

CHAPTER 14: SCIENCE AND THE CRIMINAL JUSTICE SYSTEM



There has been a landslide of problems exposed in the criminal justice system that stem from faulty science being used to convict the innocent. There are too many examples to dismiss them as exceptions, or as technical problems that apply only to specific methods. Instead, there are basic, fundamental problems in the way that scientific data are being gathered, used, and presented in our courts. It is a crisis. It means that you can live a righteous life, abiding by the law, and yet become a victim of a faulty prosecution. It also means that the real criminals are going free, to commit more crimes and spread the word that they can get away with it.

Introduction

On paper, the foundation of the US criminal justice system appears to be a triumph for the innocent. Concepts such as “innocent until proven guilty,” “a jury of peers,” and “the right to hear accusers” are safeguards to protect us, with the roots of these concepts stemming from the very inception of this country. Those of us who have never had a brush with the criminal court system can be lulled into a sense that it is functioning well. Sure, we all know that money can be important in getting a good defense, but the premise is that the innocent are rarely faced with having to defend themselves (an ironic reversal of our supposed “innocent until proven guilty” principle). Yet, for those of us with that perception, there has been a shocking embarrassment of wrongful convictions and of abuses and misuses of science that have come to light recently. Most of the information about wrongful convictions and the causes thereof has been revealed by the Innocence Project, but there are many other sources as well.

In the last decade or so:

- 1) Over 258 people in US prisons have been released after new evidence showed that they could not have committed the crime (more than 80 of these were from death row, of the approximately 6,000 people sent to death row since 1976). The most common cause of mis-conviction was mistaken identification by eyewitnesses.
- 2) Hair-matching methods, often used in court to establish that a single hair came from a particular person, have been declared nonsense. In 2003, Canada not only abandoned use of (non-DNA) hair matching methods, but began reviewing convictions that used hair matching, to see if the convictions should be overturned..
- 3) Fingerprints fell from grace. Long considered the icon of personal identification, fingerprint experts were finally subjected to proficiency tests in the mid-late 1990s and found to make false matches 10%-20% of the time.
- 4) Polygraph tests, widely used by the government though no longer allowed in many courts, were declared nonsense by a panel of the National Research Council
- 5) In a flurry of public attention in 2003, the Houston Crime lab lost its accreditation for DNA typing because of sloppy procedure

The list goes on, but these are some of the major ones. We owe many of these new revelations to DNA typing, because DNA typing exposed many of the mis-convictions, and also because it set high standards for the use of scientific data in court. As noted in the preceding (DNA) chapter, DNA typing was initially pursued enthusiastically by prosecutions but challenged by a group of scientists who felt (with some justification) that it was not being used correctly. No one on either side of the argument in those early days seemed to foresee the huge impact DNA typing would have in exposing the long history of bad science used in criminal courts.

Before delving into specific examples, we can summarize the main problems in the context of our ideal data template:

1. **Failure to gather and analyze data blindly:** This is probably the most pernicious violation of ideal data. Prosecution agencies and their consultants (labs) know whether there is a prime suspect, who it is, and know the identities of the samples being analyzed. As evidence begins to form around a suspect, it allows the entire process to continually reinforce the apparent guilt of that suspect by biasing the gathering and selection of evidence toward that person and away from others. There are many documented cases of this bias and much of it stems from a lack of blind protocols. This bias is difficult to correct, however, because many aspects of prosecutorial duties cannot be done blindly.

Since the routine use of DNA testing was implemented, 25% of the time the prosecution's prime suspect has been cleared by DNA before trial. This means that the prosecution's initial stage of gathering evidence led them to the wrong person. Because of the biases built into the prosecutorial practices, many of these 25% would have been convicted if DNA typing had not been available.

2. **Bad standards:** These problems include (i) failures to conduct blind proficiency tests of the labs, and (ii) inadequate (sometimes non-existent) reference databases.
3. **Bad protocols:** In some cases, methods have been used widely despite the lack of protocols for analyzing the data. In other cases, protocols for gathering the data have been inadequate for protecting against human and technical error, but challenged by a group of scientists who felt (with some justification) that it was not being used correctly. No one on either side of the argument in those early days seemed to foresee the huge impact DNA typing would have in exposing the long history of bad science used in criminal courts.

