

---

## Chapter 4

# Fingerprints

### 1. Introduction

On January 7, 2002, in the case *U.S. v. Llera Plaza*, Louis H. Pollack, a federal judge in the United States District Court in Philadelphia, barred any expert testimony on fingerprinting that asserted that a particular print gathered at the scene of a crime is or is not the print of a particular person. As might be imagined, this decision was met with much interest, since it seemed to call into question whether fingerprinting can be used to help prove the guilt or innocence of an accused person.

In this chapter, we will consider the ways in which fingerprints have been used by society and show how the current quandary was reached. We will also consider what probability and statistics have to say about certain questions concerning fingerprints.

### 2. History of Fingerprinting

It seems that the first use of fingerprints in human society was to give evidence of authenticity to certain documents in seventh-century China, although it is possible that they were used even earlier than

this. Fingerprints were used in a similar way in Japan, Tibet, and India. In Simon Cole's excellent book on the history of fingerprinting, the Persian historian Rashid-eddin is quoted as having declared in 1303 that "Experience shows that no two individuals have fingers exactly alike."<sup>1</sup> This statement is one with which the reader is no doubt familiar. A little thought will show that unless all the fingerprints in the world are observed, it is impossible to verify this statement. Thus, one might turn to a probability model to help understand how likely it is that this statement is true. We will consider such models below.

In the Western world, fingerprints were not discussed in any written work until 1685, when an illustration of the papillary ridges of a thumb was placed in an anatomy book written by the Dutch scientist Govard Bidloo. A century later, the statement that fingerprints are unique appeared in a book by the German anatomist J. C. A. Mayer.

In 1857, a group of Indian conscripts rebelled against the British. After this rebellion had been put down, the British government decided that it needed stricter law enforcement in its colonies. William Herschel, the grandson of the discoverer of the planet Uranus, was the chief administrator of a district in Bengal. Herschel noted that the unrest in his district had given rise to a great amount of perjury and fraud. For example, it was believed that many people were impersonating deceased officers to collect their pensions. Such impersonation was hard to prove, since there was no method that could be used to decide whether a person was who he or she claimed to be.

In 1858, Herschel asked a road contractor for a handprint, to deter the contractor from trying to contest, at a later date, the authenticity of a certain contract. A few years subsequent to this, Herschel began using fingerprints. It is interesting to note that in India, as in China, the first use of fingerprints was in civil, not criminal, identification.

At about the same time, the British were increasingly concerned about crime in India. One of the main problems was to determine

---

<sup>1</sup>Cole, Simon A., *Suspect Identities: A History of Fingerprinting and Criminal Identification*, Harvard University Press, Cambridge, MA, 2001, pp. 60-61.

whether a person arrested and tried for a crime was a habitual offender. Of course, to determine this required that some method be used to identify people who had been convicted of crimes. Presumably, a list would be created by the authorities, and if a person was arrested, this list would be consulted to determine whether the person in question had prior convictions. In order for such a method to be useful, it would have to possess two properties. First, there would have to be a way to store, in written form, enough information about a person so as to uniquely identify that person. Second, the list containing this information would have to be in a form that would allow quick and accurate searches.

Although, in hindsight, it might seem obvious that one should use fingerprints to help with the formation of such a list, this method was not the first to be used. Instead, a system based on anthropometry was developed. Anthropometry is the study and measurement of the size and proportions of the human body. It was naturally thought that once adulthood is reached, the lengths of bones do not change. In the 1880s Alphonse Bertillon, a French police official, developed a system in which eleven different measurements were taken and recorded. In addition to these measurements, a detailed physical description, including information on such things as eyes, ears, hair color, general demeanor, and many other attributes, was recorded. Finally, descriptions of any “peculiar marks” were recorded. This system was called Bertillonage and was widely used in Europe, India, and the United States, as well as other locations, for several decades.

One of the main problems encountered in the use of Bertillonage was inconsistency in measurement. The “operators,” as the measurers were called, were well trained, and many measurements of each person were taken. Nonetheless, if a criminal suspect was measured in custody, and the suspect’s measurements were already in the list, the two sets of measurements might vary enough so that no match would be made.

Another problem was the amount of time required to search the list of known offenders, in order to determine whether a person in

custody had been arrested before. In some places in India, the lists grew to contain many thousands of records. Although these records were certainly stored in a logical way, the variations in measurements made it necessary to look at many records that were “near” the place that the searched-for record should be.

The chief problem at that time with the use of fingerprints for identification was that no good classification system had been developed. In this regard, fingerprints were not thought to be as useful as Bertillonage, since the latter method did involve numerical records that could be sorted. In the 1880s, Henry Faulds, a British physician who was serving in a Tokyo hospital at the time, devised a method for classifying fingerprints. This method consisted of identifying each major type of print (like those shown in Figure 1) with a certain written syllable, followed by other syllables representing different features in the print. Once a set of syllables for a given print was determined, the set was added to an alphabetical list of stored sets of syllables representing other prints.

Faulds wrote to Charles Darwin about his ideas, and Darwin forwarded them to his cousin, Francis Galton. Galton was one of the giants among British scientists in the late 19th century. His interests included meteorology, statistics, psychology, genetics, and geography. Early in his adulthood, he spent two years exploring southwest Africa. He is known as the first modern-day proponent of eugenics; in fact, this word is due to Galton.

Galton became interested in fingerprints for several reasons. He was interested in the heritability of certain traits, and one such trait that could easily be tested were fingerprint patterns. He was concerned with ethnology, and sought to compare the various races. One question that he considered in this vein was whether the proportions of the various types of fingerprints differed among the races. He also tried to determine whether any other traits were related to fingerprints. Finally, he understood the value that such a system would have in helping the police and the courts identify recidivists.

To carry out such research, it was necessary for Galton to have access to many fingerprints. By the early 1890s, he had amassed a collection of thousands of prints. This collection contained prints from people belonging to many different ethnic groups. He also collected fingerprints from certain types of people, such as criminals. He was able to show that fingerprints are partially controlled by heredity. For example, it was found that a peculiarity in a pattern in a fingerprint of a parent might pass to the same finger of a child, or, with less probability, to another finger of that child. Nonetheless, it must be stated that his work in this area did not lead to any discoveries of great import.

One of Galton's most fundamental contributions to the study of fingerprints consisted of his publishing of material, much of which was due to William Herschel, that fully established the fact that fingerprint patterns persist over the lifetime of an individual. Of at least equal importance was his development of a method to classify fingerprints. An important attribute of his method was that it allowed fingerprint records to be quickly searched to determine if a given fingerprint were present.

Very shortly thereafter, a committee consisting of various high officials in British law enforcement was formed to compare Bertillonage and the Galton fingerprint method. The goal was to decide which method to adopt. Although Bertillonage was in use in continental Europe, India, and elsewhere, it had not yet been used in Britain. The committee also considered whether it might be still better to use both methods at once.

In their deliberations, the committee noted that the Galton fingerprint method is a much easier process than the one that is used by Bertillonage operators. In addition, a fingerprint, if it is properly taken (i.e. if the resulting impression is legible), is a true and accurate rendition of the patterns on the finger. Both of these statements lead to the conclusion that Galton's method is more accurate than Bertillonage.

