and DNA profiling are “reliable” ought to be based on: (1) studies that measure the rate at which users of the technique achieve correct results on samples of known origin, and (2) as Garrett correctly notes, expert witnesses’ ability to provide defensible estimates of the rarity of the biological features in question, so as to convey the probative value of the evidence to the fact-finder.

57 See GARRETT, supra note 7, at 86.
58 See id. at 87.
59 Id. at 89.
60 See id. at 90.
61 See id. at 86.
63 See GARRETT, supra note 7, at 86.
Garrett labels all of the other forensic techniques—microscopic hair comparison, fingerprint comparison, bite-mark comparison, shoe print comparison, and voice comparison—as “unreliable.” He defines “unreliable” as follows: “A method is unreliable if it does not produce consistent or accurate results.”64 As an example, Garrett offers microscopic hair comparison, and he notes that “[s]udies have found very high error rates in hair comparison.”65 Strictly speaking, Garrett’s definition is imprecise: he is not claiming that hair comparison “does not produce consistent or accurate results,” merely that it “does not produce consistent or accurate results” at a high enough rate to satisfy his notion of “reliability.”66 This may seem like a minor point, but it draws our attention to an important issue: “reliability”—or “validity”—should not be conceived as a dichotomous variable in which techniques are “reliable” or “unreliable.” Rather, we should understand that all forensic techniques should be expected to produce some correct results and some false results; the important issue is the ratio of correct-to-false results, which we may term the technique’s accuracy. “Accuracy” may be understood as the proportion of all results that are correct results. Once we know various techniques’ degrees of accuracy, we can exercise value judgments and describe techniques that surpass some arbitrary threshold of accuracy as “reliable” and refer to those that do not as “unreliable.”

Now that our focus is on accuracy, it becomes clear that Garrett’s “unreliable” category lumps together a variety of techniques for which the state of scientific knowledge with regard to their accuracy is quite different. Hair comparison is the only technique for which Garrett explicitly claims sufficient studies exist so that he can know the accuracy of the technique.67 Garrett claims that these studies show that the accuracy rate is poor. So hair comparison is deemed unreliable, because he believes that he has adequate knowledge that its accuracy is poor. The state of scientific knowledge with regard to the accuracy of all the remaining techniques is quite different. Garrett offers no evidence that the accuracy of any of these techniques has been measured and found to be poor. Instead, he cites the NAS Report, that found these disciplines “supported by little rigorous systematic research to validate . . . basic premises and techniques.”68 This statement by the NAS is not a report of an empirical finding of poor
accuracy. Rather, it is the conclusion of a literature review which has found a dearth of empirical studies of accuracy. In other words, the NAS is claiming that the accuracy of these techniques has not been sufficiently studied, but it is not making any substantive claim about the accuracy of these techniques.

It also becomes clear that Garrett’s “unreliable” category contains techniques—like hair comparison—for which he claims that their accuracy is known to be poor, and techniques—like shoe print comparison—for which he merely claims that their accuracy is unknown. It is not always obvious in which of these two groups a particular technique belongs. At least some rudimentary studies exist for most of these techniques. The issue is whether these studies should be considered so rudimentary that little or nothing can be inferred about the accuracy of the techniques or whether the studies should be considered sufficiently robust that such inferences can be made. For example, do the few studies we have about bite mark comparison mean that the accuracy of bite mark comparison is unknown or known to be poor? The situation is further complicated by the fact that the NAS’s statement applies to hair comparison as well—the NAS appears to view the accuracy of hair comparison as unknown, whereas Garrett claims it is known to be poor.

Of course, testifying about a technique whose accuracy is unknown is problematic as well. If such testimony was proffered at trials, then government expert witnesses testified at trials about the results of these techniques without being able to give the fact-finder credible information about their accuracy. As Garrett shows, in lieu of such credible information, the government’s expert witnesses relied on crude, vaguely defined vernacular terms like “similar,” “consistent,” and “match.” As Garrett notes, although these terms have no inherent meaning in forensic science, jurors appear to ascribe them a very high probative value.

