3

Psychological Aspects of Forensic Identification Evidence

William C. Thompson
Simon A. Cole
University of California, Irvine

Forensic scientists are often asked to compare two items of evidence (e.g., blood stains, hairs, bullets, glass fragments, toolmarks, fingerprints) to determine whether they have (or might have) a common source (Inman & Rudin, 2001). Testimony about such comparisons is called *forensic identification evidence*. This type of evidence frequently plays a crucial role in criminal trials by helping to link the defendant to the crime. For example, the defendant could be associated with an item of evidence, often called a trace, recovered from an incriminating location. Or the victim of the crime could be associated with a trace found on the defendant's body, property, or vehicle. There are a variety of types of forensic identification evidence, including DNA profiles (e.g., Imwinkelried & Kaye, 2001; National Research Council [NRC], 1992, 1996; Thompson & Krane, 2003), latent prints (i.e., finger, palm, or sole prints taken from a crime scene; Benedict, 2004; Cole, 2004; Epstein, 2002; La Morte, 2003; Mnookin, 2001; Sombat, 2002), bitemarks (Saks, 1998), toolmarks (Schwartz, 2004; Springer, 1995), hair and fiber analysis, handwriting analysis, footprints, shoe prints, and comparative bullet lead analysis (Imwinkelried & Tobin, 2003; NRC, 2004).

Forensic identification evidence raises important psychological issues. One set of issues concerns forensic experts themselves: The manner in which experts make comparisons between items of evidence, the process

of judgment and decision making that underlies their determinations that items match or do not match, and the susceptibility of their methods to bias and error are all important potential areas for psychological study. A second set of issues concerns jurors. Among the key psychological issues are jurors' ability to comprehend testimony on complex and often arcane technical issues, their ability to draw appropriate conclusions from the probabilistic and statistical data that sometimes accompany forensic evidence, the manner in which jurors evaluate forensic evidence and integrate it with other evidence in the trial, and the susceptibility of their judgments to bias and inappropriate influence.

THE CONTENT OF EXPERT TESTIMONY

Our focus in this chapter is on expert testimony that purports to make what we call a "source attribution"—that is, a determination that two physical items have (or might have) a common source. In order to make an "inference of common source" (Inman & Rudin, 2001; p. 137) a forensic scientist will examine the two items and will often test or analyze the items in various ways. If the comparison reveals inconsistent features, the analyst will report an "exclusion." If the comparison reveals some amount of consistent information and no significant or unexplainable differences, the analyst will report a "match" or "inclusion" (Inman & Rudin, p. 137). If the evidence is too limited or ambiguous to make a determination, the analyst will report that the comparison is inconclusive.

Determining Whether Items Match: Objective Standards and Subjective Judgment

In some forensic disciplines, such as DNA analysis, there are objective standards for what constitutes a match between two samples. DNA analysts use sophisticated, computer-controlled instruments that produce output showing the genetic characteristics (called *alleles*) that the instrument detects at various locations (called *loci*) on the genomic DNA found in each sample (Thompson, Ford, Doom, Raymer, & Krane, 2003). Figure 3.1 shows DNA test results of five samples: blood from a crime scene and reference samples of four suspects. This analysis includes three loci, labeled "D3S1358," "vWA," and "FGA." Each person has two alleles (shown as peaks on the graphs) at each locus, one from the maternal portion and the other from the paternal portion of the chromosome (in some instances there is a single peak because the same allele was inherited

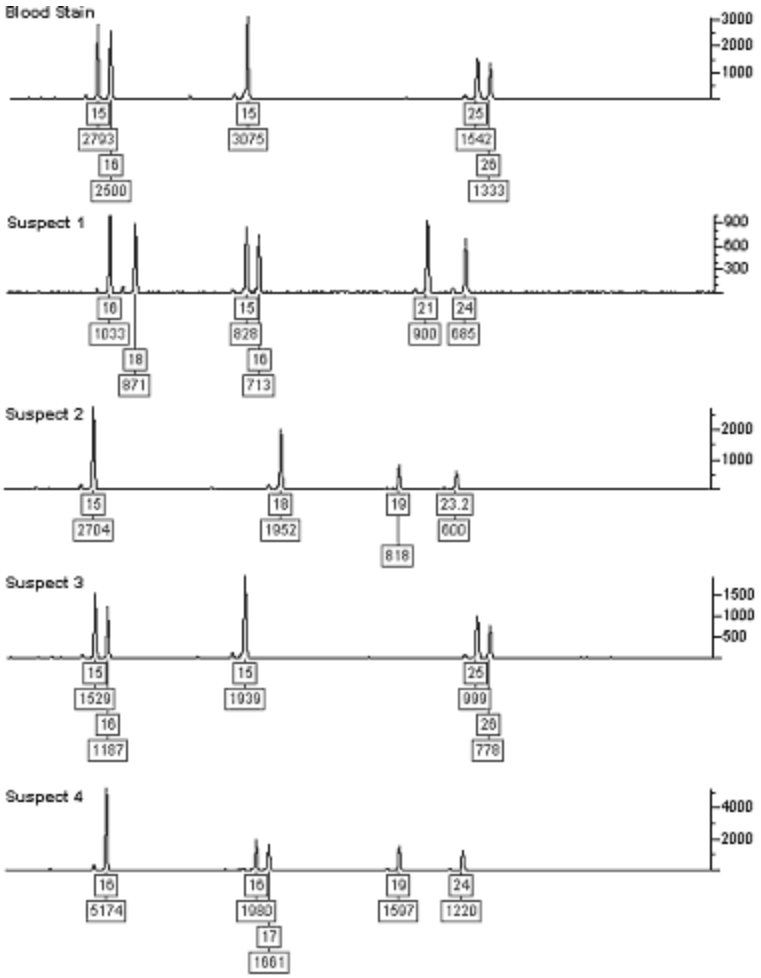

FIG. 3.1. DNA test results for five samples at three genetic loci; boxes immediately below each peak identify the allele it represents and the peak height (signal strength).

from both parents). The position of the peaks on each graph (known as an electropherogram) indicates the length of a small fragment of the DNA molecule at a specific location (*locus*) on the human genome. Forensic DNA tests examine areas of the genome where there tends to be variation among people in the length of these fragments, allowing samples from different people to be distinguished. The height of the peaks corresponds roughly to the quantity of DNA present.

The basic standard for a DNA match requires complete, one-to-one correspondence between the alleles in the two samples. As can be seen, the profile of Suspect 3 corresponds completely to that of the crime scene sample, hence it is a match that indicates Suspect 3 is a possible source of the blood at the crime scene. Suspects 1, 2, and 4 are eliminated as possible sources because one or more of their alleles differs from the crime sample.

For clear-cut test results like those shown in Fig. 3.1, interpretation is straightforward: All experts would agree that the DNA profile of Suspect 3 matches the DNA profile of the evidence, whereas the profiles of the other suspects do not match. However, DNA tests sometimes produce ambiguous results that are subject to multiple interpretations (Thompson, 1995, 1997a). Figure 3.2 shows a comparison between the DNA profile of a saliva sample from the skin of a sexual assault victim and the profile of a suspect. Experts differed over whether these two profiles match. For example, some experts thought the peak labeled "12" at locus "D3S1358" was a true allele, others thought it was merely noise in the system. The experts also differed over whether the peak labeled "OL allele" at locus "FGA" was a spurious anomaly that could be safely ignored, or whether it might be hiding another allele. When interpreting ambiguous results like those shown in Fig. 3.2, human analysts rely heavily on subjective judgments to distinguish signal from noise, explain anomalies, and account for discrepancies (Thompson, 1995, 1997a). Consequently, even though there is an objective standard for what constitutes a DNA match, analysts' interpretation of the test results in some cases still entails an element of subjective judgment (Risinger, Saks, Thompson, & Rosenthal, 2002; Thompson, Ford, et al., 2003).

Once a DNA analyst determines that two profiles match, the next step is to estimate the probability that the match could have occurred by coincidence. This is typically done by consulting databases to determine the frequency of the matching alleles in various reference populations. These frequency estimates, often called *random match probabilities* (RMPs), are then presented to the jury along with the DNA evidence.

In many forensic disciplines, standards for what constitutes a match are vague, poorly defined, or even nonexistent, and consequently the match determination rests even more heavily on subjective judgment. When asked to determine whether two bullets could have been fired from the same gun, for example, firearms examiners will typically examine the bullets under a comparison microscope to see if the striations (markings) on the bullet are similar (Schwartz, 2004, 2005). However, no standards exist to specify how many or what kind of striations must correspond before the analyst may declare the two bullets to match. According to

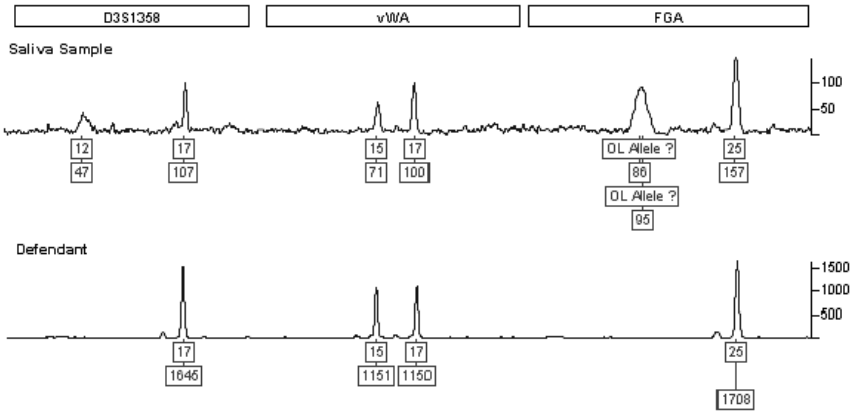

FIG. 3.2. DNA test results for saliva sample and suspect at three genetic loci.

Schwartz, analysts often declare a match notwithstanding some discrepancies between striation patterns, so long as the analyst concludes that the discrepancies are not significant. Whether a discrepancy is significant is itself a subjective determination for which no standards exist. Hence, the match determination is entirely subjective. The process leading to the determination occurs entirely within the mind of the examiners (while they look at a magnified image of the bullets). Often the only record of the determination is a written conclusion that the two bullets match or do not match.

Latent print examiners try to identify impressions of "friction ridge skin" (skin from the fingers, palms, and soles). They compare impression of unknown origin (which, following Champod, Egli, & Margot, 2004, we call *marks*) with exemplars of known origin (prints). Marks are typically faintly visible latent prints that must be developed using methods such as powders, chemical fumes, or alternative light sources. Although computers now assist human examiners in searching fingerprint databases for candidate matches, computers never make a final determination of source attribution. That determination, contrary to popular television depictions, is always made by a human examiner.