Given these remarks, it might seem strange that the committee did not recommend that fingerprints be the method of choice. However, there was still some concern about the accuracy of the classification method used by Galton. It was recommended that identification be made by fingerprints, with indexing by Bertillonage. The committee did foresee that the problems with fingerprint indexing could be overcome, and that in this case, the fingerprint method might be the sole system in use.

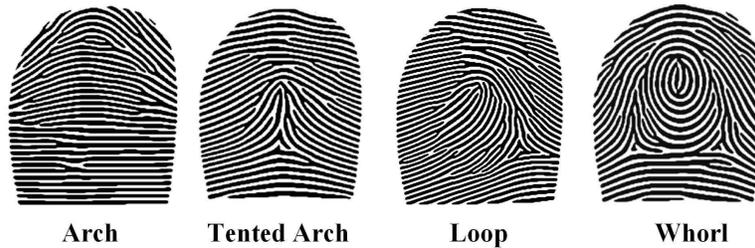
Galton continued to work on his method of classification, and in 1895, he published a system that greatly improved his previous attempts. Edward Henry, a magistrate of a district in India, worked on and modified Galton's classification method between 1898 and 1900. This modification was adopted by Scotland Yard. Regarding credit for the method, a letter from Sir George Darwin to the *London Times* had this to say: "Sir Edward Henry undoubtedly deserves great credit in recognising the merits of the system and in organising its use in a practical manner in India, the Cape and England, but it would seem that the yet greater credit is due to Mr. Francis Galton."<sup>2</sup>

In 1902, Galton published a letter entitled "Finger-Print Evidence" in the journal *Nature*, in which he discusses a pair of enlarged photographs, sent to him by Scotland Yard, of fingerprints. The first photograph came from the scene of a burglary, and the second came from the fingerprint files at Scotland Yard. Galton discusses how the use of his system allows the prosecution to explain the similarities in the two prints. The question of accuracy in matching prints obtained from a crime scene with those in a database is one that is still being considered today. Before turning to this question, we will describe Galton's method.

Galton begins by noting that in the center of most fingerprints there is a "core," which consists of patterns that he calls loops and whorls. (See Figure 1.) If no such core exists, the pattern is said to be an arch. Next, he defines a delta as the region where the parallel

---

<sup>2</sup>George Darwin, quoted in Karl Pearson, "Life, Letters, and Labors of Francis Galton," Cambridge University Press.



**Figure 1.** Four types of fingerprints. (A modified version of this figure appeared in E. Keogh, “An Overview of the Science of Fingerprints”, *Anil Aggrawal’s Internet Journal of Forensic Medicine and Toxicology*, 2001, vol. 2, no. 1 (January-June 2001).)

ridges begin to diverge to form the core. Loops have one delta, and whorls have two. These deltas serve as axes of reference for the rest of the classification. By tracing the ridges as they leave the delta(s) and cross the core, and keeping track of certain aspects such as the direction in which the loops open up, one can partition fingerprints into ten classes. We will not describe these ten classes in detail here, as the specifics are not important in what follows. Since each finger would be in one of the ten classes, there are  $10^{10}$  possible sets of ten classes. Even though the ten classes do not occur with equal frequency among all recorded fingerprints, this first level of classification already serves to distinguish between most pairs of people.

Of the ten classes, only two correspond to loops, as opposed to arches and whorls. However, about half of all observed fingerprints are loops, which suggests that the scheme is not yet precise enough. Galton was aware of this and added two other types of information to the process. The first involved using the axes of reference arising from the deltas to count ridges in certain directions. The second involved the counting and classification of what he termed “minutiae.” This term refers to places in the print where a ridge bifurcates or ends. The idea of minutiae is still in use today, although the minutiae are now sometimes referred to as “Galton points” or “points.”

There are many different types of points, and the places that they occur in a given fingerprint seems to be somewhat random. In addition, a typical fingerprint has many such points. These observations imply that if one can accurately write down where the points occur and which types of points occur, then one has a very powerful way to distinguish two fingerprints. The method is even more powerful when it is used to compare sets of ten fingerprints from two people.

### 3. Models of Fingerprints

We shall investigate some probabilistic models for fingerprints that incorporate the idea of points. The two most basic questions that one might use such models to help answer are as follows. First, in a given model, what is the probability that no two fingerprints, among all people who are now alive, are exactly alike? Second, suppose that we have a partial fingerprint, such as one that might have been recovered from a crime scene. Such partial prints are called latent prints. What is the probability that this latent print exactly matches more than one fingerprint, among all fingerprints in the world? The reason that we are interested in whether the latent print matches more than one fingerprint is that it clearly matches one print, namely the one belonging to the person who left the latent print. It is typically the case that the latent print, if it is to be of any use, will identify a suspect, i.e. someone who has a fingerprint that matches the latent print. It is obviously of great interest in a court of law as to how likely it is that someone other than the suspect has a fingerprint that matches the latent print. We will see that this second question is of central importance in the discussions going on today about the accuracy of fingerprinting as a crimefighting tool.

Galton seems to have been the first person to consider a probabilistic model that might shed some light on the answer to the first question. He began by imagining a fingerprint as a random set of ridges, with roughly 24 ridge intervals across the finger and 36 ridge intervals along the finger. Next, he imagined covering up an  $n$  by  $n$  ridge interval square on a fingerprint and attempting to recreate the

ridge pattern in the area that was covered. Galton maintained that if  $n$  were small, say at most 4, then most of the time, the pattern could be recreated by using the information in the rest of the fingerprint. However, if  $n$  were 6, he found that he was wrong more often than right when he carried out this experiment.

He then let  $n = 5$  and claimed that he would be right about one-half of the time in reconstructing the fingerprint. This led him to consider the fingerprint as consisting of a set of non-overlapping  $n$  by  $n$  squares, which he considered to be independent random variables. In Pearson's account, Galton used  $n = 6$ , although his argument is more understandable had he used  $n = 5$ . Galton claimed that any of the reconstructions, both the correct and incorrect ones, might have occurred in nature, so each random variable has two possible values, given the way that the ridges leave and enter the square, and given how many ridges leave and enter. Pearson says that Galton "proceeds to give a rough approximation to two other chances, which he considers to be involved: the first concerns guessing correctly the general course of the ridges adjacent to each square and the second of guessing rightly the number of ridges that enter and issue from the square."<sup>3</sup> Finally, Galton multiplies all of these probabilities together, under the assumption of independence, and arrives at the number 1 out of 64 billion. At the time, there were about 16 billion fingerprints in the world. (Galton claims that the odds are roughly 39 to 1 against any particular fingerprint occurring anywhere in the world. It seems to us that the odds should be 3 to 1 against.)