Garrett reports that 185 out of 250 cases (seventy-four percent) had forensic evidence present, that 169 of these had forensic trial testimony, that he was able to locate the transcript for 153 of these 169 trials, and that in

69 Id. at 2.
70 Compare id. at 161, with GARRETT, supra note 7, at 86.
71 GARRETT, supra note 7, at 90.
sixty-one percent of these 153 trials the testimony was “invalid.” The analysis of this “invalid” testimony forms Garrett’s principal contribution to the discussion of the role forensic science plays in miscarriages of justice. Elsewhere, Garrett analyzed the “invalid” testimony in greater detail. He defined six different categories of “invalid” testimony: (1) interpreting the nonprobative evidence as inculpatory; (2) discounting exculpatory evidence; (3) presenting inaccurate statistics; (4) providing frequencies or probabilities in the absence of empirical data; (5) providing nonnumerical statements of probability or frequency despite a lack of any empirical data; (6) concluding evidence did in fact come from the defendant despite no empirical data permitting such conclusions. Garrett’s data, therefore, presents a very complex picture. The forensic evidence he discusses varies along at least four important dimensions. We have forensic evidence that is inculpatory, exculpatory, and inconclusive, that reports results from at least seven different forensic techniques, resulting in trial testimony that is either valid or invalid in one or more of six different ways, over a twenty-four-year period. Making generalizations from such data is challenging to say the least.

2. The Significance of Garrett’s Findings for Forensic Science

Nonetheless, some important inferences about the state of forensic science during this twenty-four-year period can certainly be drawn. The data clearly indicates some of the problems that pervaded forensic science during this period: the absence of quantitative data with which to convey the probative value of the evidence to the fact-finder; the reliance, in lieu of such data, on verbal characterizations that impossibly overstated the probative value of the evidence, such as stating the defendant “was” the source of some forensic traces; the use of faulty statistics in lieu of such data; and the interpretation of evidence in a manner biased against defendants. The fact that all this forensic evidence—derived from techniques of either unknown or poor accuracy, couched in “invalid” testimony—was permitted to be presented to juries indicates a failure of the “gatekeeping” responsibility vested in trial courts to ensure the relevance and reliability of forensic evidence proffered in criminal trials. The data also suggests poor performance by the prosecutors who

73 Garrett, supra note 7, at 89-90.
75 Garrett, supra note 7, at 90, 254.
presented invalid testimony and relied upon it in closing arguments, defense attorneys who failed to challenge such evidence, and juries that ignored exculpatory forensic evidence. Certainly, one can conclude from Garrett’s data that during this period forensic science was tardy in conducting studies that would generate data from which to make defensible inferences about the probative value of evidence. Even more, during this period forensic science was poorly regulated, made few efforts to eliminate bias from interpretation of evidence, and American courts permitted all of this without sanction.

In sum, the great achievement of Convicting the Innocent lies in its drilling down to the level of the transcript. Garrett is able to show not merely that forensic evidence was present or even contributed to wrongful convictions but how the evidence was consciously or unconsciously distorted to fit—and strengthen—the government’s theory of the defendant’s guilt.

3. The Significance of Garrett’s Findings for Post-Conviction DNA Exonerations

Next, we might ask what this data tells us about the role of forensic science in miscarriages of justice that were exposed by post-conviction DNA exonerations. The answer, clearly, is that forensic science played an important role. However, this derives less from simply counting the proportion of cases in which forensic evidence was present, or even the proportion of cases in which forensic evidence was inculpatory, or even the proportion of cases in which the evidence derived from “unreliable” techniques or the testimony was “invalid.” A closer look at Garrett’s data clearly shows that forensic science primarily contributed to the miscarriages of justice that were exposed by post-conviction DNA exonerations in two ways.