Because of pressure distortion, printing artefacts, and various other effects, no two imprints even of the same finger are exactly alike. Therefore, latent print examiners (LPEs) may conclude that two impressions were made by the same finger despite evident differences between them. In general, LPEs look at ridge characteristics (Fig. 3.3), which are locations where the friction ridges end abruptly or bifurcate.

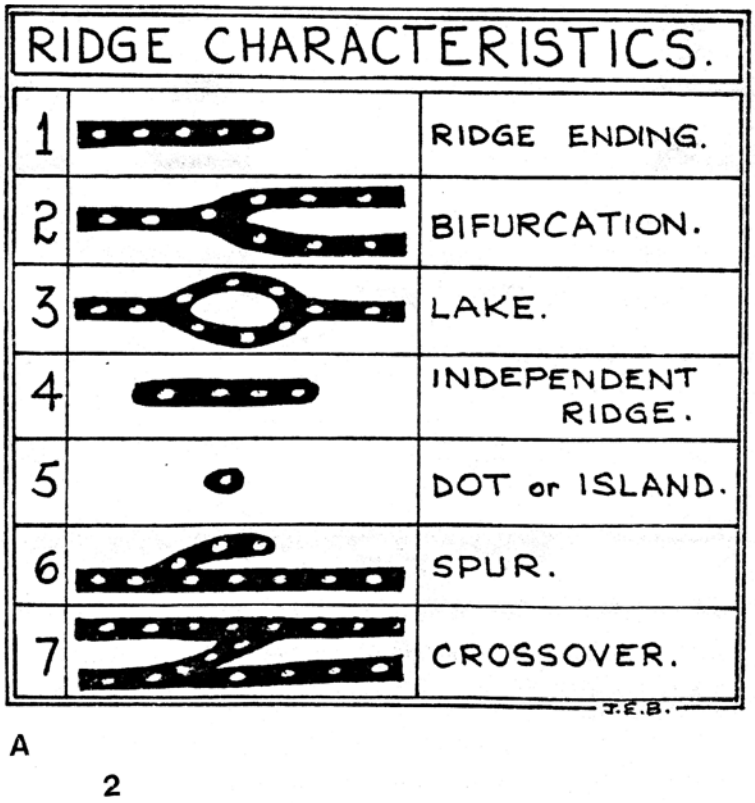

FIG. 3.3. Ridge characteristics.

There is some debate within the profession over whether examiners should look only at such ridge characteristics or whether even finer friction ridge detail, such as the shapes of ridges themselves, the attributes of the characteristics (as opposed to merely their relative location), and the location of pores, should also be utilized (Ashbaugh, 1999). The one-dissimilarity rule states that a single unexplainable dissimilarity necessitates a conclusion of exclusion (Thornton, 1977). However, LPEs must constantly make decisions about whether differences should count as unexplainable dissimilarities (or differences or discrepancies) or explainable dissimilarities (or distortions; Leo, 1998; Scientific Working Group on Friction Ridge Analysis Study and Technology, 2003a).

Assuming that the examiner believes all the ridge detail is consistent, the examiner must then decide whether there is sufficient consistent

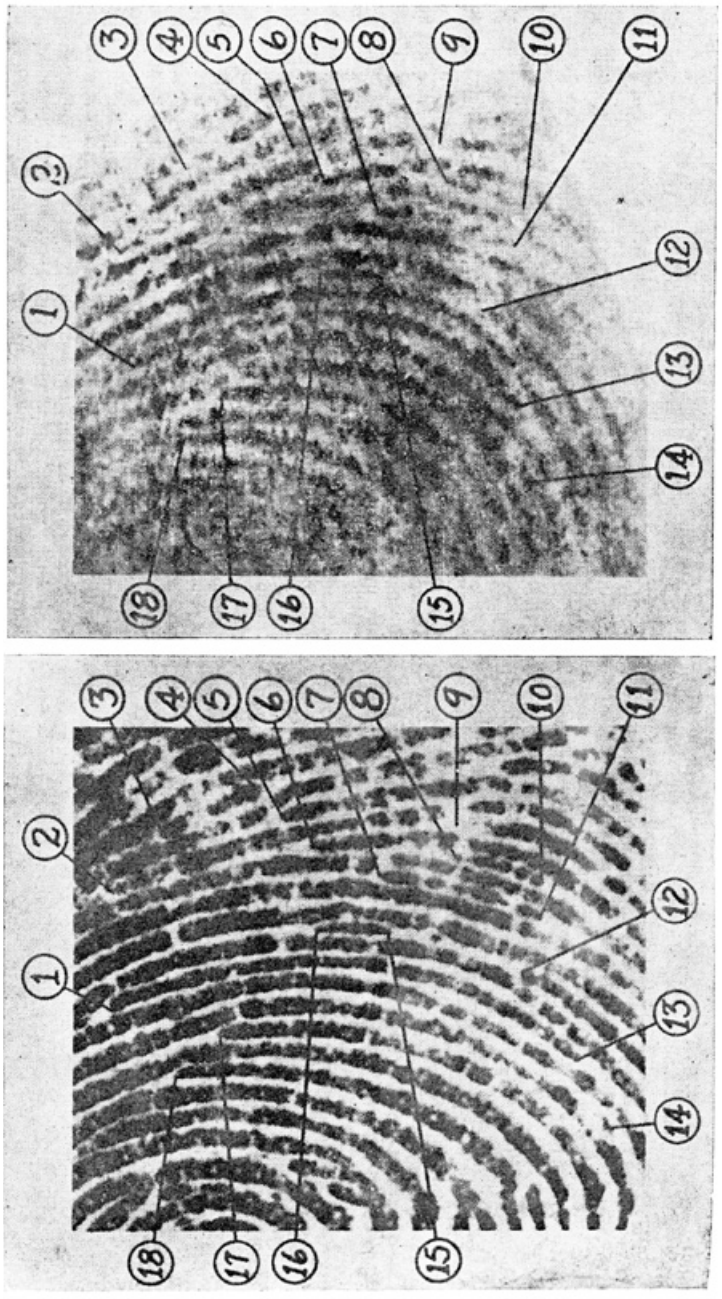

FIG. 3.4 A known-to-unknown fingerprint comparison.

ridge detail to warrant a conclusion of individualization. In contrast to the situation for DNA, no data estimating the rarity of particular ridge details or combinations of details exists (Zabell, 2005, pp. 155–156). Consequently, LPEs have no scientific basis on which to estimate the probability of a random match between two impressions, and they present no statistics in connection with their testimony. If they find sufficient consistent ridge detail they simply declare a positive identification or individualization, claiming that the potential donor pool for the mark has been reduced to one and only one area of friction ridge skin to the exclusion of all other friction ridge skin in the world.

How do they determine whether there is sufficient consistent ridge detail to warrant such a strong conclusion? LPEs follow one of two approaches (Champod, 1995; Cole, 1999). One school of thought, strongest in the continental European countries, is to set a threshold specifying the minimum number of corresponding ridge characteristics necessary to make a conclusion of individualization. (Fig. 3.4) The threshold varies from country to country but is generally between 8 and 16 matching points ("European Fingerprint Standards," 2002). There is, however, no empirical basis for these thresholds, which are simply agency or national norms. There is no way to know whether any particular threshold is high enough to justify the claim of individualization (i.e., the claim that the RMP is 0). Moreover, misidentifications have been known to occur even under the highest point thresholds (Cole, 2005, p. 1024).

A second school of thought, which is most prevalent in the Anglo-American countries, rejects arbitrary point thresholds as unscientific. Instead, it advocates that the LPE intuit when the amount and rarity of the corresponding ridge detail is sufficient to warrant a conclusion of individualization (Ashbaugh, 1999). The necessary amount and rarity of consistent friction ridge detail is not defined, other than—tautologically—by reference to the examiner's own judgment: "Individualization occurs when a latent print examiner, trained to competency, determines that two friction ridge impressions originated from the same source, to the exclusion of all others" (Scientific Working Group on Friction Ridge Analysis Study and Technology, 2002, §3.3.1).

Little is known about the manner in which examiners make these intuitive judgments, although studies have found variability among examiners in their analysis of features contained in marks (Evett & Williams, 1996; Langenburg, 2004). There have been no studies of the amount or type of ridge detail necessary to convince examiners that the detail is rare enough to reduce the random match probability to zero. This

is an unexplored psychometric question of considerable potential importance to the legal system.

It should now be clear that forensic scientists in all disciplines rely partly on subjective judgment to reach conclusions, and that conclusions in some areas are almost entirely subjective. Commentators have argued that forensic scientists' reliance on subjective judgment to make match determinations is problematic because such judgments are subject to observer effects and other forms of bias (Risinger et al., 2002; Thompson, For, et al., 2003). For example, these judgments may (consciously or unconsciously) be affected by "domain-irrelevant" information derived from the investigative process (Risinger et al.). One recent study showed that, in cases in which the fingerprint evidence was ambiguous, naive subjects comparing fingerprints were more likely to reach conclusions of identification if they had been exposed to emotionally stimulating materials, such as graphic crime-scene photographs and descriptions of violent crimes (Dror, Péron, Hind, & Charlton, 2005). Another study found that professional LPEs could be influenced to reach different conclusions when given misleading contextual information (Dror, Charlton, & Péron, 2006). A recent FBI report (Stacey, 2004) took the position that even professional LPEs are more susceptible to observer effects in high-profile cases (though we regard this conclusion as unsupported by evidence, as yet; Cole, 2005; Thompson & Cole, 2005). Clearly this is an important area for further study.

The Meaning of a Match: Class and Individual Characteristics

Forensic scientists distinguish two types of matches: those in which the items share class characteristics—features that place them in the same category or class (categories or classes populated by more than one person or thing); and those in which the items share individual characteristics—a unique combination of features. Matching blood types and DNA profiles are class characteristics to the extent that more than one person would have them.

Matching fingerprints and some matching toolmarks are said to be individual characteristics (although the claims that these disciplines can actually know when they have narrowed the potential donor pool to one have been greeted with skepticism in the scholarly literature and are sometimes challenged in court; Bunch, 2000; Cole, 2004; Nichols, 2003).

When testifying about matches between items most experts use one of the four general approaches summarized and illustrated in Table 3.1.