We will soon see other models of fingerprints that arrive at much different answers. However, it should be remembered that we are trying to estimate the probability that no two fingerprints, among all people who are now alive, are exactly alike. Suppose, as Galton did, that there are 16 billion fingerprints among the people of the world, and there are 64 billion possible fingerprints. Does the reader think that these assumptions make it very likely or very unlikely that there are two fingerprints that are the same? To answer this

---

<sup>3</sup>Pearson, *ibid.*, p. 182.

question, we can proceed as follows. Consider an urn with 64 billion labeled balls in it. We choose, one at a time, 16 billion balls from the urn, replacing the balls after each choice. We are asking for the probability that we never choose the same ball more than once. This is the celebrated birthday problem, in a world where there are 64 billion days in a year and 16 billion people. The birthday problem asks what is the probability that at least two people share a birthday. The complementary probability, i.e. the probability that no two people share a birthday, is

$$\left(1 - \frac{0}{n}\right) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right),$$

where  $n = 64$  billion and  $k = 16$  billion. This can be seen by considering the people one at a time. If 6 people, say, have already been considered and if they all have different birthdays, then the probability that the seventh person has a birthday that is different than all of the first 6 people equals

$$\left(1 - \frac{6}{n}\right).$$

(One way to obtain a rough upper bound on this product is to use the inequality

$$1 - x < e^{-x},$$

which is valid for  $x > 0$ .) For the values given by Galton, the product is less than

$$\frac{1}{10^{10^9}}.$$

This means that in Galton's model, with his estimates, it is extremely likely that there are two fingerprints that are the same.

In fact, to our knowledge, no two fingerprints from different people have ever been found that are identical. Of course, it is not the case that all fingerprints on Earth have been recorded or compared, but the FBI has a database with more than 10 million fingerprints in it, and we presume that no two fingerprints in it are exactly the same. (It must be said that it is not clear to us that all pairs of fingerprints

in this database have actually been compared. In addition, one wonders whether the FBI, if it found a pair of identical fingerprints, would announce this to the world.) In any case, if we use Galton's estimate for the number of possible fingerprints and let  $k = 10$  million, the probability that no two are alike is still very small; it is less than

$$\frac{1}{10^{339}}.$$

We can turn the above question around and ask the following question. Suppose that there are 60 billion fingerprints in the world, and suppose that we imagine they are chosen from a set of  $n$  possible fingerprints. How large would  $n$  have to be in order that the probability that all of the chosen fingerprints are different exceeds .999? An approximate answer to this question is that it would suffice for  $n$  to be at least  $10^{25}$ . Although this is quite a bit larger than Galton's estimate, there have been other, more sophisticated models of fingerprints, some of which we will now describe, have come up with estimates for  $n$  that are much larger than  $10^{25}$ . Thus, if these models are at all accurate, it is extremely unlikely that there exist two fingerprints in the world that are exactly alike.

In 1933, T. Roxburgh described a model for fingerprint classification that is much more intricate than Galton's model. This model, and many others, are described and compared in an article in the *Journal of Forensic Sciences*, written by D. A. Stoney and J. I. Thornton.<sup>4</sup> In Roxburgh's model, a vertical ray is drawn upwards from the center of the fingerprint. This idea must be accurately defined, but for our purposes, we can take it to mean the center of the loop or whorl or the top of the arch. This ray is defined to be 0 degrees. Another ray, with endpoint at the center, is revolved clockwise from the first ray. As this ray passes over minutiae, the types of the minutiae are recorded, along with the ridge numbers on which the minutiae lie. If a fingerprint has  $R$  concentric ridges,  $n$  minutiae, and there are  $T$

---

<sup>4</sup>Stoney, D. A. and J. I. Thornton, "A Critical Analysis of Quantitative Fingerprint Individuality Models", *Journal of Forensic Sciences*, vol. 31, no. 4 (1986), pp. 1187-1216.

minutia types, then the number of possible patterns equals

$$(RT)^n ,$$

since as the second ray revolves clockwise, the next minutia encountered could be on any of the  $R$  ridges and be of any of the  $T$  minutia types. Roxburgh also introduces a factor of  $P$  that corresponds to the number of different overall patterns and core types that might be encountered. Thus, he estimates the number of possible fingerprints to be

$$P(RT)^n .$$

He takes  $P = 1000$ ,  $R = 10$ ,  $T = 4$ , and  $n = 35$ ; this last value is Galton's estimate for the typical number of minutia in a fingerprint. If we calculate the above expression with these values, we obtain the number

$$1.18 \times 10^{59} .$$

Roxburgh modified the above expression for the number of possible fingerprints to attempt to account for ambiguities between various types of minutiae. For example, it is possible that a fork in a ridge might be seen as a ridge ending, depending upon whether the ridges in question meet each other or not. Roxburgh suggested using a number  $Q$  which would vary depending upon the quality of the fingerprint under examination. The value of  $Q$  ranges from 1.5 to 3, with the smaller value corresponding to a higher quality fingerprint. For each minutia, Roxburgh replaced the factor  $RT$  by the factor  $RT/Q$ . This leads to the expression

$$P((RT)/Q)^n$$

as an estimate for the number of discernable types of fingerprints, assuming their quality corresponds to a particular value of  $Q$ . Note that even if  $Q = 3$ , so that  $RT/Q = 1.33R$ , the number of discernable types of fingerprints in this model is

$$2.16 \times 10^{42} .$$

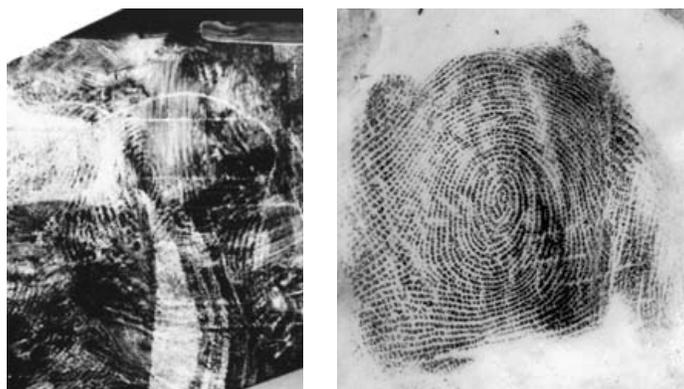


Figure 2. Examples of latent and rolled prints

Stoney and Thornton note that although this is a very interesting, sophisticated model, it has been “totally ignored by the forensic science community.”<sup>5</sup>

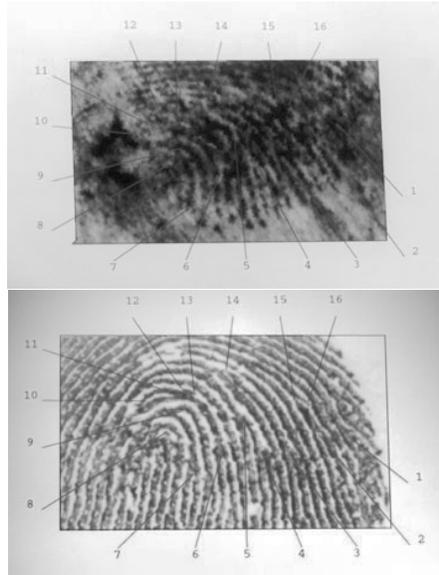
#### 4. Latent Fingerprints

According to a government expert who testified at a recent trial, the average size of a latent fingerprint fragment is about one-fifth the size of a full fingerprint. Since a typical fingerprint contains between 75 and 175 minutiae<sup>6</sup>, this means that a typical latent print has between 15 and 35 minutiae, assuming that minutiae are roughly evenly distributed across the print. In addition, the latent print recovered from a crime scene is frequently of poor quality, which tends to increase the likelihood of mistaking the types of minutiae being observed.