First, serological evidence which ought to have been interpreted as either excluding the defendant or as having nothing useful to contribute to the fact-finder’s perception of the defendant’s guilt was instead presented to the fact-finder as inculpatory. This occurred in sixty-seven cases. Second, microscopic hair comparison evidence that ought, if used at all, to have been conveyed to the fact-finder only as failing to exclude the

76 See DANIEL S. MEDWED, PROSECUTION COMPLEX: AMERICA’S RACE TO CONVICT AND ITS IMPACT ON THE INNOCENT 97 (2012).
77 GARRETT, supra note 7, at 90.
78 See id. at 90-91 (explaining that many convictions were based on a range of unreliable forensic methods, most of which are still in wide use and permitted in criminal courtrooms today).
79 Id. at 90.
defendant or perhaps as including the defendant, among a very large population that could have contributed the hair, was instead presented to the fact-finder as highly incriminating. This occurred in twenty-nine cases.

These cases strongly support the argument made by a number of scholars that bias is a potentially significant problem for forensic science.\textsuperscript{80} It is clear that the erroneous interpretations in these cases would not have been possible were the analysts not aware of police investigators’ desired outcome (i.e., who was the suspect) because otherwise they would not have known when to suppress exculpatory results and when to interpret nonprobative results as incriminating. As Garrett notes, the bias in these cases was, without exception, against the defendant, although this observation would seem to be attributable to the fact that Garrett was working with a data set of post-conviction exoneration cases.\textsuperscript{81} It is true that there were lots of other forms of problematic evidence involving other types of “invalid” testimony and other forensic techniques. But few of these other problems were frequent enough to allow us to draw generalizations. The extraordinary prevalence of serology and hair comparison in post-conviction DNA cases emerges strongly from Garrett’s data.\textsuperscript{82}

4. The Significance of Garrett’s Findings for Unexposed Miscarriages of Justice

While Garrett’s data tells us a great deal about the contribution of biased serology and microscopic hair comparison to miscarriages of justice exposed by post-conviction DNA testing, what the data tell us about the contribution of forensic science in general to miscarriages of justice is a more difficult question. The crucial point to keep in mind is that, as many scholars have noted, post-conviction DNA exonerations are an unrepresentative data set.\textsuperscript{83} That is, post-conviction DNA exonerations represent a set of wrongful convictions that were amenable to exposure through post-conviction DNA testing.\textsuperscript{84} This means that biological


\textsuperscript{81} GARRETT, supra note 7, at 90.

\textsuperscript{82} Id.


\textsuperscript{84} See Justin Brooks & Alexander Simpson, \textit{Blood Sugar Sex Magik: A Review of Post-
Forensic Science and Wrongful Convictions

Evidence probative of guilt must have been present at the crime scene; that this evidence was collected and preserved; that it was either not subjected to DNA testing, subjected to an older, less discriminating method of analysis, or erroneously interpreted; that the evidence remained available for testing through post-conviction litigation; and that the crime was serious enough that it yielded a sentence severe enough that the convict had an incentive to pursue post-conviction DNA testing. In practice, this means that sexual assault cases are highly over-represented in this data set, that homicides are over-represented as well, and so are the combination of the two: rape-murders. Thus, Garrett’s work allows us to conclude that biased serology and microscopic hair comparison evidence were major contributors to those miscarriages of justice that occurred in this selective set of cases.

However, it is clear that there are other miscarriages of justice less amenable to exposure through post-conviction DNA testing. For example, as I have shown elsewhere, wrongful convictions involving fingerprint evidence are less amenable to exposure through post-conviction DNA testing than wrongful convictions involving serology or hair.

What can we conclude about these cases from Garrett’s study? The answer to this question is less clear. One perspective would be to treat Garrett’s set of cases as deviant—to assume that Garrett has identified a problematic set of cases which are susceptible to specific problems. For example, sexual assault prosecutions are especially problematic because they are crimes that provoke high emotions and with specific evidence collection problems (e.g., eyewitnesses who are not the victim or perpetrator are rare). Or, perhaps serology and microscopic hair

---


86 Gross, supra note 83, at 179 (“Once we move beyond murder and rape cases, we know very little about any aspect of false convictions.”); Alexandra Natapoff, Misdemeanors, 85 So. CAL. L. REV. (forthcoming 2012) (manuscript at 46) (“[T]he innocence movement is centrally concerned with serious offenses, typically murder and rape, which together comprise a small fraction of the criminal system—around 2 percent.”); D. Michael Risinger, Innocents Convicted: An Empirically Justified Factual Wrongful Conviction Rate, 97 J. CRIM. L. & CRIMINOLOGY 761, 772 (2007).

