Forensic Science Needs Checks and Balances

*By Roger G. Koppl & Radley Balko**

Justice is the foundation of liberty. Thus, the proper functioning of our criminal justice system is a vital concern for all who value liberty. When figures such as David Hume first laid down the classical liberal principles which form the foundation of our criminal justice system, however, scientific evidence did not yet have the important role in criminal cases that it does today. Since the beginning of the twentieth century, that importance has grown so much that forensics is now a central function of the criminal justice system in the United States and in courts around the world.

But our system of checks and balances has not yet been updated to factor in forensic science—in theory or in practice. As a result, crime labs in most jurisdictions have a virtual monopoly over evidence analysis, and this monopoly structure has created needless and unacceptably high error rates. Today, a jury's verdict often turns on forensic evidence alone. This state of affairs, thus, calls into question whether our criminal justice system, as it stands today, can truly be called "adversarial." In many ways, forensic scientists may be more important to the outcome of a case than defense lawyers and prosecutors—which makes the profession's errors all the more alarming. In this article, we document the poor performance of forensic science and propose a system of checks and balances to fix the broken system.

Errors in Forensic Science

Despite the impression one might glean from popular culture, forensic science and medico-legal investigation are far from error-free. Persistent errors have been documented in a variety of forensics specialties, including forensic pathology, fire investigation, bite mark analysis, fingerprint analysis, and DNA typing. A few examples:

In February of this year, Mississippi exonerated two men convicted of two similar murders just a few miles apart. Levon Brooks was convicted of raping and murdering his girlfriend's three-year-old daughter in 1990. Two years later, Kennedy Brewer would be convicted of a remarkably similar crime, the rape and murder of his girlfriend's three-year-old daughter. Brooks was sentenced to life without parole. Brewer was sentenced to death. Both men were convicted almost exclusively on the testimony of Dr. Steven Hayne and Dr. Michael West (Mitchell).

Hayne is a forensic pathologist who has essentially monopolized Mississippi's autopsy business for twenty years. He has testified to performing between 1,200 and 1,800 autopsies a year, an astonishing figure given that the National Association of Medical Examiners recommends an individual doctor do no more than 325. Dr. West is a "forensic odentologist," or bite-mark analyst, who once claimed that he could trace the bite marks in a half-eaten sandwich at a murder scene back to the defendant. The two have long been criticized by Mississippi defense attorneys and medical malpractice attorneys for jiggering their conclusions to support the theories of prosecutors and plaintiffs' attorneys (Balko, "CSI: Mississippi").

In the Brooks and Brewer cases, Hayne performed the initial autopsy, then called in his longtime collaborator West to do "bite mark analysis." In both cases, West said that marks others would call "indiscriminate scratches and bruises" were really human bite marks. In both cases, West said that he could definitively trace the bite marks back to the defendants. In both cases, the jury believed him, and voted to convict (Mitchell).

In February of this year, officials in Mississippi announced that they had arrested a man named Justin Albert Johnson who confessed to both murders. A DNA match confirmed the confession. Brooks had served eighteen years in prison. Brewer had served fifteen—all of them on Death Row.

Cameron Todd Willingham was executed in Texas 2004. Charged with murdering his three small children by arson, he was convicted with forensic techniques that were current at the time of the fire in 1991, but had been discredited by the time of his execution in 2004 (Mills & Possley, 2004). A "key reference text for the Texas fire marshal's office" (Mills & Possley, 2004), found that many then-standard techniques of arson investigation have since been shown to be inaccurate. The report by the National Fire Protection Association was published on February 10, 1992, within two months of the Willingham fire. Some of these bogus techniques were used against Willingham. The presence of "crazed glass," for example, was thought to indicate that an accelerant had been used. It has since been shown that these intricately patterned cracks can also be caused by dousing hot glass with water, which obviously occurs frequently as firefighters attempt to put out fires (Mills & Possley 2004). The *Chicago Tribune* reports that, "Before Willingham died by lethal injection on Feb. 17, Texas judges and Gov. Rick Perry turned aside a report from a prominent fire scientist questioning the conviction." It would be impossible to say whether or not Willingham was guilty, but it is clear that he was convicted on bad science, and his execution was a perversion of justice.