TABLE 3.1
Four Approaches to Expert Testimony About Source Attributions

Type of Testimony	*Simple Match*	*Match Plus Statistic*	*Qualitative Statement*	*Individualization*
Typical statements in expert testimony	The samples "match" or "are consistent" and therefore "could have a common source"	The samples "match" or "are consistent" and the frequency of the matching characteristics in [a reference population]is 1 in "X"	The samples "probably have a common source" or the comparison provides "very strong evidence" that the samples have a common source	The samples have been "positively identified" as being from the same source "to the exclusion of [all other such items in the world]"

When testifying about matches on class characteristics, experts typically use one of the approaches we label "simple match," "match plus statistic," or "qualitative statement." When testifying about matches on what they believe to be individual characteristics, experts typically use the approach we call "individualization."

Simple Match (No Statistics). In some instances, analysts simply testify that two items share certain class characteristics without providing a numerical characterization of the rarity of the characteristics or the strength of the match for showing the items have a common source. We call such testimony simple match testimony. For example, the analyst may state that two fibers are both composed of blue rayon of the same diameter and are not otherwise distinguishable, but the analyst will present no statistics on the frequency or rarity of blue rayon or on the probability that fibers from a different source would share the matching characteristics. In some cases, as when an analyst testifies that a footprint was made by a size 10 Nike shoe of a certain type, a juror might reasonably be expected to make a rough commonsense-based estimate of the rarity of the class characteristic. (Or, industry data could theoretically be presented to the jury.)

When reporting a match that involves class characteristics, forensic analysts typically testify that the matching items could have a common

source. For example, an analyst might say that two fragments of glass could have come from the same broken window or that two bullets with similar metal composition could have come from the same box. During cross-examination, analysts typically concede that the samples could also have come from different sources that happen to be indistinguishable. However, anecdotal evidence suggests that expert witnesses sometimes overstate the value of a match on class characteristics, as when a hair analyst gives an opinion "that there was a transfer of hair from the Defendant to the body of" the victim even though experts accept that human hair cannot be uniquely identified (Stafford Smith & Goodman, 1996, p. 273; see also Yardley, 2001).

Match Plus Statistics. Obviously, the preceding testimony presents difficulties for a jury. The probative value of a match for proving two items have the same source may vary greatly depending on the commonality of the matching characteristics within a relevant population of persons or things. In addition, there is the potentially prejudicial nature of the word *match* itself. Jurors might infer that *match* implies something more akin to individualization (see later discussion), rather than merely the consistency between certain (perhaps quite common) attributes (American Board of Forensic Odontology, 1999). One way of aiding jurors is to accompany the testimony about a match with a statistical estimate of the rarity of the matching characteristics within a relevant population of people or things. For most types of forensic evidence, however, statistical data on the frequency of class characteristics are limited or nonexistent. The major exceptions have been serology and DNA testing. In connection with DNA and serology matches, forensic analysts nearly always present statistics on the frequency of the matching characteristics (in some reference population or populations). Indeed, courts in many jurisdictions refuse to admit evidence of a DNA match unless it is accompanied by valid statistical estimates of the frequency of the matching characteristics (Kaye & Sensabaugh, 2000; Thompson, 1997b). A DNA analyst might testify, for example, that the matching DNA profile would be found in approximately 1 person in 8 million among Caucasians, 1 person in 10 million among African Americans, and 1 person in 5 million among Hispanics (Thompson, Taroni, & Aitkin, 2003).

Analysts compute the frequency of DNA profiles based on studies of the frequency of the various genetic alleles that make up the profile in various populations. Because the alleles are assumed to be statistically independent of one another, the frequencies of the various matching

alleles are multiplied together to determine the frequency of the entire profile.

Data on the frequency of matching DNA characteristics can be presented in a variety of ways. If the frequency of the matching DNA profile is 1 in 1 million in a particular reference population, for example, jurors might be told that 1 person in 1 million would have the profile or that the random match probability (RMP) is 1 in 1 million (Koehler, Chia, & Lindsey, 1995). In the United Kingdom, the Forensic Science Service typically reports, in such a case, that the "chance of observing the DNA profile if it originated from another individual unrelated to [defendant]" is 1 in 1 million. Some laboratories in the United States convert the frequency into a likelihood ratio and report, for example, that the DNA profile found in the evidentiary sample is 1 million times more likely if the evidentiary sample came from the defendant than if it came from an another unrelated individual (Butler, 2005).

Because likelihood ratios are the most precise way to characterize the results of paternity tests and DNA comparisons involving mixed samples (Evett & Weir, 1998), juries often hear likelihood ratios in connection with such evidence. For example, a jury might be told that the results of a paternity test are *X* times more likely if the accused man is the father of a child than if a man chosen randomly (from some reference population) is the father. Or a jury might be told that the mixed DNA profile found in a bloodstain is *X* times more likely to have occurred if the stain consists of a mixture of blood from the victim and the defendant than if the stain is a mixture of blood from the victim and a randomly chosen man who is unrelated to the defendant (see Thompson, 1996, 1997b, for discussion of problematic aspects of likelihood ratio computations).

Estimates of the error rate or false-positive rate of forensic comparisons are rarely presented in criminal trials. In theory, data on false-positive rates would be highly relevant. When evaluating the strength of a match for proving that two items have a common source, the jury must consider two factors. One factor, as already discussed, is the probability of a *coincidental match*. The second factor is the probability of a *false positive*. A false positive, as we use that term here, occurs when a forensic expert erroneously reports a match between two samples that in fact do not match on the characteristics being compared. A false positive might occur due to error in the collection and handling of the samples (e.g., mislabeling), incorrect reading or misinterpretation of test results, or incorrect reporting of test results (Thompson, Taroni, & Aitkin, 2003). Although experience has shown that false positives can occur (Koehler, 1995, 1997; Peterson & Markham,

1995a, 1995b; Thompson, Taroni, & Aitkin, 2003), the rate at which they occur is difficult to estimate based on existing data, and even if the overall rate of error for a particular type of comparison were known, the relevance of the overall statistic to the probability of error in a particular case would be debatable (see, e.g., NRC, 1996: arguing that general error rates for DNA testing have little relevance to particular cases; see Koehler, 1997; Thompson, 1997a, for alternative views). Forensic experts rarely testify about error rates. When opposing counsel try to introduce such data (e.g., data on error rates in proficiency tests) courts often exclude it on grounds that its relevance to the case at hand is too tenuous. Thompson, Taroni, and Aitkin discussed the paradoxical nature of the situation with regard to DNA testing, where courts require statistics on the frequency of matching DNA profiles, but not on the probability of a false positive: "It is considered essential to know, with a high degree of scientific certainty, whether the frequency of random matches is 1 in 1000, 1 in 10,000, or one in one million, but unnecessary to have comparable estimates of the frequency of false positives" (p. 47). Thompson, Taroni, and Aitkin suggested the two types of data are treated differently in part due to a fallacious belief that data on false positives are less important.

In some disciplines, expert witnesses have testified to probabilities using numbers based on unrepresentative databases, faulty statistical inferences, or both. Expert witnesses on microscopic hair comparison, for example, have made probabilistic statements to juries based on inappropriately applying the product rule to situations in which the requirement of statistical independence is not met (*Peer review report: State v. Bromgard*, 2002; Stafford Smith & Goodman, 1996, pp. 267–271).

Qualitative Assessments Of Certainty. Some disciplines, recognizing both the need to convey to the juror an assessment of the certainty of the source attribution and the lack of any statistical data on which to base any such assessment, offer qualitative guidelines for calibrating the certainty of source attributions. For example, the American Board of Forensic Odontology promulgated the following "degrees of certainty" for bitemark testimony:

1. Source attribution to reasonable medical certainty
2. Probable source attribution
3. Possible source attribution
4. Improbable that suspect is source
5. Suspect is not source
6. Inconclusive (1999)

Although this scale brings greater clarity to the phrasing of bitemark testimony, it is unclear whether it renders such testimony more valid. According to some commentators, the analyst's decision as to where any given comparison sits along the scale rests entirely on a subjective evaluation of bitemarks (Saks, 1998).

Individualization. Thus far, we have discussed testimony about matches between items that share class characteristics. In some forensic disciplines, experts believe that when comparing two items they can identify characteristics, or sets of characteristics, that are unique. These supposedly unique features are individual characteristics. When such characteristics are found, forensic experts say they have individualized the source of the items—that the potential sources have been reduced to one and therefore that the two items being compared necessarily have a common source.

In at least one discipline, latent print identification, expert witnesses are mandated by professional guidelines to only give testimony of individualization. Current guidelines mandate that LPEs may offer only three conclusions in their reports or their testimony:

1. Individualization
2. Exclusion
3. Inconclusive (Scientific Working Group on Friction Ridge Analysis Study and Technology, 2003b, p. 358–359)

Individualization is defined as "the determination that corresponding areas of friction ridge impressions originated from the same source to the exclusion of all others (identification)" (Scientific Working Group on Friction Ridge Analysis Study and Technology, 2003a, p. 12).

The Association of Firearms and Toolmark Examiners (AFTE) "encourages" expert witnesses to phrase their conclusions as follows:

1. IDENTIFICATION—Agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool.
2. INCONCLUSIVE—A. Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.

B. Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.

C. Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for elimination.
3. ELIMINATION—Significant disagreement of all discernible class characteristics and/or individual characteristics.
4. UNSUITABLE—Unsuitable for microscopic comparison. (AFTE, 1998, p. 86)

Whether the AFTE's conception of identification is tantamount to individualization, whether the professional organization's encouragement amounts to a mandate, and to what extent these guidelines are adhered to in practice are topics of spirited debate (Nichols, 2005; Schwartz, 2005, p. 13).

Other disciplines may occasionally give evidence in terms of individualization. Indeed, even DNA analyses, which, as we have seen, are usually accompanied by statistics, may sometimes be given as individualizations. When the estimated frequency of the matching DNA profile is very low, some labs simply state "to a scientific certainty" that the samples sharing that profile are from the same person. For example, the FBI laboratory claims two samples are from the same person if the estimated frequency of the shared profile among unrelated individuals is below 1 in 260 billion. Other labs use different cutoff values for making identity claims. All of the cutoff values are arbitrary: There is no scientific reason for setting the cutoff at any particular level.[1] Moreover, these identity claims can be misleading because they imply that there could be no alternative explanation for the match, such as laboratory error or accidental cross-contamination of samples, and they ignore the fact that close relatives are far more likely to have matching profiles than unrelated individuals.