87 See Gross, supra note 83, at 174-75 (“We cannot study an event if we can’t tell when it happens. This is a severe problem for false convictions because, by definition, we don’t know when they occur.”).

comparison are particularly problematic techniques, and cases in which they were used as evidence were particularly susceptible to error.

Another perspective would be to treat Garrett’s set of cases as revealing a deeper set of causes, implying similar problems with different forensic techniques that are more prevalent in other sets of cases but less amenable to post-conviction DNA testing. From this perspective, the “invalid” testimony concerning serology and hair comparison are indicative not merely of problems with these two techniques but also of deeper problems with forensic science and the law’s use of it. For example, Garrett’s findings are revealing of the tendency of crime laboratories to exaggerate or distort the probative value of forensic evidence in general and the legal system’s tendency to permit such exaggerations and distortions. This perspective would conclude that Garrett’s data suggests that crime laboratories probably routinely provide “invalid testimony” derived from other forensic techniques, and the courts probably routinely permit that invalid testimony. We have not been made aware of the problem through exposed miscarriages of justice only because we do not have a convenient mechanism for post hoc exposure of miscarriages of justice, like post-conviction DNA testing. Thus, our greater awareness of problems in serology and hair comparison is better explained by the fact that we have improved exposure mechanisms than by the fact that there are more underlying problems with those techniques.

5. Collins’s and Jarvis’s Critique of Garrett’s Work

The difference between these two perspectives has been brought into relief by a critique of Garrett’s work by two crime laboratory advocates. Collins and Jarvis contend that the usual methods of analyzing data derived from post-conviction DNA exonerations, which involve counting the proportion of cases in which forensic science was involved, overstates the contribution of forensic science to miscarriages of justice. They argue that the presence of forensic science in a case does not necessarily imply that forensic science caused the miscarriage of justice.

This critique certainly has some merit for early analyses which simply counted whether forensic science was “present” in post-conviction DNA exonerations. But Garrett has improved on this method by counting the number of cases in which forensic science was not merely present but “unreliable” or “invalid.” However, Collins and Jarvis still contend that Garrett overstates the contribution of forensic science. In their own analysis

89 Garrett, supra note 7, 89-91.
91 See id. at 21-22.
of the post-conviction DNA exoneration data, they conclude that forensic science played a causal role in the miscarriage of justice in only 32 of the first 200 post-conviction DNA exoneration cases, or sixteen percent. 92 In a further analysis in which the total number of contributors to the wrongful conviction, rather than the total number of cases, is treated as the denominator, they conclude that only eleven percent of the total causes of wrongful convictions were forensic science. 93

Why are these figures so different from Garrett’s claim that sixty-one percent of the post-DNA exonerations contained flawed forensic testimony? The thirty-two cases in which Collins and Jarvis attribute a causal role to forensic science are limited to cases involving what they call “forensic malpractice.” 94 This is defined as “every conceivable failure that could be committed, either intentionally or accidentally, by a forensic scientist or a forensic science facility.” 95 Based on this definition, Collins and Jarvis exclude a large number of cases that were included by Garrett. For example, Collins and Jarvis categorized sixty-nine cases as instances where “[n]on-specific science failed to exclude the defendant.” 96 Although it is not clear from Collins and Jarvis’s article, 97 these are probably the serology and hair comparison cases. The argument is this: serology is accurate but not very discriminating. Suppose a simple serological analysis reveals that the perpetrator is blood type A—a type shared with approximately forty percent of the population—and the suspect is blood type A. The analysis has failed to exclude the suspect, but proper understanding of the evidence should make clear that this failure is not especially probative: an individual picked at random from the population would have a forty percent chance of not being excluded. Put another way, an individual with a perfect alibi, who could not possibly have committed