Ideal Identification Methods

Critical to many if not most criminal trial is some kind of (physical) evidence linking the suspect to a crime or crime scene. This evidence may consist of DNA, hair, eyewitness accounts, fingerprints, shoe prints, and so on. There are some features of any identification method that render it suitable for scientific inquiry:

Template of an ideal identification method:

FEATURE	WHY NEEDED	ERROR RED'N PRINCIPLE	FLAGS- INDICATORS OF ABSENCE
1) reference database	gives the population frequencies of the different characteristics, thus knowing the RMP	= 'standard' to calculate rate of a wrongful match (random match)	not mentioned or claims that match is unique
2) characteristics measured are discrete	it is clear whether a person has it or not, allowing consistent scoring	no RPA error	no description of specific characteristics
3) independent verification a) universal protocol b) characters permanent	someone else can challenge the conclusions	= replication, to detect many types of error	methods of one expert cannot be evaluated by another; no explicit protocol; characters being measured are not permanent
4) labs subjected to blind proficiency tests	provides assurance of the accuracy of the methods	= standards to estimate overall error rate	no error rate given; tests internal, not blind, undocumented

We now consider some specific examples of forensic methods that have been used to identify people over the last 50 years in U.S. courts. The following table summarizes how well they fit the ideal template (and whether the method has been discredited):

Summary of Identification methods and characteristics:

METHOD	(1) REFERENCE DATABASE	(2) CHARACTERISTICS MEASURED ARE DISCRETE	(3) INDEPENDENT VERIFICATION POSSIBLE	(4) LABS SUBJECTED TO BLIND PROFICIENCY TESTS	DISCREDITED
DNA	+	+	+	+	No
fingerprints pre-1990	+	-	-	-	Yes
fingerprints post 2000	+	+	+	+	No
hair matching					Yes
bite marks	-	-	-	-	In some cases
shoe print ID	-	-	-	-	Yes
bullet lead	?	-	-	-	Yes
dog sniffing	-	-	-	-	Yes?
eyewitness	-	-	-	-	

Detailed Discussion of Methods

Hair Matching: Bad Protocols, Bad Standards

Once the most trusted method of identification in forensics, fingerprint matching has been shown to have major problems.

The year 1911 first successful introduction of fingerprint evidence in US court, but not as the sole evidence. Thirty years later, a legal precedent was established for convictions based on fingerprint evidence alone. Although the uniqueness of a person's fingerprints was originally established for the complete set of fingerprints from all 10 digits, somewhere around this time or later, there was acceptance of the general assertion that a single fingerprint was also unique and could be used to establish identity.

Surprisingly, the main international association of fingerprint experts (which, incidentally, consisted mostly of US experts) resisted the establishment of criteria for demonstrating a match into the 1990s. That is, they refused to accept an analytical protocol for declaring a match. They instead proclaimed that each decision about a match was to be made on a case-by-case basis and should be left up to the expert reviewing the case. (There was disagreement about this point between the two main fingerprint organizations, and the British adopted a minimal set of criteria for declaring a match.)

In the years 1995-8, there were 4 voluntary proficiency tests offered to different fingerprint labs. These involved multiple fingerprint comparisons. Not all labs responded, but for those that did respond, the false positive error rate of labs was at least a few percent and was as high as 22 percent!

Hair Matching: Bad Protocols, Bad Standards

Hair matching was put to rest in 2003, both in the U.S. and Canada. In hindsight, there were many major problems with it that should have kept it from ever seeing the light of day. Specifically, there were

1. no data banks for hairs
2. no way of coding hairs
3. no protocols for analysis.

It should thus not be surprising that a full 18 of 62 wrongful convictions listed by the Innocence Project involved hair matching. What is astonishing is that hair matching was used for so long: proficiency tests from the early 1970s had found error rates of 28%-68% when labs were asked to match hairs, and different labs made different mistakes (as expected if there is no uniform protocol for doing the matches).

Dog Sniffing Identification:

Effectively a method lacking in protocols. There is no way to know what method a dogs is using for odor identification and matching. Tests using trial dogs have found that dogs are not very good at matching odors from different parts of the same person (e.g., hand versus neck).

Polygraph: Bad Protocols, Bad Standards

A report released 8 October, 2002 by the National Academy of Sciences described polygraph testing as little more than junk science. Although a 1988 federal law banned the use of such tests for employment screening in most private businesses, and polygraph data had been inadmissible in nearly all state courts, the method has been widely used in government agencies concerned with national security. There was a time when polygraph data were used in court, and the polygraph has been used for unofficial purposes in criminal investigations to help prosecutors decide who to rule out as suspects. Thus, the fact that it has been inadmissible in court has not prevented it from assuming an important role in criminal investigations.

