

BIOMETRICS PUBLICATIONS

DAUBERT HEARING ON FINGERPRINTING

WHEN BAD SCIENCE LEADS TO GOOD LAW:

THE DISTURBING IRONY OF THE DAUBERT HEARING IN THE CASE OF
U.S. V. BYRON C. MITCHELL

Dr. James L. Wayman, Director
U.S. National Biometric Test Center
College of Engineering
San Jose State University
2/2/00 Version

In my opinion, if a significant portion of one of your fingerprints is found at a crime scene, you had better be able to; 1) explain its presence; or 2) prove you were already in jail at the time the crime was committed. But I'm a scientist, not a fingerprint examiner, so I'm not paid for my opinions on these matters. Rather, I'm paid to apply the tools of science to test hypotheses such as, "No two individuals have any fingerprints, or portions of any fingerprints, in common". Proving or disproving this is really hard, because we scientists don't have access to all fingerprints from all the world's people. Consequently, we may have to use "statistical estimation". By using the word "statistical estimation", instead of the more realistic word, "mathematically-based guessing", we're hoping that most people will treat us with authority, like people used to treat physicians who actually made house calls, and not dispute these guesses. Certainly, statistical theory, when carefully and scientifically applied, can illuminate great areas of knowledge. But the forms and terminology can easily be misapplied to disguise crazy guesses and opinions. If you are a judge or serving on a jury, and I am an expert witness, I might be able to disguise my guesses with enough bogus "statistical estimation" techno-speak that you won't question them at all, even if they're absurd.

Before we can apply this erudite "statistical estimation" to fingerprinting, we must sharpen the hypothesis. In this case, exactly what do we mean by the words "fingerprint", "portion" and "in common"? "Galton ridges" are the line-like structures on the skin of the palm side of the finger past the distal (the last) joint. These structures may also include pores and will show signs of cracking, abrasion and scarring, depending upon how rough we have been on our hands recently and over the years. So the appearance of these structures is changing over time on all of us. Except on cadavers, scientists don't actually have these Galton ridges to compare and experiment with, only approximate images of these structures, called "fingerprints", perhaps acquired by rolling an inked finger on paper, or better yet, with an electronic scanner of limited resolution. So now our hypothesis is, "No two individuals can have any fingerprint images, or portions of any fingerprint images, in common at any single time".

We still haven't defined the words "portion" and "in common". The lack of a precise meaning for these terms, and the gross misuse of "statistical estimation" leading to

absurd guesses about the likelihood of an error, are the central problems with the recent government testimony in the Daubert hearing in the U.S. v Byron C. Mitchell case. This hearing took place in September in U.S. District Court in Philadelphia. Putting aside, for the moment, the problem of defining “portion” and “in common”, our hypothesis about fingerprints can easily be proved false: If the images are bad enough and the portions small enough from places outside the center of the fingerprint (perhaps only tiny segments of a couple blurry ridges), my images will be “in common” with almost anybody’s. This extreme case can be established in our lab. Using good quality images of reasonable size and finger positioning, however, we have done tens of millions of computer comparisons with exceedingly few errors, all which could be resolved by human inspection. The scientific question addressed by the government in the Daubert hearing for the Mitchell case should have been, “What is a reasonable estimation of the chance of an error when comparing fingerprint images of reasonable size, position and quality?”. The answer, based on sound science, could have been, “Reasonably low”. Unfortunately, the government’s answer, disguised in the forms and terminology of “statistical estimation”, was absurd.

Daubert v. Merrill Dow Pharmaceutical

The Daubert and Schuller families sued Merrill Dow Pharmaceuticals, claiming that the pre-natal use of a prescription drug had caused their children to be born with serious birth defects. The lower courts had ruled that scientific arguments presented by the families to show that the defects were caused by the drug did not meet the required criteria of “general acceptance” for expert evidence. The U.S. Supreme Court was asked to rule on the requirements for presentation of “scientific” evidence into a court of law. In their 1993 decision (509 U.S. 579), the court found that five conditions should be met for evidence to be admissible as “scientific”:

1.
The theory or technique has been or can be tested.
2.
The theory or technique has been subjected to peer review or publication.
3.
The existence and maintenance of standards controlling use of the technique.
4.
General acceptance of the technique in the scientific community
5.
A known potential rate of error.