In Houston, Texas problems forced the city's lab to shut down DNA testing from December 2002 to July 2006, during which time police used a private lab instead (Bromwich 2005, Khanna 2006, Glenn 2006). Before the shutdown, Josiah Sutton, was convicted of rape largely on the lab's DNA evidence, which was later shown to be inaccurate. Imprisoned at the age of sixteen, Sutton served four years before he was released (Koppl 2005). A subsequent audit of the Houston lab revealed many problems, including the risk of cross-contamination from the use of a common evidence screening area for trace, serology, and arson; failure to follow procedures for the calibration of equipment or maintain logs of repair and calibration of equipment; lack of procedure for preparing and preserving case notes; sloppy reports; and improperly and ambiguously labeled reagents (FBI Director, 2002). One particularly alarming line from the report stated, "The audit team was informed that on one occasion the roof leaked such that items of evidence came

** Roger G. Koppl is a Professor at Fairleigh Dickinson University and Director of the Institute for Forensic Science Administration. Radley Balko is a Senior Editor at Reason Magazine.*

in contact with the water" (FBI Director, 2002).

Poor DNA work can lead to false exonerations, as well. In 2001, a man in Pacific, Washington with a prior sex-crime conviction was arrested and charged with the rape of his ten-year old niece. DNA tests excluded him, however, and pointed to someone unknown. About two weeks after the original suspect accepted a deal from the prosecutors and pled guilty to the lesser charge of child molestation, the Washington State Patrol's Tacoma crime lab discovered that the original exculpatory result was tainted by cross-contamination (Teichroeb 2004). Presumably, a proper test would have included the suspect, although at this point, it is impossible to say for sure.

Brandon Mayfield is probably the most prominent American case of a false conviction, due to an incorrect fingerprint match. In 2004, the FBI arrested Mayfield as a material witness in the Madrid train bombing of March 2004. He had been identified as the source of a latent print found on a bag of detonators near the crime scene. After assigning three of its top fingerprint examiners to the case, the FBI declared a "100 percent match" to Mayfield. The Spanish National Police objected, however, and declared a match to man named Ouhnane Daoud. The Spanish authorities' suspicions were confirmed when the FBI withdrew its identification and released Mayfield (Office of the Inspector General 2006).

Florida's Seminole County provides a more recent example of erroneous fingerprint analysis. In March 2007, Tara Williamson, a fingerprint examiner for the Seminole County Sheriff's Office in Florida, wrote a memo accusing her co-worker Donna Birks of misbehavior and incompetence. Her accusations seem to have been correct. By June 2007, investigators from the Florida Department of Law Enforcement (FDLE) discovered six cases in which Birks made a positive identification from prints that should have been considered inconclusive, and a seventh case in which she identified someone who should have been excluded. (Such judgments, of course, assume that the FDLE fingerprint examiners have themselves made correct analyses.) Williamson's memo says that Birks "reported numerous identifications without verification," that she "had a trainee with three-weeks of experience verify latent print identifications," and that on one occasion she sought out a third, retired, examiner to verify an identification after two "examiners in the office were not able to verify the print"—or, in other words, disagreed with her analysis. Williamson reports that these actions violated "basic ethical guidelines" governing fingerprint examination (Williamson 2007). Birks had been promoted to latent print examiner in 1998. It is estimated that she worked on about 1,500 cases over the years. As of June 2007, the FDLE was re-examining 300 of those cases (Stutzman 2007a & 2007b, Williamson 2007).

In October 2007, a Maryland court ruled that the standard "ACE-V methodology" of fingerprint examination is not reliable enough for capital cases (*State of Maryland v. Bryan Rose*). The decision cited evidence that prominently included the Mayfield misidentification. The judge concluded "that ACE-V was the type of procedure" Maryland rules of evidence "intended to banish, that is, a subjective, untested, unverifiable identification procedure that purports to be infallible."

Studies outside the courtroom show similar cause for concern. A 1999 workshop conducted by the American Board of Forensic Odontology, for example, asked bite mark experts to match four bite marks with seven dental models. More than six in ten of participants came back with false positives (Bowers).