In a criminal case, the latent print is compared with a high quality print taken from the hand of the accused or from a database of fingerprints. Figure 2 shows a latent print and the corresponding rolled print to which the latent print was matched. Figure 3 shows another

<sup>5</sup>ibid., p. 1192.

<sup>6</sup>“An Analysis of Standards in Fingerprint Identification 1,” *Federal Bureau of Investigation Department of Justice Law Enforcement Bulletin*, vol. 1 (June 1972).



**Figure 3.** Minutiae matches

pair of prints, one latent and one rolled, from the same case. The figure also shows the claimed matching minutiae in the two prints.

The person making the comparison states that there is a match if he or she believes that there are a sufficient number of common minutiae, both in type and location, in the two prints. There have been many criminal cases in which an identification was made with fewer than fifteen matching minutiae<sup>7</sup>. There is no general agreement among various law enforcement agencies or among various countries on the number of matching minutiae that must exist in order for a match to be declared. In fact, according to Robert Epstein<sup>8</sup>, “many examiners ... including those at the FBI, currently believe that there

<sup>7</sup>see footnote 25 in Epstein, Robert, “Fingerprints Meet Daubert: The Myth of Fingerprint ‘Science’ is Revealed,” *Southern California Law Review*, vol. 75 (2002), pp. 605-658.

<sup>8</sup>*ibid.*, p. 610.

should be no minimum standard whatsoever and that the determination of whether there is a sufficient basis for an identification should be left to the subjective judgment of the individual examiner.” It is quite understandable that a law enforcement agency might object to constraints on its ability to claim matches between fingerprints, as this could only serve to decrease the number of matches obtained.

In some countries, fingerprint matches can be declared with as few as eight minutiae matches. However, there are examples of fingerprints from different people that have seven matching minutiae. In a California bank robbery trial, *U.S. v. Parks*, in 1991, the prosecution introduced evidence that showed that the suspect’s fingerprint and the latent print had ten points. The trial judge, Spencer Letts, asked the prosecution expert what the minimum standard was for points in order to declare a match. The expert announced that the minimum was eight. Judge Letts had seen fingerprint evidence entered in other trials. He said “If you only have ten points, you’re comfortable with eight; if you have twelve, you’re comfortable with ten; if you have fifty, you’re comfortable with twenty.”<sup>9</sup> Later in the same trial, the following exchange occurred between Judge Letts and another prosecution fingerprint expert:

The Witness: The thing you have there is that each department has their own goals or their own rules as far as the number of points being a make [an identification]. ...that number really just varies from department to department.

The Court: I don’t think I’m ever going to use fingerprint testimony again; that simply won’t do...

The Witness: That just may be one of the problems of the field, but I think if there was [a] survey taken, you would probably get a different number from every department that has a fingerprint

---

<sup>9</sup>Cole, *op. cit.*, p. 272.

section as to their lowest number of points for a comparison and make.

The Court: That's the most incredible thing I've ever heard of.<sup>10</sup>

According to Simon Cole, no scientific study has been carried out to estimate the probability of two different prints sharing a given number of minutiae. David Stoney and John Thornton claim that none of the fingerprint models proposed during the past century “even approaches theoretical accuracy ..., and none has been subjected to empirical validations.”<sup>11</sup> In fact, latent print examiners are prohibited by their primary professional association, the International Association for Identification (IAI), from offering opinions of identification using probabilistic terminology. A resolution, passed by the IAI at one of its meetings, states that “any member, officer, or certified latent print examiner who provides oral or written reports, or gives testimony of possible, probable, or likely friction ridge identification shall be deemed to be engaged in [unbecoming] conduct... and charges may be brought.”<sup>12</sup>

In 1993, the Supreme Court rendered a decision in the case *Daubert v. Merrell Dow Pharmaceuticals, Inc.*<sup>13</sup> The Court described certain factors that courts needed to consider when deciding whether to admit expert testimony. In this decision, the Court concentrated on scientific expert testimony; it considered the issue of expert testimony of a non-scientific nature in the case *Kumho Tire Co. v. Carmichael*<sup>14</sup>, a few years later. In the first decision, the Court interpreted the Federal Rule of Evidence 702, which defines the term “expert witness” and states when such witnesses are allowed, as requiring trial judges to determine whether the opinion of an expert witness lacks sufficient reliability, and if so, to exclude this testimony. The Daubert decision listed five factors that could be considered when determining

---

<sup>10</sup> *ibid.*, pp. 272-273.

<sup>11</sup> Stoney and Thornton, *op. cit.*, p. 1187.

<sup>12</sup> Epstein, *op. cit.*, p. 611, footnote 32.

<sup>13</sup> 509 U.S. 579 (1993).

<sup>14</sup> 526 U.S. 137 (1999).

whether scientific expert testimony should be retained or excluded. These factors are as follows:

1. “A preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.”<sup>15</sup>
2. “The court ordinarily should consider the known or potential rate of error... .”<sup>16</sup>
3. The court should consider “the existence and maintenance of standards controlling the technique’s operation... .”<sup>17</sup>
4. “‘General acceptance’ can ... have a bearing on the inquiry. A reliability assessment does not require, although it does permit, explicit identification of a relevant scientific community and an express determination of a particular degree of acceptance within that community.”<sup>18</sup>
5. “A pertinent consideration is whether the theory or technique has been subjected to peer review and publication... .”<sup>19</sup>

In the *Kumho* case, the Court held that a trial court’s obligation to decide whether to admit expert testimony applies to all experts, not just scientific experts. The Court also held that the factors listed above may be used by a court in assessing nonscientific expert testimony.

In the case (*U.S. v. Llera Plaza*) mentioned at the beginning of the chapter, the presiding judge, Louis Pollack, applied the Daubert criteria to the fingerprint identification process, as he was instructed

---

<sup>15</sup>509 U.S. 579 (1993), note 593.

<sup>16</sup>*ibid.*, note 594.

<sup>17</sup>*ibid.*

<sup>18</sup>*ibid.*, quoted from *U.S. v. Downing*, 753 F.2d 1224, 1238 (3d Cir. 1985).

<sup>19</sup>*ibid.*, note 593.

to do by the *Kumho* case. In particular, he discussed the problem with the current process employed by the FBI (and other law enforcement agencies), which is called the ACE-V system. This name is an acronym that stands for analysis, comparison, evaluation, and verification. Judge Pollack ruled that the third part of this process, in which a fingerprint expert states his or her opinion that the latent print and the comparison print (either a rolled print from a suspect or a print from a database) either match or do not match, did not measure up to several of the Daubert criteria.

With regard to the first criterion, the government (the plaintiff in the case) argued that the method of fingerprint matching had been tested empirically over a period of 100 years. It also argued that in any particular case, the method can be tested through the testimony of a fingerprint expert other than the one whose testimony is being heard. The judge rejected this argument, saying that neither of these actions could be considered as scientific tests of the method. He further noted that in the second case, the strength of the second examiner's "test" of a claimed match is diluted by the fact that in many cases, the second examiner has been advised of the first examiner's claims in advance.