One curious aspect of the tradition of phrasing latent print evidence as individualizations is that expert witnesses are banned from using probabilities in their testimony. A 1979 Resolution of the International Association for Identification (IAI), the main professional organization for LPEs in North America, stated, "Any member, officer or certified latent print examiner who provides oral or written reports, or gives

[1]The rationale for the FBI's threshold is that the probability of finding a duplicate profile among unrelated individuals in a population the size of the United States' drops below 0.05 when the frequency of the profile is below 1 in 260 billion (Budowle, Chakraborty, Carmody, & Monson, 2000). Budowle et al. acknowledged that the choice of 0.05 as a cutoff is a "policy decision." When the FBI's policy was first announced, it was touted as a "scientific breakthrough." (FBI, 1989) It would be more accurate, in our view, to call it a semantic breakthrough.

testimony of possible, probable, or likely friction ridge identification shall be deemed to be engaged in conduct unbecoming such member, officer, or certified latent print examiner" (p. 1).[2]

Typically, latent print testimony is given in one of two ways. In some cases, LPEs testify that they have identified the mark as belonging to the defendant or that the mark matched the defendant. In other cases, LPEs will testify that the defendant made the mark. In either of these scenarios, LPEs often buttress the conclusion by testifying that they are positive, that the match is a positive identification or a positive match, or that the identification of the defendant is to the exclusion of every other individual in the world. One laboratory's protocol suggests the following testimony for a garden variety latent print comparison: "The latent impression developed on exhibit ____ has been identified as the fingerprint impression of _____" (New Hampshire State Police Forensic Laboratory, 2005, p. 4).

In many cases, LPEs quantify the certainty of their conclusions as 100%. Although probabilistic conclusions are purportedly banned, this rule apparently refers only to probabilities less than 1. The absurdity of this situation was highlighted in a recent case, *Michigan v. Ballard* (2003). A LPE testified "that she was '99 percent' certain that defendant's fingerprint was found in the stolen car." A majority of a Michigan Court of Appeals panel found that this testimony had "no scientific foundation" and "no demonstrated basis in an established scientific discipline and rested solely upon Ms. Dyke's [the LPE's] personal opinion" (*Michigan v. Ballard*, p. 9).[3] The irony, of course, is that the probability Dyke offered in her testimony was not too high, but rather too low. Had Dyke testified, like many of her colleagues, that the match was 100% certain, her conclusions would likely not have garnered the court's attention.

STRENGTH OF THE SCIENTIFIC FOUNDATION

There is great variability in the scientific foundation underlying different types of forensic science testimony. Whereas DNA testing is relatively well validated through extensive programs of research on the reliability

[2]This odd mandate originated in noble intentions. Historically, the aim appears to have been to discipline LPEs to give testimony only when they were absolutely certain. This was supposed to inculcate an ethic of conservatism in the practice (Cole, 1998). The danger, of course, is that the same policy may have the effect of inducing examiners to exaggerate the probative value of their findings. In addition, it creates the false impression that latent print evidence is somehow nonprobabilistic (Champod et al., 2004; Champod & Evett, 2001).

[3]This finding was subsequently reversed on appeal.

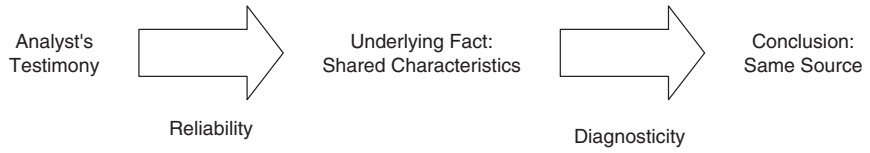

FIG. 3.5. Reliability and diagnosticity.

of the laboratory methods and the rarity of DNA profiles (NRC, 1992, 1996), many types of forensic evidence have little or no validation—leaving uncertainty about both the reliability of the procedures for determining matches and the value of a match for proving the matching items have a common source (Saks & Koehler, 2005).

When discussing the scientific foundation for forensic evidence, it is helpful to distinguish two elements that Schum and his colleagues labeled *reliability* and *diagnosticity* (Schum, 1994; Schum & DuCharme, 1971). The reliability of forensic testimony is its value for proving an underlying fact: typically that two items share class or individual characteristics. The diagnosticity of forensic testimony is the value of the underlying fact (the shared characteristics) for establishing that two items have a common source. Figure 3.5 presents this relationship.

Although many areas of forensic science are so poorly validated that no reliable data are available on either reliability or diagnosticity, that situation should improve in the near future as forensic scientists come under increasing pressure to improve their validation (e.g., Kennedy, 2003; Saks & Koehler, 2005).

Validation of DNA Tests

When DNA evidence was first introduced in U.S. courts in the late 1980s it was heralded as "the greatest advance in crime fighting technology since fingerprints" (*People v. Wesley*, 1988). After a brief honeymoon period in which DNA testimony was accepted without challenge, however, a number of scientific critics emerged who questioned both the reliability and the diagnosticity of forensic DNA (Thompson, 1993). The most heated debate concerned the rarity of DNA profiles. The methods that forensic laboratories were using to compute random match probabilities assumed the statistical independence of a number of distinct genetic markers identified by the tests. After several prominent scientists and an NRC (1992) report questioned the independence assumptions, a number

of courts held DNA evidence inadmissible on grounds that the underlying method for statistical estimation was not generally accepted in the scientific community.[4] These courts reasoned that DNA evidence is meaningless in the absence of valid statistical estimates of the random match probability. Hence, by their analysis, the scientific dispute over the methods for estimating the frequency of DNA profiles precluded admissibility of DNA evidence altogether. Interestingly, courts have never applied this analysis to other types of forensic evidence. Courts may have treated DNA differently because it appeared to be such powerful evidence or because of its novelty.

In any event, these court rulings had a positive effect on the science underlying forensic DNA testing: The prospect of negative admissibility rulings spurred much-needed research on the distribution of genetic markers in human populations (Thompson, 1997a). It also prompted forensic scientists to develop validation standards designed to assure the reliability of DNA evidence (Butler, 2005). This population research made it possible to assess the independence assumptions underlying the statistical methods. After some tweaking of the methods (NRC, 1996) the balance of scientific opinion tipped strongly in their favor and the admissibility of DNA testing was assured.

Although current DNA technology is capable of producing highly reliable results, questions are sometimes raised about the quality of laboratory work. Key issues include the potential for biased or mistaken interpretation of laboratory results and the possibility for error due to mishandling of samples. Acknowledging problems with the quality of early DNA testing procedures, a 1992 report of the NRC called for broader scrutiny of forensic DNA testing by a scientific body from outside the law enforcement community.

In response, the FBI created its own advisory body that was initially called the Technical Working Group for DNA Analysis Methods (TWGDAM) and more recently called the Scientific Working Group for DNA Analysis Methods (SWGDAM). The FBI director appoints its members. Although it has not satisfied all critics of forensic laboratory practices, this body has been credited with issuing guidelines that have improved the quality of forensic DNA work. For example, SWGDAM guidelines

[4]At that time, most states applied the Frye standard (Frye v. United States, 1923), which requires that a novel form of scientific evidence be generally accepted in the relevant scientific community before the evidence can be admitted in court.

call for each analyst to take two proficiency tests each year.

Another quality assurance mechanism is laboratory accreditation. The American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLAD-LAB) is a nonprofit organization that reviews the protocols and procedures of forensic DNA laboratories and issues a certificate of accreditation to those meeting its standards. To help assure the competence of laboratory workers, a professional organization called the American Board of Criminology has developed a certification program for DNA analysts.

Despite these efforts, problems occasionally come to light. Errors have occurred in proficiency tests, although they are infrequent. Occasional errors arising from accidental switching and mislabeling of samples or misinterpretation of results have come to light in court cases. In several cases, misinterpretation of a DNA test contributed to wrongful convictions that were later overturned when more extensive DNA tests, by other laboratories, proved the inmates' innocence (Thompson, Ford, et al., 2003).

A 1996 report of the NRC suggested that retesting of samples is the best way to address remaining concerns about the quality of laboratory work. The great sensitivity of PCR-based DNA tests makes it possible to split samples for duplicate analysis in most cases.

Latent Prints

In 1997, Stoney wrote, "From a statistical viewpoint, the scientific foundation for fingerprint individuality is incredibly weak" (p. 67). Nothing has occurred in the interim that would lead us to revise this assessment, but we would add that the scientific foundation for the accuracy of latent print identification (in our opinion, the more crucial question) is also weak, from any viewpoint. Given the extraordinary certainty of latent print testimony detailed earlier, this weakness is especially glaring. After nearly a century of courtroom use, no validation studies have been performed to assess LPEs' fundamental claim: that they can make correct source attributions (Haber & Haber, 2003). Preliminary studies have measured the accuracy of naive subjects, not LPEs (Boychuk & Vokey, 2004; Tangen, Vokey, & Allan, 2000; Torry & Vokey, 2005; Vokey, Tangen, & Boychuk, 2004).

Part of the problem is that no metric has been devised to measure the amount of information contained in a mark or the amount of corresponding information between two impressions. Although it certainly seems possible to develop systems for classifying features of marks and to do studies on the frequency of those features in various populations, no one has yet

published such studies. So there is no scientific basis for estimating the probability of a random match to a given configuration of friction ridge detail. Nevertheless, examiners assume that certain patterns are so rare as to be unique, and that they have the ability to identify such patterns.

A key psychological (and epistemological) question is whether LPEs can really make such determinations accurately. Consider for a moment the mental process required to determine that a known area of friction ridge skin (e.g., a defendant's fingertip) is the only possible source of a mark found at a crime scene, to the exclusion of all other friction ridge skin in the universe. One obvious approach would be to estimate the rarity of the configuration of ridge characteristics and then to judge the likelihood that a configuration of characteristics this rare would be duplicated in any other area of friction ridge skin anywhere in the world.[5] However, this would require examiners to make accurate estimates of extremely small random match probabilities, a daunting prospect given the difficulties people have with probability estimation (Lichtenstein, Fischhoff, & Phillips, 1982; Plous, 1993, chap. 12). It seems unlikely that examiners, relying entirely on intuitive judgment and without computational formulas or other forms of guidance, would be able to accurately determine whether, for example, the random match probability for a specific configuration of ridge characteristics is 1 in 1 million, 1 in 1 billion, or 1 in 1 trillion. Yet the differences among these estimates will have a huge impact on the likelihood of there being a duplication in the world's population of friction ridge skin, which is estimated to range from 50 to 60 billion fingers (taking into account only fingers of individuals currently living, and excluding the dead, the not yet conceived, nonhuman primates, palms, and soles). Moreover, even if examiners could determine with precision the random match probability, it would be difficult for them, without computational aids, to estimate the probability of a duplication in such a large population of fingers. (Suppose the examiner knew, for example, that a particular set of features would be found on 1 finger in 1 trillion. What is the likelihood of there being a second such finger in a population of 60 billion?) People's general tendency to underestimate

[5]The examiner would also need to consider how low the probability of a duplication would need to be to justify the claim that the probability of a duplication is (effectively) zero. As noted earlier in connection with DNA evidence, the threshold for declaring uniqueness is ultimately a policy judgment rather than a scientific question. Yet this is a question examiners must implicitly answer every time they make an intuitive judgment of individualization.

the probability of disjunctive events (Bar-Hillel, 1973) is likely to produce underestimates of the probability of a duplication, and hence may lead examiners too readily to judgments of uniqueness and individualization. In light of these problems, LPEs' claim that they can accurately determine through intuitive judgment when the random match probability has effectively been reduced to zero seems implausible.