Fundamental Principles of Forensic Science Administration

The cases reviewed above are not "isolated incidents." One of the authors here has concluded from his reviews of the evidence on error rates in forensics (including proficiency tests and controlled studies) that "forensic analysis is not sufficiently reliable" (2005a). The advocacy group The Innocence Project reports that seventy-four of the 214 cases in which DNA testing has exonerated a wrongfully convicted defendant involved the introduction of faulty forensic evidence by prosecutors. Many scholars, journalists, activists, and others have also recognized the need to improve forensic science.

The three leading proposals for reform are probably independence, masking, and oversight.

Paul Giannelli is the leading figure in favor of independence. In an important article on forensics, he argues that crime labs "should be transferred from police control to the control of medical examiner [ME] offices" (1997). Admirably, Giannelli notes that, although his proposal "is a substantial step in the right direction," it "is not a panacea".

Risinger et al. (2002) call for "masking," whereby "domain-irrelevant information" would be hidden from forensic scientists. Risinger et al. appeal to a large empirical literature in psychology. The point may be best illustrated, however, by an important study by Dror & Charlton (2006), where the co-authors employed experienced fingerprint examiners to analyze evidence from cases they had decided in the past. The subjects did not know they were looking at their own, earlier cases. In half the cases, they replaced the original case information with information suggesting a conclusion opposite to the original judgment. In half, no such contextual information was supplied. The examiners of their study reversed themselves in six of forty-eight cases. Two of the six reversals were from the twenty-four cases in which no biasing information had been given.

Peter Neufeld and Barry Scheck founded The Innocence Project, which has, to date, participated in over 200 DNA exonerations of persons wrongly convicted. An important figure in any discussion of how to improve forensic science, Neufeld has argued that "[g]overnment oversight and the creation of independent academic centers to validate technologies and techniques, encourage best practices, and enforce appropriately cautious standards for the interpretation of data could dramatically enhance the reliability of forensic science and engender greater public confidence in the outcome" (2005). Neufeld's plea for more scientific research is proper, but beyond scope of this article. His call for oversight is representative of the "repeated calls" for "oversight" noted in a 2003 *Science* editorial (Kennedy 2003).

Students of public choice theory will recognize an important problem with "oversight:" *Quis custodiet ipsos custodes?* Who will guard the guardians themselves? Koppl

(2005) identifies eight remediable features of the current institutional structure of forensic science, each of which reduces reliability. Corresponding to each flaw in the institutional structure is a suggestion for amending current institutions. The proposed suit of reforms is "competitive self regulation." Table 1 summarizes the argument of Koppl (2005).

The key to the proposal is rivalrous redundancy. In the current system, once forensic evidence has been sent to one lab, it is unlikely that another lab will review the same evidence. In this sense, each lab has a monopoly on the analysis of evidence it receives. In pure science, no lab enjoys such a monopoly. Rather, the results of any one lab may be challenged by any other. In forensic science, however, this scenario is unlikely, and this radical difference in network structure may help explain the difference in reliability that seem to exist between the two fields.

Competitive self-regulation creates a salutary rivalry among crime labs. In part, it reduces error rates by making

Table 1: Proposals of Koppl (2005) in Tabular Form

Current System	Resulting Problem	Proposed Institutional Change	Explanation or Comment
Monopoly	Sloppy, biased, and sometimes fraudulent work	Rivalrous redundancy	There should be several competing forensic labs in any jurisdiction. Subject to the constraints of feasibility, some evidence should be chosen at random for duplicate testing at other labs. The same DNA evidence, for example, might be sent to more than one lab for analysis. The forensic worker need not know whether the evidence is examined by another lab. He will know that there could be another lab, and sometimes is.
Dependence	Bias	Independence	Crime labs should be independent of police and prosecutors.
Poor quality control	Persistently poor work	Statistical review	Statistical review would support improved quality control. For example, if a given lab produces an unusually large number of inconclusive findings, its procedures and practices should be examined.
Information sharing	Conscious and unconscious bias	Information hiding	Evidence should be prepared for testing so as to shield the lab doing a test from all extraneous knowledge of the case particulars.
No division of labor between forensic analysis and interpretation	Error from false interpretations of legitimate results	Division of labor between forensic analysis and interpretation	When this measure is combined with the provision of forensic counsel for the defense, errors of interpretation are less likely to go unchallenged.
Lack of forensic counsel	False convictions	Forensic counsel for the indigent	Forensic science decides many criminal cases and yet we do not have a right to forensic counsel similar to our right to legal counsel.
Lack of competition among forensic counselors	Poor quality forensic counsel	Forensic vouchers	A voucher system would give forensic counselors to the indigent an incentive to provide high-quality services to their clients.
Public ownership	Weak financial incentives to provide high-quality work	Privatization	Unlike public labs, private labs would be subject to meaningful fines and civil liability. In the US, the federalist structure of government means federal regulation and oversight are easier when labs are private.