On the point of testing, it is interesting to note that in 2000, the National Institute of Justice (NIJ), which is an arm of the Department of Justice, had solicited proposals for research projects to study the reliability of fingerprinting. This solicitation was mentioned by the judge in his ruling and was also taken as evidence by the defense that the government did not know whether fingerprinting was reliable.

The second Daubert criterion concerns the "known or potential rate of error" of the method. In their arguments before the court, the government contended that there were two types of error—methodology error and practitioner error. One of the government's witnesses, when asked to explain methodology error, stated that "an error rate is a wispy thing like smoke, it changes over time..."<sup>20</sup> The

---

<sup>20</sup> *U.S. v. Llera Plaza*, January 7, 2002, at 47.

judge said that “the full import of [this] testimony is not easy to grasp.” He summarizes this testimony as saying that if a method, together with its limitations, has been defined, then there is no methodology error. All of the error is practitioner error. The other government witness, Stephen Meagher, a supervisory fingerprint specialist with the FBI, also testified that if the scientific method is followed, then the methodology error rate will be zero, i.e. all of the error is practitioner error. We will have more to say about practitioner error below.

Judge Pollack also found problems concerning the third Daubert criterion, which deals with standards controlling a technique’s operation. There are three types of standards discussed in the judge’s ruling. The first is whether there is a minimum number of Galton points that must be matched before an overall match is declared. In the ACE-V process, no minimum number is prescribed, and in fact, in some jurisdictions, there is no minimum. The second type of standard concerns the evaluation of whether a match exists. The government and defense witnesses agreed that this decision is subjective. The judge concluded that “it is difficult to see how fingerprint identification—the matching of a latent print to a known fingerprint—is controlled by any clearly describable set of standards to which most examiners subscribe.”<sup>21</sup> Finally, there is the issue of the qualifications of examiners. There are no mandatory qualification standards that must be attained in order for someone to become a fingerprint examiner; nor are there any uniform certification processes.

Regarding the fourth Daubert criterion, the judge had this to say:

General acceptance by the fingerprint examiner community does not ... meet the standard... . First, there is the difficulty that fingerprint examiners, while respected professionals, do not constitute a ‘scientific community’ in the Daubert sense... . Second, the Court cautioned in *Kumho*

---

<sup>21</sup>ibid. at 58.

*Tire* that general acceptance does not help show that an expert's testimony is reliable where the discipline itself lacks reliability. The failure of fingerprint identifications fully to satisfy the first three Daubert factors militates against heavy reliance on the general acceptance factor. Thus, while fingerprint examinations conducted under the general ACE-V rubric are generally accepted as reliable by fingerprint examiners, this by itself cannot sustain the government's burden in making the case for the admissibility of fingerprint testimony under Federal Rule of Evidence 702.<sup>22</sup>

The conclusion of the judge's ruling was as follows:

For the foregoing reasons:

- A. This court will take judicial notice of the uniqueness and permanence of fingerprints.
- B. The parties will be able to present expert fingerprint testimony (1) describing how any latent and rolled prints at issue in this case were obtained, (2) identifying, and placing before the jury, such fingerprints and any necessary magnifications, and (3) pointing out any observed similarities and differences between a particular latent print and a particular rolled print alleged by the government to be attributable to the same persons. But the parties will not be permitted to present testimony expressing an opinion of an expert witness that a particular latent print matches, or does not match, the rolled print of a particular person and hence is, or is not, the fingerprint of that person.<sup>23</sup>

---

<sup>22</sup>*ibid.* at 61.

<sup>23</sup>*ibid.* at 69.

The government asked for a reconsideration of this ruling. Not surprisingly, it felt that its effectiveness in both the trial at hand and in future trials would be seriously compromised if witnesses were not allowed to express an opinion on whether or not a latent print matches a rolled print. The government asked to be allowed to submit evidence that would show the accuracy of FBI fingerprint examiners. The defendants argued that the judge should decline to reconsider his ruling, and Judge Pollack stated that their argument was solid: “Neither of the circumstances conventionally justifying reconsideration—new, or hitherto unavailable facts or new controlling law—was present here.”<sup>24</sup> Nonetheless, the judge decided to grant a reconsideration hearing, arguing that the record on which he made his previous ruling was testimony presented two years earlier in another courtroom. “It seemed prudent to hear such live witnesses as the government wished to present, together with any rebuttal witnesses the defense would elect to present.”<sup>25</sup>

At this point in our narrative, it makes sense to consider the various attempts to measure error rates in the field of fingerprint analysis. Lynn and Ralph Haber, who are consultants at a private company in California and also adjuncts at the University of California at Santa Cruz, have obtained and analyzed relevant data from many sources.<sup>26</sup> These data include both results on crime laboratories and individual practitioners. We will summarize some of their findings here.

The American Society of Crime Laboratory Directors (ASCLD) is an organization that provides leadership in the management of forensic science. It is in their interest to evaluate and improve the quality of operations of crime laboratories. In 1977, the ASCLD began developing an accreditation program for crime laboratories. By 1999, 182 labs had been accredited. One requirement for a lab to be accredited

---

<sup>24</sup> *U.S. v. Llera Plaza*, March 13, 2002, at 11.

<sup>25</sup> *ibid.*

<sup>26</sup> Haber, Lynn, and Ralph Norman Haber, “Error Rates for Human Latent Fingerprint Examiners,” in *Advances in Automatic Fingerprint Recognition*, Nalini K. Ratha, ed., New York, Springer-Verlag, 2003.

is that the examiners working in the lab must pass an externally administered proficiency test. We note that since it is the lab, and not the individual examiners, that is being tested, these proficiency tests are taken by all of the examiners as a group in a given lab.

Beginning in 1983, the ASCLD began administering such a test in the area of fingerprint identification. The test, which is given each year to all labs requesting accreditation, consists of pictures of 12 or more latent prints and a set of ten-print (rolled print) cards. The set of latent prints contains a range of quality and is supposed to be representative of what is actually seen in practice. For each latent print, the lab was asked to decide whether it is “scorable,” i.e. whether it is of sufficient quality to attempt to match it with a rolled print. If it is judged to be scorable, then the lab is asked to decide whether or not it matches one of the prints on the ten-print cards. There are “correct” answers for each latent print on the test, i.e. the ASCLD has decided, in each case, whether or not a latent print is scorable, and if so, whether or not it matches any of the rolled prints.

The Habers report on results from 1983 to 1991. During this time, the number of labs that took the exam increased from 24 to 88; many labs took the tests more than once. A new test was constructed each year. Assuming that in many cases the labs have more than one fingerprint expert, this means that hundreds of these experts participated in the test at least once during this period.

Each lab returned one answer for each question. There are four types of errors that can be made on each question of each test. A scorable print can be ruled unscorable or vice versa. If a print is correctly judged to be scorable, it can be erroneously matched to a rolled print, or it can fail to be matched at all, even though a match exists. Of these four types of errors, the second and third are more serious than the others, assuming that we take the point of view that erroneous evidence against an innocent person should be strenuously guarded against.

