A fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process. As noted above, AFTE has adopted a theory of identification, but it does not provide a specific protocol. It says that an examiner may offer an opinion that a specific tool or firearm was the source of a specific set of toolmarks or a bullet striation pattern when "sufficient agreement" exists in the pattern of two sets of marks. It defines agreement as significant "when it exceeds the best agreement demonstrated between tool marks known to have been produced by different tools and is consistent with the agreement demonstrated by tool marks known to have been produced by the same tool." The meaning of "exceeds the best agreement" and "consistent with" are not specified, and the examiner is expected to draw on his or her own experience. This AFTE document, which is the best guidance available for the field of toolmark identification, does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.

Although some studies have been performed on the degree of similarity that can be found between marks made by different tools and the variability in marks made by an individual tool, the scientific knowledge base for toolmark and firearms analysis is fairly limited. For example, a report from Hamby, Brundage, and Thorpe\(^6\) includes capsule summaries of 68 toolmark and firearms studies. But the capsule summaries suggest a heavy reliance on the subjective findings of examiners rather than on the rigorous quantification and analysis of sources of variability. Overall, the process for toolmark and firearms comparisons lacks the specificity of the protocols for, say, 13 STR DNA analysis. This is not to say that toolmark analysis needs to be as objective as DNA analysis in order to provide value. And, as was the case for friction ridge analysis and in contrast to the case for DNA analysis, the specific features to be examined and compared between toolmarks cannot be stipulated a priori. But the protocols for DNA analysis do represent a precisely specified, and scientifically justified, series of steps that lead to results with well-characterized confidence limits, and that is the goal for all the methods of forensic science.

ANALYSIS OF HAIR EVIDENCE

The basis for hair analyses as forensic evidence stems from the fact that human and animal hairs routinely are shed and thus are capable of being

transferred from an individual to the crime scene, and from the crime scene to an individual. Forensic hair examiners generally recognize that various physical characteristics of hairs can be identified and are sufficiently different among individuals that they can be useful in including, or excluding, certain persons from the pool of possible sources of the hair. The results of analyses from hair comparisons typically are accepted as class associations; that is, a conclusion of a “match” means only that the hair could have come from any person whose hair exhibited—within some levels of measurement uncertainties—the same microscopic characteristics, but it cannot uniquely identify one person. However, this information might be sufficiently useful to “narrow the pool” by excluding certain persons as sources of the hair.

Although animal hairs might provide useful evidence in certain cases (e.g., animal poaching), animal hair analysis often can lead to an identification of only the type of animal, not the specific breed\textsuperscript{66}; consequently, most (90 to 95 percent) of hair analyses refer to analyses of human hair. Human hairs from different parts of the body have different characteristics; Houck cautions strongly against drawing conclusions about hairs from one part of the body based on analyses of hairs from a different body part.\textsuperscript{67}

Houck and Bisbing recommend as minimal training for hair examiners a bachelor’s degree in a natural or applied science (e.g., chemistry, biology, forensic science), on-the-job training programs, and an annual proficiency test.\textsuperscript{68}

Sample Data and Collection

Sample hairs received for analysis initially are examined macroscopically for certain broad features such as color, shaft form (e.g., straight, wavy, curved, kinked), length, and overall shaft thickness (e.g., fine, medium, coarse).

In the second stage of analysis, hairs are mounted on microscopic slides using a mounting medium that has the same refractive index (about 1.54) as the hair, to better view the microscopic features (see next section). One hair or multiple hairs from the same source may be mounted on a glass microscope slide with an appropriate cover slip, as long as each mounted hair is clearly visible. It is most important that questioned and known hairs are mounted in the same type of mounting medium.

During this examination, the hair analyst attempts to identify the part of the body from which the hair might have come, based on certain de-
finable characteristics that distinguish hairs from various body locations. Occasionally, suspects can be eliminated on the basis of these simple microscopic characteristics.