fraud and corruption more difficult. It does so, however, by providing each actor in the system with an external epistemic check currently lacking.

The case of the FBI forensic scientist Jacqueline Blake is a fine illustration. For over two years, Blake systematically failed to run her negative controls when performing PCR/STR DNA analysis. The negative control tests the reagents and equipment used in such analyses, but without including a DNA sample.

It is difficult to guess Blake's true motives. The OIG report on her case, however, seems to suggest that she was a well meaning person who was simply not up to the job. "Some Laboratory employees have speculated that the reason that she failed to process the negative controls was because she lacked confidence in her ability to master PCR/STR testing" (OIG 2004). In this case, as with the Houston Crime Lab, we find a crime lab whose failures went undetected for years because there were no external epistemic controls of the sort we take for granted in pure science.

Events in the Seminole County fingerprint scandal reinforce the point. Recall that fingerprint examiner Tara Williamson wrote a memo accusing Birks of misconduct and incompetence. The investigations initiated in response to that memo revealed that she had made her own errors, including improper verifications of Birks's work (Stutzman 2007b). Williamson has been demoted to dispatcher and is no longer given fingerprint work. It seems clear that she was a well-motivated individual. But good intentions did not prevent her from making errors. Error reduction requires that each forensic lab be subject to an external check, without which even highly motivated actors may unwittingly commit repeated errors.

Redundancy is Cost-Reducing

Redundant testing would seem to be a costly suggestion. Where we now have one crime lab, shall we build three? A closer looks, however, shows that redundancy would *reduce* the taxpayer cost of administering the criminal justice system. One of the authors explains in his forthcoming work (reference below) why rivalrous redundancy would require little or no increase in our basic forensics infrastructure. No grand capital expansion is needed; the central point being the low cost of forensic tests relative to the costs of forensic error.

Using 2002 data, the author (forthcoming) estimated the cost of adding two redundant fingerprint examinations (for a total of three) to each felony case with fingerprint evidence which goes to trial. The average felony sentence in 2002 was about five years. The cost of incarcerating a prisoner was about $20,000 per year. Even discounting future values to calculate a present value, the costs of incarceration for a false felony conviction were about $100,000 in 2002. This value is 1,000 times greater than the $100 cost of two fingerprint tests. The imagined redundancy would eliminate almost all false positive errors in fingerprint examination. Thus, this form of redundancy would save money if the false positive error rate in fingerprint is anything over one in a thousand, or 0.1%.

(The break-even point in the study was 0.115%.) The true rate of false positive errors is likely to be at least 0.8%, and probably more. The author's calculation is thus extremely conservative because it counts only the taxpayer cost of incarcerating the wrongfully convicted, ignoring all other aspects of the social cost of putting the wrong person in jail.

CONCLUSION

In sum, the authors believe that competitive self-regulation would eliminate most errors in our criminal justice system, while reducing taxpayer costs. Paraphrasing, and amending, Madison in *Federalist* 51:

> Competitive self-regulation supplies, by opposite and rival interests, the defect of better motives and greater wisdom. It makes each lab a check on the other, ensuring that the private interest and understanding of every lab may be a sentinel over the public rights.

It is important that we act effectively to ensure not only the reliability of forensic science but also continued public trust in the most vitally scientific element of our criminal justice system. For that to happen, we must tear down the monopoly structure, which has given us needlessly high error rates, and bring forensic science within the fold of our system of checks and balances.

References

Balko, Radley "CSI: Mississippi," *Reason*, October 2007.

Bowers, Michael. 2006. "Problem-Based Analysis of Bitemark Misidentifications: The Role of DNA," Forensic Science International, 159 (supplement 1): S104-s109.