Trial judges still retain some discretionary power over what scientific evidence does and does not get presented in a trial. Justice Blackmun, writing for the unanimous Court said, “...the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable”. Justice Rehnquist, although voting with the rest, dissented on this particular point, worrying that the court should not impose on judges “...either the obligation or the authority to become amateur scientists in order to perform that role”. So now the above five requirements are the “law of the land” and must be met if evidence is to be introduced into any trial as

“scientific”.

U.S v. Byron C. Mitchell

In 1998, Byron Mitchell was arrested for robbery. The arrest was supported by the apparent match of his fingerprints with small portions of two fingerprints found on the getaway car. His public defenders argued that fingerprint comparison techniques did not meet the five criteria for admissibility established by the U.S. Supreme Court in the Daubert decision, particularly the fifth: that the potential rate of error is known. The Mitchell defense petitioned the court for a Daubert hearing to determine the admissibility of fingerprint match as “scientific” evidence. The government defense of fingerprinting was lead by the U.S. Department of Justice with assistance of government contractors. The hearing began in July of this year.

The Government’s “Statistical Estimation”

Mitchell’s fingerprints had been matched by fingerprint experts. There are no data available on the error rates of these experts, but they are widely acknowledged to be very low. Arranging for a test of a suitable size to reveal even one error would be very expensive and time consuming, so the government proposed testing a computer fingerprint matching system instead. Because these systems do not seem to perform as well as humans, substituting a computer for humans will lead to a higher error estimate, but such “conservative” estimates do make for good science.

To establish an estimate of the chance of an error by the computer system, the government concocted two tests. In the first test, 50,000 fingerprint images were compared to each other. That is, each of the images was compared to all other images, including itself. In computer fingerprint systems, a comparison of fingerprint image A to fingerprint image B leads to a different “score” than the comparison of the prints in reverse order (B to A). Consequently, these 50,000 data points lead to about 2 1/2 billion comparisons. The comparison of images to themselves lead, of course, to extremely high scores, which researchers called the “perfect match” score. Because in life fingerprints are always changing, no real comparison of two different images of the same finger will ever yield such a high score. By adopting, as the definition of “in common”, the score obtained by comparison of identical images, the government very strongly biased any results in the government’s favor.

Now the government did something even worse: They looked at all the scores between different fingerprint images and declared them to follow a “bell curve”. There are potentially an infinite number of curves that could fit the data, some better than others. There are simple tests available to show if the “bell curve”, or any other curve, roughly fits the data. No such tests, which might have eliminated the “bell curve” assumption, were performed, however. Now, the government simply pulled out a college-level textbook on statistical estimation and, based on the “bell curve” assumption, found the probability of two different prints being “in common”, as previously and unreasonably defined, to be one in 10⁹⁷. This number, 10⁹⁷, is extremely large. We have no word for this number in any language, as it is beyond

human comprehension. In the entire history of mankind, there have been only about 1011 fingerprints. It is possible that in the entire future of all mankind there will never be 1097 fingerprints. Yet, the government is comfortable with predicting the fingerprints of the entire history and future of mankind from a sample of 50,000 images, which could have come from as few as 5,000 people. They have disguised this absurd guess by claiming reliance on “statistical estimation”.

There was an additional logical problem that the government needed to address: The crime scene fingerprint images, called “latent prints”, showed only a small portion of the finger. So to test the error rate for latent prints, the government researchers artificially cropped the size of the original 50,000 images, in effect changing the position of the finger in the images. The precise way in which this is done could have profound impact on the projected error rates, but the government doesn’t reveal exactly their method. Further, the latent prints in the Mitchell, or any police, case would have been naturally “cropped” in a completely different way. The government’s laboratory research gets quite sketchy at this point, but in court, the government claimed error rates between 1 in 1027 and 1 in 1097, presumably using the same flawed methodology as in the first test. The government did not try to run any real crime scene prints against the same 50,000 database to determine comparison scores and establish error probabilities for latent prints in real cases.