A “control” or “comparison” group of hairs must be collected from a known hair source. A known head hair sample should consist of hairs from the five different areas of the scalp (top, front, back including nape, and both sides). Known hair samples should be obtained by a combination of pulling and combing from the sampled region. Ideally, a total of 50 hairs should be obtained from the scalp. A known pubic hair sample or a sample from any other somatic region should ideally consist of 25 hairs obtained by pulling and combing from different regions. A comparison can still be performed with less than the recommended number of hairs, but this may increase the likelihood of a false exclusion.69

Features from human hair analyses can be divided broadly into “major characteristics” and “secondary characteristics.” The former category includes features such as color, treatment (e.g., dyed, bleached, curled, permed), pigment aggregation (e.g., streaked, clumped, patchy), and shaft form (e.g., wavy, straight, curly). Other major characteristics may include pigment distribution (e.g., uniform, peripheral, clustered), medulla appearance, if present (e.g., continuous, interrupted, or fragmented—and opaque or translucent), hair diameter, medullary index, and presence or absence of cortical fusi (e.g., root or shaft). Secondary characteristics include cuticular margin (e.g., smooth, serrated, looped, or cracked), pigment density (e.g., absent, sparse, heavy), pigment size (e.g., absent, fine, coarse), tip shape (e.g., tapered, cut, rounded, frayed, split), and shaft diameter (e.g., narrow or wide).70

Studies of Accuracy in Identification

In 1974, investigators Gaudette and Keeping described a system of hair analysis and used it in a study of pairwise comparisons among 861 hairs from 100 different persons.71 They acknowledged that “the hair samples were not chosen from the population at random, but were selected so that the probability of two hairs being similar would be greater, if anything, than in the population at large.”72 From their assignment of probabilities, the authors estimated that the chance of asserting a difference between two

---

70 Ibid.
72 Ibid., p. 65.
hairs from the same person is small, about 1 in 4,500. This assignment of probabilities has since been shown to be unreliable. Moreover, the study does not confirm the chance of asserung a match between two dissimilar hairs, and the authors acknowledge that, "due to the fact that so many of the characteristics coded are subjective—for example, color, texture—it was not possible to get complete reproducibility between two or more examiners coding the same hair."

Barnett and Ogle raised four concerns with the Gaudette and Keeping study: (1) it relied on idealized (not from real life) test scenarios; (2) there was no objective basis for selecting the features; (3) the statistical analysis of data from the study was questionable; and (4) there was a possible examiner bias. Gaudette attempted to address these concerns through a further study. However, this additional study involved only three hair examiners, in addition to the author. The author concluded that:

... whereas hair is not generally a basis for positive personal identification, the presence of abnormalities or unusual features or the presence of a large number of different unknown hairs all similar to the standard can lead to a more positive conclusion. The problem, at present, lies in finding suitable additional characteristics [of hair, for effecting individualization]. Although there is basic agreement as to the value of the macroscopic and microscopic characteristics used, other characteristics are either unreliable or controversial. Physical characteristics such as refractive index, density, scale counts, tensile strength, and electrical properties have been proposed by some workers but have been attacked by others, and the general consensus is that they are of little use in hair comparison.

In 1990, Wickenheiser and Hepworth attempted a study to address examiner bias in a small study with only two examiners. They reported that "no incorrect associations were made by either examiner." But a study with only two examiners cannot offer accurate and precise estimates of bias in the population of examiners.

An attempt at an objective system for identifying "matches" among hair samples is presented in Verma et al., based on a neural network.

---

75 Barnett and Ogle, op. cit.
76 Barnett and Ogle, op. cit.
78 Wickenheiser and Hepworth, op. cit., p. 1327.
According to the authors of this article, "The system accurately judged whether two populations of hairs came from the same person or from different persons 83 percent of the time." The article states that 83 percent was obtained by testing the neural network on all possible pairs among 9 samples of hairs from 9 people (i.e., 81 combinations, of which 9 are "true matches" and 72 are "true mismatches"). Their Table 3 can be summarized as follows:

<table>
<thead>
<tr>
<th></th>
<th>System said &quot;same&quot;</th>
<th>System said &quot;different&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same person</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Different persons</td>
<td>9</td>
<td>64</td>
</tr>
</tbody>
</table>

Because the total of these 4 numbers is 82, not 81, one presumes a typographical error in the table; as stated, the number of correct calls is (5 + 64)/81 = 0.83, or 85 percent. (If one of the counts, 5 or 64, is off by 1, the percentage would be 84 percent.) However, the table also shows that the neural network claimed 9 of the 73 different pairs as "same," for a false positive rate of 9/73 = 12 percent, and 4 sets of hairs from the same person as "different," for a false negative rate of 4/9 = 44 percent. With such high error rates, one would want to study improvements to such systems before putting them into routine practice.