Of course it is not clear that analysts actually make their determinations of individualization in the manner posited here: that is, by first estimating the rarity of the matching ridge detail and next estimating the likelihood of a duplication among the world population of fingers. They may rely on a simpler heuristic strategy. The Interpol European Expert Group on Fingerprint Identification (IEEGFI, 2004, §8.12.7) noted that, if the task is individualization, "The scientific problem that the fingerprint examiner is facing is to single out the donor of the print out of a potential of over 60 billion fingerprints." And yet, presented with a print pair, the examiner faces a potentially overwhelming temptation to simply ask, "Do I *think* it's him or not?" (original emphasis; IEEGFI, 2004, § 12.7). Although IEEGFI claims examiners can resist this temptation through training and sheer self-discipline, LPEs have offered no proof whatsoever that they can. Nor is there any evidence that analysts who adopt the approach recommended by the IEEGFI perform fingerprint identifications more accurately than those who do not.

What foundation, then, does latent print identification have? First, there are statistical models that suggest that friction ridge skin patterns are highly variable and that exact duplication is unlikely (Pankanti, Prabhakar, & Jain, 2002; Stoney, 2001). This is useful—if duplication could be demonstrated, then the value of latent print evidence would be greatly reduced—but it is of limited utility. The principal problem is that little is known beyond the broad-brush statement that exact duplication is unlikely. The further question, of how different the most similar friction ridge patterns within a given population are, remains unanswered and, without a metric for similarity, unanswerable.

Second, there is anatomical research on the development of friction ridge skin (Babler, 1975, 1978, 1983, 1987, 1990, 1991; Bonnevie, 1924; Cummins & Midlo, 1943; Wertheim & Maceo, 2002; Whipple, 1904). This literature establishes that embryonic temperature and pressure are significant variables in the development of friction ridge skin. The further conclusion "that the process of prenatal development causes an *infinite variation* [italics added] of individual friction ridge details" (Moenssens, 2003, p. 32), does not seem, to us, to be warranted by the

research. In any case, understanding the development of friction ridge skin is of limited value in establishing whether an expert community can make correct source attributions from trace impressions of that skin.

Third, after nearly a century of courtroom use, latent print identification has produced a relatively small number of exposed false-positive errors (Cole, 2005). The value of this datum is of course undermined by the unlikelihood of latent print false-positive errors being exposed. Although the test of time remains a primary resource for courts' and LPEs' continued confidence in latent print identification, it makes for a weak foundation for any knowledge claim, let alone claims as strong as those made by LPEs (Cole, 2004, 2006).

Indeed, although it might be said that the practice of latent print identification has some scientific foundation in the variability of friction ridge skin, latent print expert testimony might be said to have no scientific foundation at all. If we assume, for example, that a LPE has evaluated a mark and a print and concluded that all the friction ridge detail in the mark is consistent with the print, the LPE has no scientific data with which to assess the probative value of that conclusion and convey it to the jury. The latent print profession's historical solution of simply rounding the probative value up to 1 cannot be sustained.

Other Disciplines

With the notable exception of forensic DNA analysis, most forensic identification disciplines can offer only startlingly weak scientific foundations for their testimony. Although most areas boast substantial scientific literatures concerning the detection, recovery, and classification of crime scene traces, it is the scientific foundation for source attribution testimony that is typically lacking. Toolmark identification, like latent prints, lacks a scientific foundation to support the inordinately strong expert testimony that is mandated within the profession. As Schwartz (2005) observed, "Firearms and toolmark examiners have taken only the most minimal steps towards developing the necessary statistical empirical foundations for their identity claims" (p. 4).

Other forensic identification techniques, such as bitemark analysis, handwriting identification, and microscopic hair comparison, have similar problems. Like latent prints and toolmarks, bitemark identification has a solid scientific foundation that establishes that bitemarks can provide probative information, but little research that establishes the validity, or measures the accuracy, of bitemark identification in practice. Bowers

(2002) concluded, "Reliability of dental opinion historically is based on intuition derived from the expert's 'experience,' not scientific data" (p. 259). In addition, proficiency test data have revealed disappointing accuracy rates among practicing forensic odontologists (Bowers). Bitemark identification distinguishes itself from latent prints and toolmarks primarily by its more modest testimonial claims.

Handwriting identification has shown poor results on proficiency tests (Risinger, 2002). Many other areas of forensic analysis have generated poor results on (nonblind) proficiency tests (Peterson & Markham, 1995a, 1995b). No controlled empirical studies of microscopic hair comparison have been performed (Stafford Smith & Goodman, 1996, p. 234). However, a study in which hair comparison conclusions were compared to results of mitochondrial DNA tests on the same evidence showed high rates of disagreement (Houck & Budowle, 2002; Risinger & Saks, 2003; for a dissenting interpretation, see Houck, 2004).

JURY RESEARCH

A number of studies have examined mock jurors' reactions to forensic evidence (for reviews see Kaye & Koehler, 1991; Koehler, 2001; Thompson, 1989). One line of research examined how statistics on the RMP affect the weight that jurors give to a forensic match (Faigman & Baglioni, 1988; Goodman, 1992; Smith, Penrod, Otto, & Park, 1996; Thompson & Schumann, 1987a). These studies asked jurors to revise an initial (prior) estimate of a suspect's probability of guilt after receiving forensic evidence implicating the suspect. The studies compared jurors' posterior judgments to the posteriors specified by Bayesian models. In general, jurors' judgments of the posterior probability of guilt were lower than Bayesian posteriors. This finding was taken as evidence that jurors are more conservative than they should be when revising judgments in light of a forensic match or, in other words, that people give less weight to forensic evidence than they should.

Although this finding is consistent with a large body of research showing that people tend to be conservative in Bayesian updating tasks (Griffin & Tversky, 1992; Kraemer & Weber, 2004; Slovic, Fishhoff, & Lichtenstein, 1977; Slovic & Lichtenstein, 1971), there are several important caveats. First, subjects' apparent conservatism in these early studies may have been due, in part, to the inadequacy of the Bayesian models. In these models the likelihood ratio depended on a single variable—the random match probability. Although these likelihood ratios may have

reflected the diagnostic value of the forensic match, they failed to capture any uncertainty about its reliability. In other words, the Bayesian models against which subjects' judgments were compared implicitly assumed the forensic tests were error-free. This is a big assumption and one that subjects probably did not share (Navon, 1978; Schklar & Diamond, 1999). Second, many of the studies may have had scaling problems in their dependent measures. The studies generally asked jurors to estimate the likelihood of the suspect's guilt on a scale of probability (0–1.00) or as a percentage (0–100%). It is possible that the apparent conservatism arose simply from a reluctance to use the endpoints of these scales.

Two more recent studies (Nance & Morris, 2002; Schklar & Diamond, 1999) employed better Bayesian models that incorporated both the probability of a coincidental match and the probability of a false positive due to laboratory error. These studies found that judgments were still, on average, somewhat more conservative than Bayesian norms, thereby bolstering the evidence that conservatism is a valid phenomenon.

However, not all subjects in the early studies made conservative judgments. The pattern of responses suggested that some subjects were responding inappropriately to the forensic evidence due to what some researchers labeled fallacious interpretation of the statistical data (Thompson, 1989; Thompson & Schumann, 1987b) and other researchers called "semantic confusion" (Koehler, 1996). Thompson and Schumann labeled one error the "prosecutor's fallacy." Victims of this fallacy equate the random match probability with the probability the matching items have a different source. If the defendant matches the perpetrator of a crime on a characteristic found in 2% of the population, for example, victims of the fallacy assume there is only a 2% chance the defendant is not the perpetrator and therefore a 98% chance defendant is guilty. This reasoning is erroneous, of course, because it fails to consider the prior probability that the defendant is the perpetrator. Erroneous inference of this type might well lead people to give more weight to evidence of a forensic match than they should, particularly if the prior probability of the suspect's guilt is low (Thompson).

Some jurors also made judgments consistent with a second error that Thompson and Schumann (1987) called the "defense attorney's fallacy." When the suspect and perpetrator matched on a characteristic found in 2% of the population, for example, they apparently reasoned that 2% of the population comprises thousands of people and concluded that there is little or no relevance in the defendant's membership in such a large group. What this reasoning misses, of course, is that the forensic evidence

drastically narrows the range of people who might be guilty without eliminating the defendant.

Nance and Morris (2002) reported that some of their subjects committed a third type of error. They equated the conditional probability that a suspect would match if he was not the source with the probability of the suspect's guilt. When told that the defendant and perpetrator matched on a characteristic found in 4% of the population, for example, they concluded irrationally that this meant there was a 4% chance the defendant was the perpetrator. The Nance and Morris study did not include deliberation, so it is unclear whether this misunderstanding, which was observed in 5% of jurors, would survive exposure to other points of view during deliberation.

Jurors also appear to have difficulty aggregating or combining information about random match probabilities with information about the probability of a false positive. In most cases, the probability of an erroneous match being reported if the suspect is not the source can be estimated with fairly good accuracy by simply adding together the random match probability (RMP) and the false positive probability (FPP). However, research suggests that jurors do not combine the RMP and FPP in an additive manner. In a provocative study, Koehler et al. (1995) found that jurors gave far more weight to DNA evidence when they were told the FPP was .02 and the RMP was 0.000000001 (1 in 1 billion) than when they were told the FPP was .02 and they were given no information about the RMP. This finding is counternormative. The low RMP should have made little difference to the jurors because it was dwarfed by the high FPP[6] and hence the low RMP should not significantly change the value of the DNA evidence. Koehler et al. suggested that jurors might have been unduly influenced by the flashy one-in-a-billion statistic or that they might have averaged rather than summed the RMP and FPP when assigning weight to the DNA evidence. Schklar and Diamond (1999) also found evidence of misaggregation. In their study, jurors gave significantly more weight to DNA evidence when they were told the RMP was 0.000000001 and the FPP was 0.02 (or vice versa) than when told the combined probability that the laboratory would report an incorrect match due to either a random match or a false positive was 0.02.