92 Id. at 25.
93 Id.
94 Their coining of this term is criticized in Norah Rudin & Keith Inman, Who Speaks for Forensic Science?: The Conviction and Exoneration of a Straw Man, CAL. ASS’N OF CRIMINALISTS NEWS, Fourth Quarter 2008, at 10, 11.
95 Collins & Jarvis, supra note 90, at 24.
97 On January 26, 2012, I wrote to the authors through their website, CRIME LAB REPORT, http://www.crimelabreport.com (last visited June 1, 2012), where a pre-publication version of this study appeared. I received two replies. One stated: “The e-mail message could not be delivered because the user’s mailfolder is full.” The second reply stated: “The CRIME LAB REPORT editors have received your message. Thank you for contacting us.” I wrote Mr. Jarvis at an email address listed on the AMERICAN SOCIETY OF CRIME LABORATORY DIRECTORS LABORATORY ACCREDITATION BOARD, http://www.ascld-lab.org (last visited June 1, 2012), but received no response.
the crime, would have a forty percent chance of not being excluded.

Suppose that the same defendant is subsequently convicted, based in part on the serological evidence. Suppose further that DNA typing is subsequently developed, the convict requests post-conviction DNA typing, the test is performed, and it excludes the defendant. DNA profiling is far more discriminating than blood type. The temptation is to say that the serological evidence was “wrong” or “false” or “flawed,” but, of course, this is somewhat unfair. The serologist never said that the suspect was the perpetrator. True, the serological evidence was used against a defendant who turned out to be innocent. However, the serologist presumably allowed for the fact that the suspect may have been among the many innocent people that a blood-type analysis failed to exclude. The same argument may be mounted in defense of hair comparison. Microscopic hair comparison is not considered to be very discriminating. Thus, it should not be surprising that many suspects might be included by microscopic hair comparison, but excluded by DNA profiling.

This is a sound argument if the expert witness properly conveys the probative value of the evidence to the fact-finder. In serology, that would mean conveying not only that the analysis failed to exclude the defendant, but also the proportion of innocent persons the test would also be expected to fail to exclude. For microscopic hair comparison, the problem would be trickier because responsible estimates of the population frequencies of the attributes analysts rely upon are not available. Therefore, the analyst would be forced to employ vague verbal characterizations of the discriminating power of the observed features. The testimony would be something like: “These attributes of the crime scene hairs were shared by the sample hair from the suspect; however, many people in the population also share these attributes.”

However, Garrett’s findings show that the forensic expert witnesses did not properly convey the probative value of the evidence to the fact-finder. Rather, as noted above, Garrett shows that most of the serology cases were not simply “failures to exclude” but cases in which “masking” problems meant that the evidence should have been reported as having no probative value whatsoever—that the test failed to exclude one-hundred

---

100 See GARRETT, supra note 7, at 92-93.
percent of the population. Or, put another way, that there was a one-
hundred percent probability that the test would fail to exclude an innocent
person. In an additional set of cases, exculpatory evidence was reported as
either non-probative or inculpatory. Similarly, Garrett shows that in
many of the hair comparison cases, the probative value of the evidence was
exaggerated. It is misleading to categorize cases in which the probative
value of the evidence was skewed against the defendant as “[n]on-
specific science failed to exclude the defendant.”

The above cases probably account for the bulk of the difference
between Garrett’s and Collins and Jarvis’s counts of the numbers of
wrongful convictions attributable to forensic science. Collins and Jarvis
also exclude cases in which the forensic evidence proffered against the
defendant derived from “disciplines . . . rarely, if ever, practiced in
America’s crime laboratories.” These include voice-print analysis and
dog-scent tracking. In my opinion, these analyses fit better in a category
called “forensic science” than in any other commonly used category of
evidence. Certainly, in the categorization typically used for the analysis
of wrongful convictions, dog scent fits better with “forensic science” than
with any of the other categories. But, in any event, the number of these
cases is small, so this difference in classification accounts for little of the
discrepancy between Garrett’s numbers and those of Collins and Jarvis.