Interviews With Suspects: Bad Protocols

Interviews have commonly not been videotaped or transcribed, so accounts of what was said have been based on recollections; the conduct of interviews has also been variable. Nonetheless, law enforcement officials often use their “recollections” of what a suspect said during an interview. Claims of confessions or incriminating statements may thus have been in error. Information may also have been passed to the suspect that was then used to indicate intimate knowledge of the crime. The use of physical force during interviews was banned by the Supreme Court only by 1936.

Eyewitness Identification: Not Blind, Bad Protocols, Bad Standards.

This is the most baffling of all evidence used in courts. Eyewitness identification of a suspect is the most powerful evidence there is for swaying a jury. And it is among the most fallible of all evidence: 52 of 62 wrongful convictions tabulated by the Innocence Project involved mistaken ID, the most common error attributed to a wrongful conviction.

It has been known for over a century that eyewitness accounts are less than perfect. A 1902 experiment conducted in class (involving a gun – not something we'd do now) revealed that the best witnesses were wrong on 26% of important details; the worst had an 80% error rate. In a more recent California State University experiment with a staged attack on the professor, only 40% of the class later identified the attacker, and 25% attributed the attack to a bystander.

Errors by eyewitness testimony in court have been documented for decades, and some of them are profound. There have been several cases in which half a dozen or more eyewitnesses identified the same person, and it was the wrong person. In many cases, there is not even a close resemblance between the right and the wrong person.

How can this happen? How can many people independently arrive at the same wrong identification?

Let's start with a single eyewitness. Some psychologists use a 3-sequence model of memory:

1. **acquisition** – the events are recorded in your brain
2. **retention** – the acquired events remaining in your brain are lost at some rate
3. **retrieval** – the events are recalled by you

The acquisition phase is known to be imperfect – you never record all aspects of an event. That is, your memory starts out like a photograph with gaps. The more distracted or stressed you are at the time, the less you acquire. Thus, a person being raped or facing a gun/knife will have a much faultier acquisition than a person facing a non-threatening situation.

The retention phase is also imperfect. However, not only can you lose the memory of an event you had once acquired, you can also add events that never happened. Your memory is dynamic, and you are constantly building it, often filling in the holes. This rebuilding of the memory is where problems arise with eyewitnesses. In particular, your memory is very prone to subtle suggestions which eventually bias what you remember. Here are two problems that confound witness identification.

- Subtle hints can influence a witness to choose a particular suspect, even if that suspect is not the right one. These hints can be as innocuous as the police merely asking the witness to “take a careful look at #3” for example. Anything that makes one suspect stand out from the others can be a subtle influence on the witness (the way a picture of a suspect is taken, what they are wearing, etc.). Bad protocols and an absence of blind testing contribute to this bias. (The absence of blind exists if the police are influencing the witness and know who is their preferred suspect). The witness should not experience any outside factors that will push their choice toward a particular suspect.

- Familiarity transfers from one setting to another. If an eyewitness has had some previous exposure to a suspect in a setting unrelated to the crime, it is common for that familiarity to be transferred over in the witness's mind to the new context. In one case, the witness mistakenly identified a man who lived on her block (but whom she had only seen at a distance on a couple of occasions). In an experiment listed above, 25% of witnesses chose the person who had been a bystander to the crime – no doubt because that person was familiar to them, even though not committed the act. Another effect of familiarity is also problematic. Once a witness has been exposed to a set of suspects, any subsequent exposure to one of the suspects reinforces the witness's memory of THAT suspect. For example, if a witness is shown two lineups (at different times), and one suspect is common to both lineups, that suspect is likely to be chosen because of the familiarity. This problem is difficult to eliminate completely, because the familiarity may have been obtained before the witness saw the crime, so police procedure could not, in that case, prevent it.

These are the problems that one witness can experience, even without knowing it. But how can several people all make the same mistake? When several people all make the same mistake, it is a clear indication that police protocols are bad – it generally means that the police are influencing witnesses, or witnesses are influencing each other.

Lab Tests: Lack of Blind

Although some of the specific examples mentioned above stem from lab errors, there are generic problems that apply to some labs, independent of the method of analysis. Perhaps the most widespread problem is a lack of blind analysis. Knowing which samples belong to which people allows fabrication of evidence (of which there have been many cases (Fred Zain's string of fraud resulted in so many convictions that prosecutors sought him out, as did the news program "60 minutes" when he was exposed). For honest technicians, the absence of blind encourages honest mistakes and selective replication of results (you repeat the test if the results don't fit your preconceived ideas about guilt). And as revealed in the Castro Case (early DNA testing), the lack of blind causes people to over interpret results and make them fit preconceived notions.