Studies have examined jurors' reactions to a number of variations in the way statistical evidence is presented and have found that logically inconsequential differences in the format of the evidence can have significant effects. For example, several studies found that jurors give more weight to

[6]The sum of .02 and .00000001 is very close to .02.

forensic evidence when the RMP is presented as a conditional probability (Thompson & Schumann, 1987) or likelihood ratio (Nance & Morris, 2002) than when it is presented as a frequency (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000; Lindsey, Hertwig, & Gigerenzer, 2003).

This finding may arise in part because conditional probabilities and likelihood ratios induce more judgments consistent with fallacious reasoning (Nance & Morris, 2002; Thompson & Schumann, 1987). But it is probably also due to a phenomenon that Koehler labeled exemplar cueing (Koehler, 2001; Koehler & Macchi, 2004). In a clever series of studies, Koehler demonstrated that people give less weight to evidence of a forensic match when the match statistics are presented in a manner that cues or suggests the possibility of a coincidence. For example, Koehler and Maachi found that people gave less weight to matching evidence when the frequencies were presented in a manner that suggested multiple people matching (2 in 1,000) than when the frequencies were presented in a mathematically equivalent manner that did not (0.2 in 100). According to exemplar cueing theory, people assign weight to the evidence of a match according to how readily they can imagine others matching, rather than by any more formal process, and hence can be influenced by variations in presentation format that are logically inconsequential.

Overall, the studies suggest that people are not intuitive Bayesians and that their judgments of forensic evidence may be influenced by variables that would be given no weight in a Bayesian model. Moreover, the research suggests that people sometimes use suboptimal strategies for combining different types of statistical evidence, and for combining forensic evidence with other evidence in a case, that may cause them to undervalue or overvalue forensic evidence.

Latent Prints

As explained previously, latent print evidence is typically given in very strong terms. What impact this has on a jury is not clear. Most commentators would probably agree that fingerprint evidence enjoys a strong presumption of accuracy among jurors. As a Utah Court of Appeals judge put it, "In essence, we have adopted a cultural assumption that a government representative's assertion that a defendant's fingerprint was found at a crime scene is an infallible fact, and not merely the examiner's opinion" (*State v. Quintana,* 2004, Thorne, J., concurring).

Illsley (1987) conducted the most comprehensive jury research on fingerprint evidence so far. Illsley surveyed 1,000 potential jurors who were

serving jury duty at four different courts in Utah (one federal, three state). Not surprisingly, Illsley found that jurors think very highly of fingerprint evidence. Ninety-three percent agreed with the statement "Fingerprint identification is a science," and only 2% disagreed. Eight-five percent agreed with the statement "Fingerprints are the most reliable means of identifying a person," and only 8% disagreed. These responses suggest that jurors bring very favorable preconceptions to the evaluation of latent print evidence. At the same time, Illsley's study suggests that this favorable predisposition does not necessarily translate to "infallibility"[7]; three quarters of respondents agreed with the statement "It is possible for a fingerprint expert to make a mistake when comparing two fingerprints."

LEGAL REFORMS

A major problem with forensic identification science, at present, is the willingness of experts to present conclusions with unwarranted certainty. As discussed, experts frequently make assertions about the accuracy of their methods and results that are unsupported, or inadequately supported, by scientific research. This mismatch between expert testimony and underlying science might be addressed in two ways. First, experts might be encouraged, or required, to moderate the strength of their conclusions so that their testimony stays within scientifically supportable bounds. Second, the scientific foundations of the field might be improved to that point that more confident assertions are justified. In this section we discuss several possible ways to bring about these reforms.

Professional Standards

Self-regulation by forensic scientists has not been particularly successful at addressing the problems discussed here. Although professional societies have promulgated standards for a number of forensic science disciplines, the

[7]Many LPEs claim that the technique is "infallible" (Cole, 2005, p. 987). Laboratory studies of juror perception of latent print evidence are just getting underway (Dahl, Brimacombe, & MacLean, 2005). Reardon, Danielsen, and Meissner (2005) found that fingerprint evidence was the most important form of evidence in simulated cases, outweighing eyewitness and alibi evidence. They found that jurors were sensitive to the number of corresponding ridge characteristics and to the quality of marks. Curiously they found that the expert's declaration of a match had no influence on the jurors' evaluation of the evidence. This would suggest that Judge Pollak's proposed remedy (*United States vs. Llera plaza i,* 2002) of allowing LPEs to attest to similarities but not to matches would not have made a difference.

standards generally focus more on assuring uniformity in procedures and testimony than assuring that testimony is well grounded in science. The standards promulgated by latent print and toolmark examiners, which require experts to express conclusions with absolute certainty, are excellent examples of the disconnect that can arise between standards and science.

The limitations of professional self-regulation may arise, in part, from the institutional and social context in which forensic science is practiced. Thompson (1997a, p. 408) offered an analysis of the context of forensic DNA testing that may also apply to other forensic disciplines:

> The tests are developed by individuals and organizations that have a professional, and in some instances, a financial stake in their rapid acceptance in the courtroom. Those who develop and promote the tests also design and perform the bulk of the research to validate them. The test procedures, validation and casework generally receive outside scrutiny only from scientists who are themselves involved in the adversary process as consultants and expert witnesses for litigants. Thus, forensic … testing is not an independent area of academic science; it is a technical field in which procedures are designed, shaped and performed specifically to play a role in litigation.
>
> Moreover, the field is dominated by experts and laboratories whose primary clients are law enforcement agencies and whose typical role in litigation is to provide evidence supporting criminal prosecutions. In this context, pressures exist to distort science to serve law enforcement goals.… The desire to be helpful to law enforcement leads forensic scientists to design, validate and perform … tests in ways that strike a compromise between scientific rigor and other goals, such as maintaining the analysts' discretion to resolve ambiguities in accordance with other information about a case.

The limitations of self-regulation are also apparent in the frequent failure of forensic scientists to detect and expose fraudulent conduct by their colleagues. Cases in which forensic scientists were proven to have engaged in scientific fraud, such as fabrication of test results, are surprisingly common (Giannelli, 1997; Kelly & Wearne, 1998; see also cases compiled at http://www.corpus-delicti.com/forensic_fraud.html). A striking feature of these fraud cases is how few were exposed by forensic scientists. Most of the cases were exposed only after extraordinary circumstances, such as postconviction DNA exonerations, revealed the innocence of a person convicted by the fraudulent evidence (Scheck, Neufeld, & Dwyer, 2000). If self-regulation is inadequate for policing outright scientific fraud, it is unlikely to be effective for controlling testimony that, though not intentionally dishonest, is exaggerated and misleading.

In recent years there has been a trend toward accreditation of forensic laboratories. A nonprofit organization called the American Society of

Crime Laboratory Directors Laboratory Accreditation Board (ASCLD-LAB) is the leading accrediting body. It sends outside experts to review the protocols and procedures of laboratories seeking accreditation. Accreditation is undoubtedly helpful in assuring that laboratories meet minimal standards for training, equipment, documentation, and reporting. However, the expert panels that perform the reviews consist almost exclusively of forensic scientists from other, similar laboratories. Thus, for example, LPEs from Laboratory A evaluate the latent print procedures of Laboratory B, and vice versa. Consequently, the accreditation process may be more helpful for assuring consistency with established practices in the field than for evaluating the validity and appropriateness of those practices. A better approach might be to create independent review panels that include academic scientists as well as forensic practitioners.

Legislation

A variety of bills have been proposed, and some passed into law, for regulating forensic science at the state and federal levels. New York established a state forensic science commission with the power to oversee the operation of state crime laboratories. Following a major scandal involving the Houston Police Crime Lab (Bromwich, 2005), Texas recently passed similar legislation (McVicker, 2005). Several states require state-operated laboratories to be accredited by ASCLD-LAB. Congress has required DNA laboratories to meet certain standards to be eligible for federal funding. Although this legislation is helpful, it has not yet touched the issues we focus on here—that is, the mismatch in many forensic disciplines between courtroom assertions and scientific foundation. These issues may be too technical, too specific, and too dependent on the evolving state of science to be appropriate subjects for legislation.

Admissibility Standards

Perhaps the most likely pathway to reform is more active involvement of trial court judges in policing forensic testimony. The power of judges, in their role as gatekeepers, to exclude invalid testimony is widely acknowledged and was made explicit, for federal courts, by the U.S. Supreme Court's ruling in *Daubert v. Merrell Dow Pharmaceuticals* (1993). Before *Daubert*, most courts applied some version of the general acceptance test (also called the *Frye* [1923] standard) under which scientific evidence was admissible if it was "generally accepted" in the relevant scientific community. Under the *Frye* standard, it was difficult to exclude problematic testimony,

such as the certainty claims of latent print and toolmark examiners, because those claims were (and are) "generally accepted" by forensic practitioners (notwithstanding the absence of scientific proof).[8]

Many scholars expected that forensic science would face more exacting judicial scrutiny under the *Daubert* (1993) standard (Faigman, Saks, & Porter, 1994; Jonakait, 1994; Saks, 2000). After *Daubert,* challenges were indeed raised to several types of forensic testimony that had long been established as admissible evidence (Fradella, O'Neill, & Fogarty, 2004), including handwriting analysis (Hartfield, 2002) and latent print identifications (Cole, 2004). However, the judicial gatekeepers have, to date, been surprisingly lenient in what they will let pass—admitting forensic testimony even when its scientific foundation is clearly flimsy (Cole; Risinger, 2000). This judicial leniency may arise in part from ignorance of science. Based on a survey of 400 state court judges, one research group concluded that "many of the judges surveyed lacked the scientific literacy seemingly necessitated by *Daubert*" (Gatowski et al., 2001, p. 433). Another study found that judges performed poorly when evaluating the merits of scientific experiments, often failing to appreciate serious threats to validity (Kovera & McAuliff, 2000). Improving judicial education in scientific methods, and particularly the requirements of scientific validation, would clearly be helpful.

But ignorance of science is not the only problem. The judicial leniency in admitting forensic identification evidence may arise in part from sympathy for prosecutors, who are the major proponents of forensic evidence (Risinger, 2000). It may also arise in part because courts themselves, rather than the scientific community or scientific institutions, have become the principal sources of legitimation for many forensic identification techniques (Cole, 2004). Another factor may be judges' concerns that excluding such testimony will deny to law enforcement the vital benefits of whole categories of forensic evidence (Fradella et al., 2004).