101 See id. 87, 94-95.
102 See id.
103 See id. at 96-99.
104 See id. at 86.
105 Collins & Jarvis, supra note 90, at 24.
106 See id. at 15, 20; see also Rudin & Inman, supra note 94, at 10, 12.
107 Voice analysis is a frequent subject of publications in forensic science journals. See, e.g.,
Geoffrey Stewart Morrison, Forensic Voice Comparison and the Paradigm Shift, 49 SCI. & JUST.
298, 301, 307 (2009). Dog scent, though perhaps less frequently, is still present. See, e.g., Lisa M.
Harvey & Jeffery W. Harvey, Reliability of Bloodhounds in Criminal Investigations, 48 J. FORENSIC
SCI. 811, 811 (2003); Gertrud A.A. Schoon, A First Assessment of the Reliability of an Improved
Scent Identification Lineup, 43 J. FORENSIC SCI. 70, 70 (1998); Gertrud A.A. Schoon & J.C.
DeBruin, The Ability of Dogs to Recognize and Cross-Match Human Odours, 69 FORENSIC SCI. INT’L
111, 112 (1994). There is also a Scientific Working Group (“SWG”) for scent dogs called
SWGDOG. See About Scientific Work Groups, SWGFAST, http://www.swgfast.org/About
SWGhs.htm (last visited June 1, 2012) (“Since the early 1990s, American and International
forensic science laboratories and practitioners have collaborated in Scientific Working Groups . .
. to improve discipline practices and build consensus standards.” (emphasis added)); see also
Andrew E. Taslitz, Does the Cold Nose Know? The Unscientific Myth of the Dog Scent Lineup, 42
HASTINGS L.J. 15, 54 (1990) (arguing that dog scent evidence should legally be treated as
scientific expert evidence). But see Rudin & Inman, supra note 94, at 12 (agreeing with Collins
and Jarvis that “most criminals would not consider bloodhounds part of our profession”).
In purporting—in a clever rhetorical conceit—to “exonerate” forensic science from its “wrongful conviction” by Garrett, Collins and Jarvis identify who they believe to be the true perpetrator: the legal system. As both a general matter and in specific cases, Collins and Jarvis argue that forensic science usually “gets it right,” while the legal system is to blame for miscarriages of justice. They contend that “innocence activists” like Garrett prefer to focus on forensic science as a cause of miscarriages of justice because “many of the legislators with whom innocence activists hope to curry favor are practicing lawyers themselves” and because “blaming lawyers is boring.” Thus, in the case of the wrongful conviction of Ray Krone, for example, Collins and Jarvis blame prosecutorial misconduct and “bad lawyering” because a bite mark examiner consulting for the government concluded that the bite mark excluded Krone. The government ignored this evidence, sought a second opinion from another bite mark examiner who implicated Krone, concealed the exculpatory opinion from the defense, and Krone was convicted. Garrett blames “forensic science” because, after all, a forensic scientist provided evidence that was undoubtedly crucial in persuading the jury to convict Krone. Collins and Jarvis reply that forensic science “got it right” because a forensic scientist did report a conclusion that the evidence excluded Krone.