The percentage of answers with errors of each of the four types were 8%, 2%, 2%, and 8%, respectively. What should we make of these error rates? We see that the more serious types of errors had lower rates, but we must remember that these answers are consensus answers of the experts in a given lab. For purposes of illustration, suppose that there are two experts in a given lab and they agree on an answer that turns out to be incorrect. Presumably they consulted each other on their answers, so we cannot multiply their individual error rates to obtain their group error rate, since their answers were not independent events. However, we can certainly suppose that if a lab error rate is 2%, say, then the individual error rate of at least one of the experts at the lab who took the test is at least 2%.

In 1994, the ASCLD asked the IAI for assistance in creating and reviewing future tests. The IAI asked a company called Collaborative Testing Services (CTS) to design and administer these tests. The format of these tests is similar to the earlier ones, but all of the latent prints are scorable, so there are only two possible types of errors for each question. In addition, individual fingerprint examiners who wish to do so may take the exam by themselves. The Habers report on the error rates for the examinations given from 1995 through 2001. Of the 1685 tests that were graded by CTS, 95 of them, or more than 5%, had at least one erroneous identification, and 502 of the tests, or more than 29%, had at least one missed identification.

Since 1995, the FBI has administered its own examinations to all of its fingerprint examiners. These examinations are similar in nature to the ones described above, but there are a few differences worthy of note. These differences were described in Judge Pollack's reconsideration ruling, in the testimony of Allan Bayle, a fingerprint examiner for 25 years at Scotland Yard.<sup>27</sup>

Mr. Bayle had reviewed copies of the internal FBI proficiency tests before taking the stand. He found the latent prints utilized in those tests to be, on

---

<sup>27</sup> *U.S. v. Llera Plaza*, March 13, 2002, at 24.

the whole, markedly unrepresentative of the latent prints that would be lifted at a crime scene. In general, Mr. Bayle found the test latent prints to be far clearer than the prints an examiner would routinely deal with. The prints were too clear—they were, according to Mr. Bayle, lacking in the “background noise” and “distortion” one would expect in latent prints that were not identifiable; according to Mr. Bayle, at a typical crime scene only about ten per cent of the lifted latent prints will turn out to be matched. In Mr. Bayle’s view the paucity of non-identifiable prints: “makes the test too easy. It’s not testing their ability. It doesn’t test their expertise. I mean I’ve set these tests to trainees and advanced technicians. And if I gave my experts these tests, they’d fall about laughing.”

Approximately 60 FBI fingerprint examiners took the FBI test each year in the period from 1995 to 2001. On these tests, virtually all of the latent prints had matches among the rolled prints. Since many of the examiners took the tests most or all of these years, it is reasonable to suppose that they knew this fact, and hence would hardly ever claim that a latent print had no match. The results of these tests are as follows: there were no erroneous matches, and only three cases where an examiner claimed there was no match when there was one. Thus, the error rates for the two types of error were 0% and 1%.

It seems clear that the error rates of the crime labs for the various types of error are small, but not negligible, and the FBI’s rates are suspect for the reasons given above. Given that in many criminal cases fingerprint evidence forms a crucial part of the prosecution’s case, it is reasonable to ask whether the above data, were it to be submitted to a jury, would make it difficult for the jury to find the defendant guilty “beyond a reasonable doubt,” which is the standard that must be met in such cases.

The question of what this last phrase means is a fascinating one, and the answers show how hard it is to use probabilistic language in the legal world. The U.S. Supreme Court recently weighed in on this issue, and the majority opinion is thorough in its attempt to explicate the history of the usage of this phrase. The Court agreed to review two cases involving instructions given to juries by judges. Standard instructions to juries state that “guilt beyond a reasonable doubt” means that the jurors need to be convinced “to a moral certainty” of the defendant’s guilt. In one case, “California defended the use of the moral-certainty language as a “commonsense and natural” phrase that conveys an “extraordinarily high degree of certainty.”<sup>28</sup> In the second case, a judge in Nebraska “included not only the moral-certainty language but also a definition of reasonable doubt as ‘an actual and substantial doubt.’ The jurors were instructed that ‘you may find an accused guilty upon the strong probabilities of the case, provided such probabilities are strong enough to exclude any doubt of his guilt that is reasonable.’”<sup>29</sup> The Supreme Court upheld both sets of instructions. The decision regarding the first set was unanimous, while in the second case, two justices dissented, noting that “the jury was likely to have interpreted the phrase ‘substantial doubt’ to mean that ‘a large as opposed to a merely reasonable doubt is required to acquit a defendant.’”<sup>30</sup>

The Court went on to note that the meaning of the phrase “moral certainty” has changed over time. In the mid-19th century, the phrase generally meant a high degree of certainty, whereas today, some dictionaries define the phrase to mean “based on strong likelihood or firm conviction, rather than on the actual evidence.”<sup>31</sup> Although the Court upheld both sets of instructions, the majority opinion stated that the Court did not condone the use of the phrase “moral certainty.”

---

<sup>28</sup>Linda Greenhouse, “High Court Warns About Test for Reasonable Doubt,” *New York Times*, March 22, 1994.

<sup>29</sup>*ibid.*

<sup>30</sup>*ibid.*

<sup>31</sup>*American Heritage Dictionary of the English Language*, 1992.

In a concurring opinion, Justice Ruth Bader Ginsburg noted that some Federal appellate circuit courts have instructed trial judges not to provide any definition of the phrase “beyond a reasonable doubt.” Justice Ginsburg said that it would be advisable to construct a better definition than the one used in the instructions in the cases under review. She cited one suggested in 1987 by the Federal Judicial Center, a research arm of the Federal judiciary. Making no reference to moral certainty, that definition says in part, “Proof beyond a reasonable doubt is proof that leaves you firmly convinced of the defendant’s guilt.”<sup>32</sup>

It may very well be the case that after wading through the above verbiage, the reader has no clearer an idea (and perhaps even has a less clear idea) than before of what the phrase “beyond a reasonable doubt” means. However, juries are given this phrase as part of their instructions, and in the case of fingerprint evidence, they deserve to be educated about error rates involved. We leave it to the reader to ponder whether evidence produced by a technique whose error rate seems to be at least 2% is strong enough to be beyond a reasonable doubt.

On March 13, 2002, Judge Pollack filed his second decision in the Llera Plaza case. The judge’s ruling was a partial reversal of the original one. His ruling allowed FBI fingerprint examiners to state in court whether there is a match between a latent and a rolled print, but nothing was said in the ruling about examiners not in the employ of the FBI. The judge’s mind was changed primarily because of the testimony of Mr. Bayle who, ironically, was a witness for the defense. Although, as noted above and in the judge’s decision, there are shortcomings in the FBI’s proficiency testing of its examiners, the judge was convinced by the facts that the ACE-V system used by the FBI is essentially the same as the system used in Great Britain and that Mr. Bayle believes in this system without reservation.