Bromwich, Michael R. 2005. *Second Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room*. Downloaded on 24 May 2007 from http://www.hpdlabinvestigation.org/reports/050531report.pdf

Dror, Itiel and David Charlton. 2006. "Why Experts Make Errors," *Journal of Forensic Identification*. 56(4): 600-616.

FBI Director. 2002. "An Audit of Houston Police Department Crime Laboratory-DNA/Serology Section," December 12-13, 2002.

Giannelli, Paul C. 1997. "The Abuse of Evidence in Criminal Cases: The Need for Independent Crime Laboratories," *Virginia Journal of Social Policy & the Law*, 4: 439-478.

Glenn, Mike. 2006. "Under the microscope: HPD's evidence closely watched this time around," *Houston Chronicle*, 11 July 2006. Downloaded on 24 May 2007 from http://www.chron.com/disp/story.mpl/special/crimelab/4038116.html.

The Innocence Project, "Forensic Science Misconduct," http://www.innocenceproject.org/understand/Forensic-Science-Misconduct.php, accessed on March 10, 2008.

Kennedy, Donald. 2003. "Forensic Science: Oxymoron?" *Science*, 302: 1625.

Khanna, Roma. 2006. "Crime lab's DNA section, shut in '02, will reopen," *Houston Chronicle*, 21 June 2006. Downloaded on 24 May 2007 from http://www.chron.com/disp/story.mpl/special/crimelab/3988198.html.

Koppl, Roger. Forthcoming. "Romancing Forensics: Legal Failure in Forensic Science Administration," in Lopez, Edward, edited, *Government Failure in the Legal System: A Public Choice Review of the Law*, Independent Institute.

Koppl, Roger. 2005. "How to Improve Forensic Science," *European Journal of Law and Economics*, 20(3): 255-286.

Mills, S. and M. Possley. 2004. "Texas Man Executed on Disproved Forensics: Fire that Killed his 3 Children could have been Accidental," *Chicago Tribune*, 9 December 2004.

Mitchell, Jerry. "Innocence Project Letter Blasts Steven Hayne's Work," *Jackson Clarion-Ledger*, February 14, 2008.

Neufeld, Peter. 2005. "The (Near) Irrelevance of *Daubert* to Criminal Justice and Some Suggestions for Reform," *American Journal of Public Health*, 95(supplement 1): S107-S113.

Office of the Inspector General, United States Department of Justice. 2006. *A Review of the FBI's Handling of the Brandon Mayfield Case: Unclassified Executive Summary*. Available at http://0225.0145.01.040/oig/special/s0601/PDF_list.htm.

Office of the Inspector General, United States Department of Justice. 2004. *The FBI Laboratory: A Review of Protocol and Practice Vulnerabilities*, Available at http://www.usdoj.gov/oig/special/0405/final.pdf.

Risinger, Michael, Michael J. Saks, William C. Thompson, and Robert Rosenthal. 2002. "The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion," *California Law Review* 90: 1-56.

State of Maryland v. Bryan Rose, Memorandum Decision, 19 October 2007, Circuit Court for Baltimore County, K06-545.

Stutzman, Rene. 2007a. "Are Innocents Imprisoned? Fingerprint Errors Found," *Orlando Sentinel*, 4 May 2007.

Stutzman, Rene. 2007b. "Fingerprint scandal costs analyst her job," *Orlando Sentinel*, 7 June 2007.

SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology). 2002a. "Friction Ridge Examination Methodology for Latent Print Examiners." Downloaded from http://www.swgfast.org/Friction_Ridge_Examination_Methodology_for_Latent_Print_Examiners_1.01.pdf on 11 June 2007.

SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology). 2002b. "Quality Assurance Guidelines for Latent Print Examiners." Downloaded from http://www.swgfast.org/Quality_Assurance_Guidelines_for_Latent_Print_Examiners_3.0.pdf on 11 June 2007.

Teichroeb, Ruth. 22 July 2004. "Rare Look Inside State Crime Labs Reveals Recurring DNA Test Problems," *Seattle Post-Intelligencer*.

Williamson, Tara. 12 March 2007. Memorandum to Captain Randy Pittman. Available from the author.