IV. Forensic Science as “Corrector” of Miscarriages of Justice

Clearly, Krone is an odd case upon which to base the “exoneration” of forensic science. If nothing else, the case shows a weakness of bite-mark analysis given that two different examiners—who were deemed qualified prior to the examination—reached different conclusions based on the same data. But the broader conclusion to be drawn from this case is how artificial it is to treat “forensic science” and “law” as separate entities whose contributing roles in miscarriages of justice must be accounted for separately, as if the two had nothing to do with one another. It seems,
instead, that one can only understand “forensic science” in the context of its use in “law,” and one can only understand “lawyering” in the context of the evidence that lawyers use to support proof. The disparity between Garrett’s position and that of Collins and Jarvis, therefore, like the earlier dispute between Saks and Koehler and defenders of forensic science, illustrates the limitations of blaming discrete “causes” like “eyewitness identification,” “false confession,” “prosecutorial misconduct,” “bad lawyering,” and “forensic science” for miscarriages of justice. Further, this dispute illustrates the limitations of trying to quantify the impact of forensic science on wrongful convictions relative to other causes. As Professor Natapoff notes: “This posture assumes that if those discrete pieces of evidence were stronger, the convictions would be sound. . . . Rather than focusing on discrete pieces of evidence such as confessions or fingerprints, innocence skepticism should be aimed at the entire procedural apparatus.”

To be clear, I am not questioning the value of what we might call “the categorizing project.” This project has certainly succeeded in giving us a rough sense of what the contributors to miscarriages of justice are, and the data is sufficient to warrant the inference that our criminal justice system has weaknesses in each of these areas. It is understandable that authors of books about miscarriages of justice find that the standard categories offer a convenient way of organizing the material. So, certainly, this “categorical” way of thinking about miscarriages of justice has its uses. But even those who find it useful presumably understand that it is, ultimately, a simplification. It is well understood, moreover, that data limitations make drawing further inferences about the relative role of each of these areas or their general accuracy problematic. In real cases, these discrete “causes” interact with one another in complex and contradictory ways. Finding patterns in these interactions is far more challenging than counting up cases in which a particular “cause” was present or contributed to a

114 See id. at 132.
115 See id. at 32, 132 (studying the influence of various causes on wrongful convictions and then concluding that such influence cannot be quantified or empirically determined).
116 Natapoff, supra note 86, (manuscript at 49).
miscarriage of justice.

CONCLUSION

In my conclusion, I want to suggest that one way of moving beyond the debates like the one between Garrett and Collins and Jarvis is to move beyond the oversimplified framework of discrete causes and begin the more challenging task of thinking about miscarriages of justice as the products of complex interactions between different items of evidence. One possible criticism of Convicting the Innocent is that it is atheoretical—that it is not grounded in a theory of miscarriages of justice. This is a criticism that, as Professor Leo has noted, applies to almost the entire field of the study of miscarriages of justice.119 I agree with Professor Leo that the most promising theoretical approach is that of Lofquist, an approach that is very similar—though with somewhat different terminology and less grounded in sociology—to that of Findley and Scott.120 Lofquist draws on Perrow’s “Normal Accidents Theory” to posit a model of miscarriages of justice that is like a shipping collision.121 Pieces of investigative evidence are like the pieces of data—collected from instruments, vision, hearing, other individuals—that are fed to a ship’s pilot. Sometimes a pilot may form a false theory of where her ship is. When new evidence that might potentially disconfirm that theory is presented to the pilot, rather than revising her theory, she may devise ways of interpreting the new evidence so as to make it consistent with the erroneous theory. As this process continues, an entirely erroneous edifice of evidence is constructed. When the error is finally exposed—and the ship runs into the shoals—lots of evidence that may have seemed to have integrity in isolation turns out to be have been misinterpreted in light of the pilot’s false theory.

Lofquist suggested thinking about miscarriages of justice in a similar way, and some scholars have explored this approach as being particularly applicable to miscarriages of justice involving forensic science.122 If we apply Lofquist’s approach to the debate between Garrett and Collins and Jarvis, we end up with a very different discussion. Our concern is no longer

119 See Leo, supra note 117, at 214.
whether it was “forensic science” or “bad lawyering” that “caused” the wrongful convictions. Instead, because we now know\(^\text{123}\) that the defendants were innocent, we can conceive of forensic evidence as potentially disconfirming evidence that, in this case, failed to dissuade the apparatus of the state—the police, prosecutors, juries, and judges—from their original, erroneous theory of the crime.