---

<sup>32</sup>Greenhouse, *loc. cit.*

As an interesting footnote to this case, after Judge Pollack announced his second ruling, the NIJ cancelled its original solicitation, described above, and replaced it by a “General Forensic Research and Development” solicitation. In the guidelines for this proposal under “what will not be funded,” we find the phrase “proposals to evaluate, validate, or implement existing forensic technologies.” This is a somewhat strange way to respond to the judge’s worries about whether the method has been adequately tested in a scientific manner.

## 5. The 50K Study

At the beginning of Section 3, we stated that in order to decide whether fingerprints are useful in forensics, it is of central importance to be able to estimate how likely it is that a latent print will be incorrectly matched to a rolled print. In 1999, the FBI asked the Lockheed Martin Company to carry out a study of fingerprints. In a pre-trial hearing in the case *U.S. v. Mitchell*<sup>33</sup>, Stephen Meagher, whom we have introduced earlier, explained why he commissioned the study. The primary reason for carrying out this study, he said, was to use the FBI database of over 34 million sets of 10 rolled prints to see how well the automatic fingerprint recognition computer programs distinguished between prints of different fingers. The results of the study could also be used, he reasoned, to strengthen the claim that no two fingerprints are alike. Thus, this study was not originally conceived as a test of the accuracy of matching latent and rolled prints. Nonetheless, as we shall see, this study touched on this second issue.

Together with Bruce Budlowe, a statistician who works for the FBI, Meagher came up with the following design for the experiment. The overall idea was to compare every pair of rolled prints in the database, to see if the computer algorithms could distinguish among different prints with high accuracy. It was decided that carrying this

---

<sup>33</sup> *U.S. v. Mitchell*, July 7, 1999.

out for the whole database was not reasonable (about  $5.8 \times 10^{16}$  comparisons would be required), so they instead chose 50,000 rolled fingerprints from the FBI's master file. These prints were not chosen at random; rather, they were the first 50,000 that were of the pattern "left loop" from white males. It was decided to restrict the fingerprints in this way because according to Meagher, race and gender have some effect on the size and types of fingerprints. By restricting in this way, the resulting set of fingerprints are probably more homogeneous than a set of randomly chosen fingerprints would be, thereby making it harder to distinguish between pairs from the set. If the study could show that each pair could be distinguished, then the result is more impressive than a similar result accomplished using a set of randomly chosen prints.

At this point, Meagher turned the problem of design over to the Lockheed group, where the design and implementation of the study were carried out by Donald Zeisig, an applied mathematician and software designer, and James O'Sullivan, a statistician. Much of what follows comes from testimony that Zeisig gave at the pre-trial hearing in *U.S. v. Mitchell*.

Two experiments with this data were performed. The first began by using two different software programs that each generated a measure of similarity between two fingerprints based on their minutiae patterns. A third program was used to merge these two measures. A paper by David Kaye<sup>34</sup> delved into various difficulties presented by this study. Information about this study was also provided by the fascinating transcripts of the pre-trial hearing mentioned above<sup>35</sup>.

We follow Kaye in denoting the measure of similarity between fingerprints  $f_i$  and  $f_j$  by  $x(f_i, f_j)$ . Each of the fingerprints was compared with itself, and the function  $x$  was normalized. Although this normalization is not explicitly defined in either the court testimony or the Lockheed summary of the test, we will proceed as best we

---

<sup>34</sup>Kaye, David, "Questioning a Courtroom Proof of the Uniqueness of Fingerprints", *International Statistical Review*, vol. 71, no. 3 (2003), pp. 521-533.

<sup>35</sup>Daubert Hearing Transcripts, at [www.clpex.com/Mitchell.htm](http://www.clpex.com/Mitchell.htm)

can. It seems that the values of  $x(f_i, f_j)$  were all multiplied by a constant, so that  $x(f_i, f_i) \leq 1$  for all  $i$ , and there is an  $i$  such that  $x(f_i, f_i) = 1$ . One would expect that a measure of similarity would be symmetric, i.e. that  $x(f_i, f_j) = x(f_j, f_i)$ , but this is never mentioned in the report, and in fact there is evidence that this is not true for this measure.

The value of  $x(f_i, f_j)$  is then computed for all  $2.5 \times 10^9$  ordered pairs of fingerprints. If this measure of similarity is of any value, it should be very small for all pairs of non-identical fingerprints and large (i.e. close to 1) for all pairs of identical fingerprints.

Next, for each rolled print  $f_i$ , the 500 largest values of  $x(f_i, f_j)$  are recorded. One of these values, namely when  $j = i$ , will presumably be very close to 1, but the other 499 values will probably be very close to 0. At this point, the Lockheed group calculated the mean and standard deviation of this set of 500 values (for each fixed value of  $i$ ). Presumably, the mean and the standard deviation are both positive and very close to 0 (since all but one of the values is very small and positive).

Next, Zeisig and O'Sullivan assume that the distribution, for each  $i$ , is normal, with the calculated mean and standard deviation. No reason is given for making this assumption, and we shall see that it gives rise to some amazing probabilities. Under this assumption, one can change the values of  $x(f_i, f_j)$  into values of a standard normal distribution by subtracting the mean and dividing by the standard deviation. The Lockheed group calls these normalized values  $Z$  scores. The reader can see that if this is done for a typical set of 500 values of  $x(f_i, f_j)$ , with  $i$  fixed, one should obtain 499  $Z$  scores that are fairly close to 0 and one  $Z$  score, corresponding to  $x(f_i, f_i)$ , that is quite large.

It is then pointed out that if one takes 500 independent values from the standard normal distribution, the expected value of the largest value obtained should be about 3. This value is easy to estimate by simulation; we simulated 50,000 repetitions of the maximum

of 500 independent standard normal values, and found the mean of the maximum values to be 3.04. Thus, Zeisig and O'Sullivan would be worried if any of the non-mate  $Z$  scores (i.e.  $Z$  scores corresponding to pairs  $(f_i, f_j)$  with  $i \neq j$ ) were much greater than 3. In fact, except for three cases, which will be discussed below, all of the non-mate  $Z$  scores were less than 1.83. This fact casts much doubt on whether the distribution in question is normal.

The three non-mate  $Z$  scores that were larger than 1.83 corresponded to the  $(i, j)$ -pairs (48541, 48543), (48543, 48541), and (18372, 18373). The scores in these cases were 6.98, 6.95, and 3.41. When Zeisig and O'Sullivan found these high  $Z$  values, they discovered that in all three cases, the pairs were different rolled prints of the same finger. In other words, the sample of 50,000 fingerprints were from at most 49,998 different people. It is interesting to note that the ordered pair (18373, 18372) must have had a  $Z$  score of less than 1.83, even though the pair corresponds to two prints of the same finger. We'll have more to say about this below. This shows that it is possible for two different prints of the same finger to generate a  $Z$  score which is in the same range as a score generated by two prints of different fingers.