As an example of the absence of blind analysis by labs, here are two letters sent from the Chicago Police Department to the FBI, requesting DNA typing. All names are omitted from our text; where names were included in the letters, a description is given in square brackets [].

Letter 1: From Chicago Police Crime Lab to F.B.I. DNA Laboratory Division, 10 August, 1989

Dear [name of Director of F.B.I. lab],

I am writing to request that DNA typing be performed on several items of serological evidence. The names of the people involved are: [name of female victim] F/W (the victim) and [name of male suspect] M/B (the suspect). The evidence I am sending you consists of the following:

- Blood standard from [name of victim]
- Blood standard from [name of suspect]
- Extract from swab
- Extract from victim's pants
- Extract from victim's bra

All three of these extracts were found to be semen/spermatozoa positive and the two extracts from the clothing were found to have ABO, PGM and PEP A activity consistent with that of the suspect. I am also enclosing a copy of my laboratory report stating these results.

The facts of the case are that on 25 May 1989, the victim was grabbed from behind, pulled into the woods and sexually assaulted. The victim never got a good look at her offender and therefore is not able to make a positive I.D. of the suspect. The suspect [name] had just been released from the ILLINOIS DEPARTMENT OF CORRECTIONS after serving time for the same type of crime in the same area. At this time the suspect has not been charged.

Thank you very much for your assistance in this matter. Please feel free to contact me if you need more information.

Letter 2: From Chief of Detective Division, Chicago Dept. of Police to F.B.I. DNA lab

Dear [name, Commanding Officer, F.B.I. DNA lab],

In early January, 1990, detectives assigned to the Chicago Police Department's Detective Division, Area Three Violent Crimes Unit were assigned to investigate the particularly brutal Aggravated Criminal Sexual Assault, Robbery and Kidnapping of one [name of victim], recorded under Chicago Police Department Records Division Number N-005025. On January 10, 1990, one [name of suspect] M/N/ 31 years, FBI [#], C.P.D. Record Number [#], was arrested and charged with this and other offenses.

Blood and saliva samples of the offender and victim were obtained and tendered to Technician [name of technician] of the C.P.D. Criminalistics Unit. A sexual assault kit (Vitullo Kit) was also completed and submitted for the victim.

The undersigned requests that the recovered specimens and evidence be evaluated and subjected to DNA comparison testing. Although the offender has been identified and charged, we feel this comparison would greatly enhance the prosecution of [name of suspect], who was arrested after a week long crime spree.

If any additional information is needed, kindly contact Detective [name], star [#], Area Three Violent Crimes Unit, 3900 South California, Chicago, Illinois 60632, Telephone #(312)-744-8280, or the office of the undersigned.

Sincerely,
[name]

Combining Different Sources of Error

A declared match between a suspect and a forensic sample may not be real for several reasons. First, the RMP (random match probability) indicates how possible it is that the match is coincidence. But the RMP calculation assumes that the suspect and sample do indeed match. The match could be erroneous because of any number of human and technical errors in the process of gathering, labeling, and testing the samples. There are thus several reasons that the sample may not have come from the suspect despite the declared match.

A lot of ink and words have been exchanged over the best way to calculate the RMP in DNA typing. For the most informative DNA typing method (STR), the RMPs are typically 1 in a million (much larger with mitochondrial DNA profiles and Y-STR profiles). With larger and larger reference databases, the uncertainty in those calculations has gone down. But the impressively low RMPs are not the whole story. Something realized over a decade ago by J. Koehler (then at U. Texas) is that labs make mistakes, and the lab error rate (LER) needs to be factored into the probability of an erroneous or spurious match.

From the forensic perspective, we want to know the chance that the suspect is not the source of the sample. For the match to be ‘real,’ the lab and forensic team must not have erred in processing AND the match must not be due to chance. In the simplest case (LER independent of RMP), the match is ‘real’ with probability.

$$(1.0 - LER) * (1-RMP) = 1 - LER - RMP + RMP*LER. \text{ (match is real)}$$

If both LER and RMP are small (e.g., less than 0.1), the probability that the match is not real is very nearly

$$LER + RMP \text{ (match is not real)}$$

In other words, you add the different possible causes of a spurious match to get the overall rate that the match is not real.

The implications are profound when you consider the history of argument over RMP calculations. Most labs do not divulge their error rates (nor subject themselves to blind proficiency tests), but Koehler’s estimate of error rates was around 2%. Some labs are certain to be better than others, but even if the error rate is 0.2%, this value is large enough to render the RMP meaningless in comparison, at least when the RMP is 1 in 10,000 or less. So the emphasis should be on lab error rates (and reducing them) rather than on vanishingly small RMPs.