There may be a variety of reasons for these failures. In some cases, the forensic technique may simply have lacked sufficient discriminating power, as in the case of a properly conducted and reported serological test that fails to exclude an innocent suspect. Or, a properly conducted forensic test may have been ignored by state actors, either through willful misconduct or, more likely, because it contradicted other evidence in which those actors placed even greater trust. But, in other cases, as Garrett’s research shows, it may be the forensic analyst herself who misinterprets the evidence so that it did not contradict the theory committed to by police and prosecutors, which the analyst had apparently committed to as well. This appears to be what occurred in the cases in which serologists ignored the “masking” issue and interpreted nonprobative or exculpatory results as inculpatory. This is why forensic reformers argue for the “sequential unmasking” of forensic analyses in order that analysts cannot commit to the state’s theory of the crime.\(^\text{124}\)

Rather than asking to what extent forensic science “caused” miscarriages of justice, we might ask to what extent forensic science lived up to its potential to derail impending miscarriages of justice. Even more than other forms of evidence, people think, quite reasonably, that “science” is—or should be\(^\text{125}\)—an independent actor in the criminal justice system. The function of forensic science should be to correct false theories developed by police as well as to support true theories. It therefore seems reasonable to consider the extent to which forensic science functions as an independent check on state actors who put forth erroneous theories of crimes (figures 1-2). Of course, answering such a question empirically will be quite challenging. As a starting point, we would need to know the potential of various forensic techniques to correct false theories of crimes. Even forensic techniques that are considered “unreliable,” like microscopic

\(^{123}\) Well, we do not actually know, but we have very strong reason for believing, and, even if we are wrong about a handful of cases, it is extremely unlikely that we are wrong about many of the 250 post-conviction DNA exonerations.

\(^{124}\) See supra note 81 and accompanying text.

\(^{125}\) “Should be” may be more appropriate than “is” because most crime laboratories in the United States are part of law enforcement agencies—despite the recent recommendation of a National Research Council committee that crime laboratories be independent—and this is unlikely to change. See COMM. OF IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY., NAT’L RESEARCH COUNCIL, supra note 68, at 23.
hair comparison, have some such potential. There are, presumably, innocent suspects who would be excluded by microscopic hair comparison. But we have little information about the discriminating power of all but a few forensic techniques.

We would also need data on the number of times forensic evidence convinced state actors to alter their theory of a crime. To my knowledge, little such data exists. One anecdotal piece of such data derives from the FBI’s early uses of DNA testing. It was reported that DNA testing excluded the primary suspect around one-third of the time.\textsuperscript{126} Indeed, the scarcity of such data may itself be viewed as an indication of the security and confidence forensic scientists feel in their role in the criminal justice system and as an indication of how they conceive that role. If forensic scientists conceived of demonstrating their ability to correct erroneous theories of crimes as a way of demonstrating their value to the criminal justice system, they might have historically made greater efforts to document and count such instances. A more disturbing interpretation of this fact, of course, is that forensic scientists know or intuit that their value to the criminal justice system lies in their ability to support, rather than contradict, the theories of crimes formed by state actors. This is how we might ideally assess the contribution of forensic science to wrongful convictions. Of course, we do not have the data to do such an assessment retrospectively about the post-conviction DNA exonerations or any other set of past cases. So I do not wish to fault Garrett for not adopting an approach for which data was not available. Clearly, Garrett did an admirable job telling us what he could base on the available data. But, perhaps we could collect such data prospectively. Setting up data collection in this manner would also emphasize conceiving of forensic science as an independent check on the theories of crimes formed by state actors.

\textsuperscript{126} Peter Neufeld & Barry C. Scheck, Commentary in EDWARD CONNORS ET AL., CONVICTED BY JURIES, EXONERATED BY SCIENCE: CASE STUDIES IN THE USE OF DNA EVIDENCE TO ESTABLISH INNOCENCE AFTER TRIAL, at xxviii (1996) (noting that out of 8000 sexual assault cases with conclusive DNA test results, approximately 2000 excluded the primary suspect).