In short, nothing in the government study or testimony gives us any indication of the likelihood that the crime scene fingerprints were falsely identified as belonging to the defendant Mitchell or, more broadly, that any latent fingerprints might be falsely identified. In my opinion, the government failed completely to answer the fundamental question, “What is a reasonable estimation of the chance of an error when comparing fingerprint images of reasonable size, position and quality?” They could have done so simply by designing better experiments. If we had a good answer, we’d only have to establish that the crime scene prints in the Mitchell, or any, case were of reasonable size, position and quality to roughly estimate the possibility of error.

In fact, false fingerprint matches are not unknown and have been introduced as faulty evidence in criminal trials. See www.onin.com/fp for details of such occurrences in Illinois and Scotland.

Probability and Statistical Estimation in Legal Cases

There is a history in American jurisprudence of human identification based on the gross misuse of statistical and probability theory. In the famous 1968 *People v. Collins* case, Malcolm and Janet Collins were convicted of robbery based on the testimony by a college math instructor that the chances of some other couple committing the crime was 1 in 1.2 million. The decision was reversed by the California Supreme Court on the grounds that the probability-based arguments were without foundation, and erroneous and misleading to the point of distracting the jury. Writing about the case in 1969, University of Houston Law Professor Alan D. Cullison, states “...it would be unsound for courts to reject expert probability testimony on the basis of the invalidity of probability theory itself...A more cogent

basis for broadside objection to expert probability testimony is that the applications of probability theory to fact-finding problems in law cases have in the past been, crude, misleading and often just plain erroneous.”

More recently, questionable use of probability theory in human identification has involved forensic DNA analysis. Referring to disagreements in National Research Council (NRC) studies of DNA analysis error rates, UC Berkeley Statistics Professor Peter Bickel (current chair of the NRC’s Committee on Applied and Theoretical Statistics and a member of the National Academy of Science) writes in the Proceedings of the National Academy of Science,

The existence of two reports (1992 and 1996), close in time, which disagree on aspects of methodology illustrates what scientists have always known but what the law sometimes wishes to ignore: that scientists can differ in their expert judgment of the accuracy of numbers predicted from data by model-based formulae. In this case, the focus of the disagreement is on the question of the extent to which models of population genetics can be applied in estimating the probability that the DNA of the suspect and DNA found on the victim match perfectly at each and every one of the preselected set of loci. This probability has to be computed under the assumption that the match occurred “by chance alone”. That assumption is not enough to allow us to compute or rather estimate this probability. To finally arrive at a formula, further assumptions are made: treating the FBI and other databases effectively as random samples from the relevant population and, more significantly, that (certain statistical independence assumptions) are satisfied or are perturbed in a correctable way. Given that no laboratory error has been committed, there is, I believe, little disagreement between the committees or within the scientific community that the match probabilities referred to above are small, typically of order smaller than 1 in 1,000. But many scientists would not agree that the modeling assumptions made above can be verified to hold so precisely that the match probabilities can be ascertained to an order of 1 in a billion.

Prof. Bickel’s arguments about error rate estimation in DNA analysis apply equally well to our discussion of fingerprinting. I am of that group that do not agree that the required assumptions about fingerprints hold so precisely that error rates on the order even of 1 in a billion can be ascertained, let alone 1 in 10⁹⁷.

Conclusions

In September, the U.S. Court of Appeals released their findings in the Daubert hearing of the U.S. v. Mitchell case, holding that fingerprinting meets the necessary criteria for admissibility as evidence. This is the correct decision. Fingerprinting is an established science, subjected to peer review and publication, with general acceptance and standards for its practice. Error rates are difficult to measure, precisely because they are so low. So I am pleased with the outcome. I’m saddened, however, that the government’s case had to rest on such shoddy science. I’d certainly prefer to see good law resulting from good science. We must strive to do better.