To the extent such concerns exist, we believe they are misplaced. In the long run, stricter standards for admissibility will do more to strengthen than to harm forensic science. The experience with DNA evidence is

[8]When applying the *Frye* (1923) standard courts occasionally looked beyond practitioners of a particular discipline to the broader scientific community. Hence, astrology is inadmissible even though it might be generally accepted by practicing astrologers. But courts have rarely looked beyond forensic scientists when evaluating the admissibility of forensic identification evidence. The exceptions (*People v. Kelly,* 1976; *Williamson v. Reynolds,* 1995) involve forms of evidence (voice spectrography or microscopic hair comparison) that lack a strong presumption of belief.

instructive. The rulings in the early 1990s that excluded DNA evidence because the underlying statistics were poorly validated did not spell the doom of DNA evidence. To the contrary, these rulings had important positive effects: "It was the prospect of negative admissibility rulings that spurred much-needed research on the problem of population structure, research that otherwise might not have been done" (Thompson, 1997a, p. 423). We believe similar positive benefits would arise from rulings excluding other poorly validated forms of forensic identification evidence. Although such rulings might hinder prosecution in a few cases, they would force forensic practitioners to either moderate their testimony or strengthen their scientific validation so that their testimony rests on a stronger scientific footing. Because courts are the primary users of forensic identification evidence, they retain the greatest amount of leverage to effect reform. We believe stricter judicial standards for the admissibility of forensic identification evidence are the surest and fastest pathway to reform of the field of forensic identification science.

ACKNOWLEDGMENT

This project was funded in part by the National Science Foundation (Award #SES-0347305). The views expressed, and responsibility for any errors, are the authors'.

REFERENCES

American Board of Forensic Odontology. (1999). *Bitemark methodology guidelines*. Retrieved from http://www.abfo.org/bitemark.htm

Ashbaugh, D. R. (1999). *Quantitative-qualitative friction ridge analysis: An introduction to basic and advanced ridgeology*. Boca Raton, FL: CRC Press.

Association of Firearms and Toolmark Examiners. (1998). Theory of identification as it relates to Toolmarks. *AFTE Journal, 30*(1), 86.

Babler, W. J. (1975). Early prenatal attainment of adult metacarpal-phalangeal rankings and proportions. *American Journal of Physical Anthropology, 43,* 327–332.

Babler, W. J. (1978). Prenatal selection and dermatoglyphic patterns. *American Journal of Physical Anthropology, 48,* 21–28.

Babler, W. J. (1983, Spring). How is epidermal ridge configuration determined. *Newsletter of the American Dermatoglyphics Association, 2,* 3–4.

Babler, W. J. (1987). Prenatal development of dermatoglyphic patterns: Associations with epidermal ridge, volar pad and bone morphology. *Collegium Antropologicum, 11,* 297–304.

Babler, W. J. (1990). Prenatal communalities in epidermal ridge development. In N. M. Durham & C. C. Plato (Eds.), *Trends in dermatoglyphic research* (pp. 54–68). Dordrecht, The Netherlands: Kluwer.

Babler, W. J. (1991). Embryological development of epidermal ridges and their configurations. *Birth Defects Original Article Series, 27*(2), 95–112.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9,* 396–406.

Benedict, N. (2004). Fingerprints and the Daubert standard for admission of scientific evidence: Why fingerprints fail and a proposed remedy. *Arizona Law Review, 46,* 519–549.

Bisbing, R. E., Deadman, H., Houck, M. M., Palenick, S., & Rowe, W. F. *Peer review report: State v. Bromgard.* (2002).

Bonnevie, K. (1924). Studies on papillary patterns of human fingers. *Journal of Genetics, 15,* 1–112.

Bowers, C. M. (2002). Identification from bitemarks: Scientific issues. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Science in the law: Forensic science issues* (pp. 244–293). St. Paul, MN: West Group.

Boychuk, J., & Vokey, J. R. (2004, April). *On the psychophysics of fingerprint identification.* Paper presented at the Banff Annual Seminar in Cognitive Science, Banff, Alberta, Canada.

Bromwich, M. (2005). *Third report of the independent investigator for the Houston Police Department crime laboratory and property room.* Houston, TX: Fried, Frank, Harris, Shriver, & Jacobson, LLP.

Budowle, B., Chakraborty, R., Carmody, G., & Monson, K. L. (2000). Source attribution of a forensic DNA profile [Electronic version]. *Forensic Science Communications, 2*(3).

Bunch, S. G. (2000). Consecutive matching criteria: A general critique. *Journal of Forensic Sciences, 45,* 955–962.

Butler, J. M. (2005). *Forensic DNA typing: Biology and technology behind str markers* (2nd ed.). San Diego, CA: Academic Press.

Champod, C. (1995). Edmond locard: Numerical standards and "probable" identifications. *Journal of Forensic Identification, 45,* 136–163.

Champod, C., Egli, N., & Margot, P. A. (2004, October). *Fingermarks, shoesole and footprint impressions, tire impressions, ear impressions, toolmarks, lipmarks, bitemarks: A review.* Paper presented at the 14th Interpol Forensic Science Symposium, Lyon, France.

Champod, C., & Evett, I. W. (2001). A probabilistic approach to fingerprint evidence. *Journal of Forensic Identification, 51,* 101–122.

Cole, S. (1999). What counts for identity? The historical origins of the methodology of latent fingerprint identification. *Science in Context, 12,* 139–172.

Cole, S. A. (1998). Witnessing identification: Latent fingerprint evidence and expert knowledge. *Social Studies of Science, 28,* 687–712.

Cole, S. A. (2004). Grandfathering evidence: Fingerprint admissibility ruling from *Jennings* to *Llera Plaza* and back again. *American Criminal Law Review, 41,* 1189–1276.

Cole, S. A. (2005). More than zero: Accounting for error in latent fingerprint identification. *Journal of Criminal Law and Criminology, 95,* 985–1078.

Cole, S. A. (2006). Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponents' discourse. *Law and Policy, 28,* 109–135.

Cummins, H., & Midlo, C. (1943). *Finger prints, palms and soles: An introduction to dermatoglyphics.* Philadelphia: Blakiston.

Dahl, L. C., Brimacombe, C. A. E., & MacLean, C. (2005, March). *Fingerprint versus eyewitness evidence: Equally important in the forensic drama?* Paper presented at the American Psychology Law Society, La Jolla, CA.

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to make erroneous identifications. *Forensic Science International, 156,* 74–78.

Dror, I. E., Péron, A. E., Hind, S.-L., & Charlton, D. (2005). When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology, 19,* 799–809.

Epstein, R. (2002). Fingerprints meet *Daubert:* The myth of fingerprint "science" is revealed. *Southern California Law Review, 75,* 605–657.

European Fingerprint Standards. (2002). *Fingerprint Whorld, 28*(107), 19.

Evett, I. W., & Weir, B. S. (1998). *Interpreting DNA evidence: Statistical genetics for forensic scientists.* Sunderland, MA: Sinauer.

Evett, I. W., & Williams, R. L. (1996). A review of the sixteen points fingerprint standard in England and Wales. *Journal of Forensic Identification, 46,* 49–73.

Faigman, D., & Baglioni, A. (1988). Bayes' theorem in the trial process. *Law and Human Behavior, 12,* 1–17.

Faigman, D., Saks, M. J., & Porter, E. (1994). Check your crystal ball at the courthouse door, please: Exploring the present, worrying about the future of scientific evidence. *Cardozo Law Review, 15,* 1799.

Federal Bureau of Investigation. (1998, February). DNA profiling advancement. *FBI Law Enforcement Bulletin, 67*(2), 24.

Fradella, H. F., O'Neill, L., & Fogarty, A. (2004). The impact of Daubert on forensic science. *Pepperdine Law Review, 31,* 323–361.

Frye v. United States, 293 F. 1013 (D.C. Ct. App. 1923).

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a
post-*Daubert* World. *Law and Human Behavior, 25,* 433–458.

Giannelli, P. (1997). The abuse of scientific evidence in criminal cases: The need for independent crime laboratories. *Virginia Journal of Social Policy and the Law, 4,* 439–478.

Goodman, J. (1992). Jurors' comprehension and assessment of probabilistic evidence. *American Journal of Trial Advocacy, 16,* 361–389.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Haber, L., & Haber, R. N. (2003). Error rates for human fingerprint examiners. In N. K. Ratha & R. Bolle (Eds.), *Automatic fingerprint recognition systems* (pp. 339–360). New York: Springer.

Hartfield, L. C. (2002, November). Daubert/Kumho challenges to handwriting analysis. *The Champion, 26,* 24–36.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science, 290,* 2261–2262.

Houck, M. M. (2004). Forensic science, no consensus. *Issues in Science and Technology, 20*(2), 6–8.
Houck, M. M., & Budowle, B. (2002). Correlation of microscopic and mitochondrial DNA hair comparisons. *Journal of Forensic Sciences, 47*(5), 1–4.
Illsley, C. (1987, July). *Juries, fingerprints, and the expert fingerprint witness*. Paper presented at the International Symposium on Latent Prints, Quantico, VA.
Imwinkelried, E. J., & Kaye, D. H. (2001). DNA typing: Emerging or neglected issues. *Washington Law Review, 76*, 413–474.
Imwinkelried, E. J., & Tobin, W. A. (2003). Comparative bullet lead analysis (CBLA) evidence: Valid inference or ipse dixit? *Oklahoma City University Law Review, 28*, 43–72.
Inman, K., & Rudin, N. (2001). *Principles and practice of criminalistics: The profession of forensic science*. Boca Raton, FL: CRC Press.
International Association for Identification. (1979). Resolution vii. *Identification News, 29*, 1.
Interpol European Expert Group on Fingerprint Identification II. (2004). *Method for fingerprint identification*. Retrieved from http://www.interpol.int/Public/Forensic/fingerprints/Working Parties/IEEGFI2/default.arp
Jonakait, R. N. (1994). The meaning of Daubert and what that means for forensic science. *Cardozo Law Review, 15*, 2103–2116.
Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society, Series A, 154*, 75–81.
Kaye, D. H., & Sensabaugh, G. F., Jr. (2000). Reference guide on DNA evidence. In J. Cecil (Ed.), *Reference manual on scientific evidence* (Vol. 2, pp. 485–576). Washington, DC: Federal Judicial Center.
Kelly, J. F., & Wearne, P. K. (1998). *Tainting evidence: Inside the scandals at the FBI crime lab*. New York: Free Press.
Kennedy, D. (2003). Forensic science: Oxymoron? *Science, 302*, 1625.
Koehler, J. J. (1995). The random match probability in DNA evidence: Irrelevant and prejudicial? *Jurimetrics, 35*, 201–219.
Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review, 67*, 859–886.
Koehler, J. J. (1997). Why DNA likelihood ratios should account for error (even when a National
Research Council report says they should not). *Jurimetrics, 37*, 425–437.
Koehler, J. J. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior, 25*, 493–513.
Koehler, J. J., Chia, A., & Lindsey, J. S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? *Jurimetrics, 35*, 201–219.
Koehler, J. J., & Macchi, L. (2004). Thinking about low-probability events. *Psychological Science, 15*, 540–546.
Kovera, M. B., & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? *Journal of Applied Psychology, 85*, 574–586.
Kraemer, C., & Weber, M. (2004). How do people take into account weight, strength and quality of segregated vs. aggregated data? Experimental evidence. *Journal of Risk and Uncertainty, 29*(2), 113–142.