Now things get murky. The smallest  $Z$  score of any fingerprint paired with itself was stated to be 21.7. This high value is to be expected; the reader will recall that for any fingerprint  $f_i$ , the 500 values correspond to 499 small  $Z$  scores and one very large  $Z$  score. However, the conclusion drawn from this statement is far from clear. If one calculates the probability that a standard normal random variable will take on a value greater than 21.0, one obtains a value of less than  $10^{-97}$ . The Lockheed group states its conclusion as follows<sup>36</sup>: "The probability of a non-mate rolled fingerprint being identical to any particular fingerprint is less than  $10^{-97}$ ."

David Kaye points out that the real question is not whether a computer program can detect copies of photographs of rolled prints,

---

<sup>36</sup>Kaye, op. cit. , p. 530

as is done in this study when a rolled print is compared with itself. Rather, it is whether such a program can, for each finger in the world, put all rolled prints of that finger in one category and make sure that no rolled prints from any other finger fall into that same category. Kaye notes that although there were so few repeated fingers in the study that one cannot determine the answer to this question with any great degree of certainty, one of the three pairs noted above, of different rolled prints of the same finger, produced a  $Z$  score that would occur about once in every 3000 comparisons, assuming the comparisons generate scores that are normally distributed. This means that if one were to make millions of comparisons between pairs of rolled prints of different fingers, one would find thousands of  $Z$  scores as high as the one corresponding to the pair (18372, 18373). This would put the computer programmer in a difficult situation. To satisfy Kaye, the program would have to be assigned a number  $Z^*$  with the property that if a  $Z$  score were generated that was above this value, the program would state that the prints were of the same finger, while if the generated  $Z$  score were below this value, the program would state that the two prints were of different fingers. But we can see that there can be no such  $Z^*$  value that will always be right. If  $Z^* > 3.41$  (the value corresponding to the pair (18372, 18373)) then the program would declare that the thousands of the pairs of prints of different fingers mentioned above are in fact prints of the same finger. If  $Z^* < 3.41$ , then the program would declare that the pair (18372, 18373) are prints of different fingers.

As we noted above, the pair (18373, 18372) was not flagged as having a large  $Z$  score. The reason for this is that when the three non-mate pairings mentioned above were flagged, it was not yet known that they corresponded to the same fingers. However, one does wonder whether the Lockheed group looked at the  $Z$  score of this pair, once the reversed pair was discovered to have a high  $Z$  score. In any event, the  $Z$  score of this pair is not given in the summary of the experiments. Robert Epstein, an attorney for the defense in *U.S.*

*v. Mitchell*, noticed this fact as well and asked Donald Zeisig, during cross-examination, what the  $Z$  score of this pair was. It turns out that the  $Z$  score was 1.79. This makes things still worse for the matching algorithm.

First, there were other non-mate pairs with larger  $Z$  scores. Second, one might expect that the  $Z$  score of a pair would be roughly the same in either order (although it isn't clear that this should be so). In any event, a  $Z$  score of 1.79 does not correspond to an extremely unlikely event; thus, the algorithm might fail, with some not-so-small probability, to detect an identification between two fingerprints (or else might, with some not-so-small probability, make false identifications). In fact Epstein, in his cross-examination, noted that the pair (12640, 21111) had the  $Z$  values 1.83 and 1.47 (depending upon the order), even though it was later discovered that both of this prints were of the same finger. When asked by Epstein, Zeisig agreed that there could possibly have been other pairs of different prints of the same finger (which must have had low  $Z$  values, since they were not flagged).

The second experiment that the Lockheed group performed was an attempt to find out how well their computer algorithms dealt with latent fingerprints. To that end, a set of "pseudo" latent fingerprints was made up, by taking the central 21.7% of each of the 50,000 rolled prints in the original data set. This percentage was arrived at by taking the average size of 300 latent prints from crime scenes versus the size of the corresponding rolled prints.

At this point, the experiment was carried out in essentially the same way as the first experiment. Each pseudo latent  $l_i$  was compared with all 50,000 rolled prints, and a score  $y(l_i, f_j)$  was determined. For each latent  $l_i$ , the largest 500 values of  $y(l_i, f_j)$  were used to construct  $Z$  scores. As before, the  $Z$  score corresponding to the pair  $(l_i, f_i)$  was expected to be the largest of these by far. Any non-mate  $Z$  scores that were high were a cause for concern.

The two pairs (48541, 48543) and (18372, 18373) did give high  $Z$  scores, but it was already known at this point that these pairs corresponded to different rolled images of the same finger. There were three other pairs whose  $Z$  scores were above 3.6. One pair, (21852, 21853) gave a  $Z$  score of 3.64. The latent and the rolled prints were of fingers 7 and 8 of the same person. Further examination of this pair determined that part of finger 8 had intruded into the box containing the rolled print of finger 7. The computer algorithm had found this intrusion, when the pseudo latent for finger 8 was compared with the rolled print of finger 7. This is a somewhat impressive achievement.

One other pair, (12640, 21111), generated large  $Z$  scores in both orders. At the time the summary was written, it had not yet been determined whether these two prints were of the same finger. The Lockheed group compared all 20 fingerprints (taken from the two sets of 10 rolled prints corresponding to this pair) with each other. Not surprisingly, the largest scores were generated by prints being compared with themselves. The second highest score for each print was generated when that print was compared with the corresponding print in the other set of 10 rolled prints, and these second-highest scores were quite a bit higher than any of the remaining scores. This is certainly strong evidence that the two sets of 10 rolled prints corresponded to the same person.

The second experiment does not get at one of the central issues concerning latent prints, namely how the quality of the latent print affects the ability of the fingerprint examiner (or a computer algorithm) to match this latent print with a rolled one. Figures 2 and 3 show that latent prints do not look much like the central 21.7% of a rolled print. Yet it is just these types of comparisons that are used as evidence in court. It would be interesting to conduct a third experiment with the Lockheed data set, in which care was taken to create a more realistic set of latent prints.

**Exercise.**

1. By the middle of the 20th century, the FBI had compiled a set of more than 10 million fingerprints. Suppose that there are  $n$  fingerprint patterns among all of the people on Earth. Thus,  $n$  is some number that does not exceed 10 times the number of people on Earth, and it equals this value if and only if no two fingerprints are exactly alike.
  - (a) Suppose that all  $n$  fingerprint patterns are equally likely. Estimate the number  $f(n)$  of random fingerprints that must be observed in order that the probability that two of the same pattern are observed exceeds .5. Hint: To do this using a computer, try different values of  $n$  and guess an approximate relationship between  $n$  and  $f(n)$ .
  - (b) Under the supposition in part a), given that  $f(n) = 10$  million, estimate  $n$ . Note that it is possible to show that if not all  $n$  fingerprint patterns are assumed to be equally likely, then the value of  $f(n)$  decreases.
  - (c) Suppose that  $n < 60$  billion (so that at least two fingerprints are alike). Estimate  $f(n)$ .
  - (d) Suppose that  $n = 30$  billion, so that, on the average, every pattern appears twice among the people who are presently alive. Using a computer, choose 10 million patterns at random, thereby simulating the set compiled by the FBI. Was any pattern chosen more than once? Repeat this process many times, keeping track of whether or not at least one pattern is chosen twice. What percentage of the time was at least one pattern chosen at least twice?
  - (e) Do the above calculations convince you that no fingerprint pattern appears more than once among the people who are alive today?