Houck et al. indicate that proficiency testing is conducted regularly for hair experts in crime laboratories. Collaborative Testing Services offers hair and fiber proficiency tests annually. Unfortunately, mass production of test samples such as hair is problematic. Because known samples exhibit a range of characteristics within each of the major and secondary characteristics, it is not possible to provide comparable samples to multiple examiners.

Scientific Interpretation and Reporting of Results

The success of hair analyses to make a positive identification is limited in important ways. Most hair examiners would opine only that hairs exhibiting the same microscopic characteristics "could" have come from a
particular individual. Moreover, the "best" or most reliable characteristics will vary by case. For example, "color" may be a critical determinant in a case where it is artificial, because that introduces additional independent variables, such as the time since treatment and the actual hair color, while a natural hair might provide less information.

However, several members of the committee have experienced courtroom cases in which, despite the lack of a statistical foundation, microscopic hair examiners have made probabilistic claims based on their experience, as occurred in some DNA exoneration cases in which microscopic hair analysis evidence had been introduced during trial. Aitken and Robertson discuss some probabilistic concepts with respect to hair analysis.

The availability of DNA analysis has lessened the reliance on hair examination. In a very high proportion of cases involving hair evidence, DNA can be extracted, even years after the crime has been committed. Although the DNA extraction may consist of only mitochondrial DNA (mtDNA), such analyses are likely to be much more specific than those conducted on the physical features of hair. For this reason, cases that might have relied heavily on hair examinations have been subjected more recently to additional analyses using DNA. Because of the inherent limitations of hair comparisons and the availability of higher-quality and higher-accuracy analyses based on mtDNA, traditional hair examinations may be presented less often as evidence in the future, although microscopic comparison of physical features will continue to be useful for determining which hairs are sufficiently similar to merit comparisons with DNA analysis and for excluding suspects and assisting in criminal investigations.

Summary Assessment

No scientifically accepted statistics exist about the frequency with which particular characteristics of hair are distributed in the population. There appear to be no uniform standards on the number of features on which hairs must agree before an examiner may declare a "match." In one study of validity and accuracy of the technique, the authors required exact agreement on seven "major" characteristics and at least two agreements among six "secondary" characteristics. The categorization of hair features depends heavily on examiner proficiency and practical experience.

An FBI study found that, of 80 hair comparisons that were "associ-

---

ated through microscopic examinations, 9 of them (12.5 percent) were found in fact to come from different sources when reexamined through mtDNA analysis. This illustrates not only the imprecision of microscopic hair analyses, but also the problem with using imprecise reporting terminology such as "associated with," which is not clearly defined and which can be misunderstood to imply individualization.

In some recent cases, courts have explicitly stated that microscopic hair analysis is a technique generally accepted in the scientific community. But courts also have recognized that testimony linking microscopic hair analysis with particular defendants is highly unreliable. In cases where there seems to be a morphological match (based on microscopic examination), it must be confirmed using mtDNA analysis; microscopic studies alone are of limited probative value. The committee found no scientific support for the use of hair comparisons for individualization in the absence of nuclear DNA. Microscopy and mtDNA analysis can be used in tandem and may add to one another's value for classifying a common source, but no studies have been performed specifically to quantify the reliability of their joint use.

ANALYSIS OF FIBER EVIDENCE

Fibers associated with a crime—including synthetic fibers such as nylon, polyester and acrylic as well as botanical fibers such as ramie or jute, which are common in ropes or twines—can be examined microscopically in the same way as hairs, and with the same limitations. However, fibers also can be analyzed using the tools of analytical chemistry, which provide a more solid scientific footing than that underlying morphological examination. In some cases, clothing and carpets have been subjected to relatively distinctive environmental conditions (e.g., sunlight exposure or laundering agents) that impart characteristics that can distinguish particular items from others from the same manufacturing lot. Fiber examiners agree, however, that none of these characteristics is suitable for individualizing fibers (associating a fiber from a crime scene with one, and only one, source) and that fiber evidence can be used only to associate a given fiber with a class of fibers.

87 Houck and Budowie, op. cit.
90 See, e.g., R.R. Breese. 1987. Evaluation of textile fiber evidence: A review. Journal of Forensic Sciences 32(2):510-521 See also SWGMAT. 1999. Introduction to forensic fiber examination. Forensic Science Communications 1(1). Available at www.fbi.gov/hq/lab/forencis/fsc_communications/april1999/houcktoc.htm, which includes the following summation in Section 5.4: "It can never be stated with certainty that a fiber originated from a particular textile because..."