La Morte, T. M. (2003). Sleeping gatekeepers: *United States v. Llera Plaza* and the unreliability of forensic fingerprinting evidence under *Daubert*. *Albany Law Journal of Science and Technology, 14,* 171–214.

Langenburg, G. M. (2004). Pilot study: A statistical analysis of the ace-v methodology—analysis stage. *Journal of Forensic Identification, 54*(1), 64–79.

Leo, W. F. (1998). Distortion versus dissimilarity in friction skin identification. *Journal of Forensic Identification, 48*(2), 125–129.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating statistical DNA evidence. *Jurimetrics, 43,* 147–163.

McVicker, S. (2005, June 20). Commission created to look into crime labs. *The Houston Chronicle, Section B,* p. 5.

Michigan v. Ballard, 2003 Mich. App. LEXIS 547 (Mich. Ct. App. 2003).

Mnookin, J. L. (2001). Fingerprint evidence in an age of DNA profiling. *Brooklyn Law Review, 67*(1), 13–70.

Moenssens, A. (2003). Fingerprint identification: A valid reliable "forensic science?" *Criminal Justice, 18*(2), 31–37.

Nance, D. A., & Morris, S. B. (2002). An empirical assessment of presentation formats for trace evidence with a relatively large and quantifiable random match probability. *Jurimetrics Journal, 42,* 403–448.

National Research Council. (1992). *DNA technology in forensic science*. Washington, DC: National Academy Press.

National Research Council. (1996). *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press.

National Research Council. (2004). *Forensic analysis: Weighing bullet lead evidence*. Washington, DC: National Academy Press.

Navon, D. (1978). Statistical and metastatistical considerations in analyzing the desirability of human Bayesian conservatism. *British Journal of Mathematical and Statistical Psychology, 304,* 205–212.

New Hampshire State Police Forensic Laboratory. (2005). *Latent print examination: Interpretation and reporting of conclusions* (3rd ed., Vol. IDU-001-03). Author.

Nichols, R. G. (2003). Firearm and toolmark identification criteria: A review of the literature, Part II. *Journal of Forensic Sciences, 48,* 318–327.

Nichols, R. (2006). The scientific foundations of firearms and tool mark identification: A response to recent challenges. *California Association of Criminalists News,* 8–27.

Pankanti, S., Prabhakar, S., & Jain, A. K. (2002). On the individuality of fingerprints. *IEEE Transactions on PAMI, 24,* 1010–1025.

People v. Kelly, 549 P.2d 1240 (Cal. 1976).

People v. Wesley, 553 N.Y.S.2d 643 (N.Y. Sup. Ct. 1988).

Peterson, J. L., & Markham, P. N. (1995a). Crime laboratory proficiency testing results, 1978–1991, I: Identification and classification of physical evidence. *Journal of Forensic Sciences, 40,* 994–1009.

Peterson, J. L., & Markham, P. N. (1995b). Crime laboratory proficiency testing results, 1978–1991, II: Resolving questions of common origin. *Journal of Forensic Sciences, 40,* 1009–1029.

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.

Reardon, M., Danielsen, E., & Meissner, C. A. (2005, March). *Investigating juror perceptions of fingerprint evidence in criminal cases*. Paper presented at the American Psychology-Law Society, La Jolla, CA.

Risinger, D. M. (2000). Navigating expert reliability: Are criminal standards of certainty being left on the dock? *Albany Law Review, 64,* 99–152.

Risinger, D. M. (2002). Handwriting identification. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Science in the law: Forensic science issues* (pp. 113–193). St. Paul, MN.: West Group.

Risinger, D. M., & Saks, M. J. (2003). A house with no foundation. *Issues in Science and Technology, 20,* 35–39.

Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The *Daubert/Kumho* implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review, 90,* 1–56.

Saks, M. J. (1998). Merlin and Solomon: Lessons from the law's formative encounters with forensic identification science. *Hastings Law Journal, 49,* 1069–1141.

Saks, M. J. (2000). Banishing *ipse dixit*: The impact of *Kumho Tire* on forensic identification science. *Washington and Lee Law Review, 57,* 879–900.

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309,* 892–895.

Scheck, B., Neufeld, P., & Dwyer, F. (2000). *Actual innocence*. New York: Doubleday.

Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior, 23*(2), 159–184.

Schum, D. A. (1994). *Evidential foundations of probabilistic reasoning*. New York: Wiley.

Schum, D. A., & DuCharme, W. M. (1971). Comments on the relationship between the impact and the reliability of evidence. *Organizational Behavior and Human Performance, 6,* 111–131.

Schwartz, A. (2004). A challenge to the admissibility of firearms and toolmark identifications: Amicus brief prepared on behalf of the defendant in *United States v. Kain*, Crim. No. 03-573-1 (E.D. Pa.). *Journal of Philosophy, Science and Law, 4,* 1–32.

Schwartz, A. (2005). A systemic challenge to the reliability and admissibility of firearms and toolmark identification. *Columbia Science and Technology Law Review, 6,* 1–42.

Scientific Working Group on Friction Ridge Analysis Study and Technology. (2002). *Friction ridge examination methodology for latent print examiners* (ver. 1.01 ed.).

Scientific Working Group on Friction Ridge Analysis Study and Technology. (2003a). *Glossary—consolidated* (ver. 1.0 ed.). Retrieved from http://www.swgfast.org/Glossary-consolidated_ver_1.pdf

Scientific Working Group on Friction Ridge Analysis Study and Technology. (2003b). Standards for conclusions (ver. 1.0 ed.). *Journal of Forensic Identification, 54,* 358–359.

Slovic, P., Fishhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior, 20,* 49–82.

Sombat, J. M. (2002). Note, latent justice: Daubert's impact on the evaluation of fingerprint identification testimony. *Fordham Law Review, 70,* 2819.

Springer, E. (1995). Toolmark examinations: A review of its development in the literature. *Journal of Forensic Sciences, 40,* 964–968.

Stacey, R. B. (2004). A report on the erroneous fingerprint individualization in the Madrid train bombing case. *Journal of Forensic Identification, 54,* 706–718.

Stafford Smith, C. A., & Goodman, P. D. (1996). Forensic hair comparison analysis: Nineteenth century science or twentieth century snake oil? *Columbia Human Rights Law Review, 27,* 227–291.

State v. Quintana, 103 P.3d 168 (Utah Ct. App. 2004).

Stoney, D. A. (1997). Fingerprint identification: Scientific status. In D. L. Faigman, D. H. Kaye, M. J. Saks, & J. Sanders (Eds.), *Modern scientific evidence* 2 (pp. 55–78). St. Paul, MN: West Group.

Stoney, D. A. (2001). Measurement of fingerprint individuality. In H. C. Lee & R. E. Gaensslen (Eds.), *Advances in fingerprint technology* (pp. 327–387). Boca Raton, FL: CRC Press.

Tangen, J. M., Vokey, J. R., & Allan, L. G. (2000, July). *What's in a fingerprint? A PCA approach to fingerprint identification and categorisation.* Paper presented at the Joint Meeting of the Experimental Psychology Society and the Canadian Society for Brain, Behaviour, and Cognitive Science, Cambridge, England.

Thompson, W. C. (1989). Are juries competent to evaluate statistical evidence? *Law and Contemporary Problems, 52,* 9–41.

Thompson, W. C. (1993). Evaluating the admissibility of new genetic identification tests: Lessons from the "DNA war." *Journal of Criminal Law and Criminology, 84,* 22–104.

Thompson, W. C. (1995). Subjective interpretation, laboratory error and the value of forensic DNA evidence: Three case studies. *Genetica, 96,* 153–168.

Thompson, W. C. (1996). DNA evidence in the O. J. Simpson trial. *University of Colorado Law Review, 67,* 827–857.

Thompson, W. C. (1997a). Accepting lower standards: The National Research Council's

second report on forensic DNA evidence. *Jurimetrics, 37,* 405–424.

Thompson, W. C. (1997b). Forensic DNA evidence. In B. Black & P. Lee (Eds.), *Expert evidence: A practitioner's guide to law, science and the FJC manual* (pp. 195–266). St. Paul, MN: West Group.

Thompson, W. C., & Cole, S. A. (2005). Lessons from the Brandon Mayfield case. *The Champion, 29,* 42–44.

Thompson, W. C., Ford, S., Doom, T., Raymer, M., & Krane, D. (2003). Evaluating forensic DNA evidence: Elements of a competent defense review. *The Champion, 27,* 16–25.

Thompson, W. C., & Krane, D. E. (2003). DNA in the courtroom. In J. Moriarty (Ed.), *Psychological and scientific evidence in criminal trials* (pp. 11-1–11-75). St. Paul, MN: West Group.

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior, 11,* 167–187.

Thompson, W. C., Taroni, F., & Aitkin, C. G. G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences, 48,* 1–8.

Thornton, J. (1977). The one-dissimilarity doctrine in fingerprint identification. *International Criminal Police Review, 32,* 89–95.

Torry, M. E., & Vokey, J. R. (2005, May). *Fingerprint matching and naive observers*. Paper presented at the Banff Annual Seminar in Cognitive Science, Banff, Alberta, Canada.

United States v. Llera Plaza, 179 F. Supp. 2d 492 (E.D. Pa. 2002), *vacated,* 188 F. Supp. 2d 549.

Vokey, J. R., Tangen, J. M., & Boychuk, J. (2004, June). *On the identification of latent fingerprints.* Paper presented at the Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science, St. John's, Newfoundland and Labrador, Canada.

Wertheim, K., & Maceo, A. (2002). The critical stage of friction ridge pattern formation. *Journal of Forensic Identification, 52,* 35–85.

Whipple, I. L. (1904). The ventral surface of mammalian chiridium with special reference to the conditions found in man. *Zeitschrift füür Morphologie und Anthropologie, 7,* 261–368.

Williamson v. Reynolds, 904 F. Supp. 1529 (E.D. Okla. 1995).

Yardley, J. (2001, May 8). Flaws in chemist's findings free man at center of inquiry. *The New York Times,* p. A1.

Zabell, S. L. (2005). Fingerprint evidence. *Journal of Law and Policy, 13,* 143–170.