J Forensic Sci, July. 2003, Vol. 48, No. 4
Paper ID JFS2003022_484
Published 19 May 2003
Available online at: www.astm.org

Authors' Response

Sir:

In response to Professor Saks's critique we first wish to provide some background to our work. Numerous challenges have been made in the courts against presenting expert forensic testimony on the grounds of whether the particular modality of forensic evidence has a scientific basis (1). Our research group has been engaged, for over two decades, in the domain of processing handwriting by computer (2). Both the legal debate and our research experience motivated us to undertake computer-based studies of handwriting individuality. Having worked with millions of samples of handwriting encountered by postal services we were familiar with the automatic analysis of computer-scanned handwriting. We had explored how to take advantage of the individual characteristics of a writer in deciphering an address but had not done, prior to the work reported in Ref (3), a study of whether handwriting was individual.

We decided to study as to what extent it can be said that no two people write alike, while considering the obvious fact that no one person writes exactly the same way twice. Multiple samples of the same writer were collected to account for the differences of the same writer. One goal of our research was to provide the scientific methodology to conduct the study. For this purpose we performed two different statistical pattern recognition tests known as identification and verification. In identification the goal is to determine from a handwriting sample its writer from a known set of writers. In verification the goal is to determine whether two handwriting samples were from the same individual or not. Identification accuracy is dependent upon the set of individuals considered and the result is not inferable to the population at large, e.g., a correct identification rate determined over a set of 1000 individuals (or subsets thereof) may not be the same when determined over a set of million individuals. On the other hand, verification accuracy is statistically inferable over the entire population when formulated as a test of whether the dissimilarity (or difference) between two samples belongs to one of only two distributions: same writer or different writer. Obtaining samples from more writers will refine the two distributions with the result converging to the true verification error rate. Thus the verification rate with 1000 writers is not likely to change much if we had a million writers, except for the confidence interval becoming narrower.

The fact that the accuracies in identification and verification were not quite 100% does not mean that perfect scores are not achievable. On the contrary, the fact that we reached high accuracies (98% for two-writer identification and 96% for verification) with simple measurements of handwriting encourages us that a more sophisticated analysis (better handwriting features) would only push the accuracies higher. Years of research in automating pattern recognition functions, e.g., handwriting recognition, face recognition, and even playing chess, show that cognitive skills of expert human beings exceed that of any computer. Therefore, newer versions of such machines are constantly being invented to reach human performance levels. Furthermore, our understanding is that to date even highly reliable DNA evidence falls short of the 100% mark in identification.

Professor Saks brings up six issues regarding the design choices in our experiment. We address each of the issues in the following.

1. *Samples are from a diverse population:* In order for our verification accuracy to be statistically valid we needed to have the samples from a diverse population. Samples from a biased population would make the verification study invalid. A population biased to a particular cohort group would be a useful study to perform in the future—to see whether the identification rates would be different from that of a more diverse group. Thus we stand by our design choice that a diverse population was needed for this initial study.
2. *Extrapolation of results from a limited sample to a general population:* This issue was addressed earlier in the discussion on identification and verification where we pointed out that verification results can indeed be extrapolated to the general population; by using more writers we would get a tighter confidence interval on the verification rate. With regards to whether our sample size was adequate, consider the fact that if 1000 writers provided three handwriting samples each, then we can generate 3000 same writer pairs and 4 495 500 different writer pairs. In the Benedictin Epidemiology, which was central to Ref (1), 13 different studies were conducted where the exposed total varied from 72 to 2720 and the nonexposed totals varied from 589 to 49 113 (4). Thus the sample size considered in our study is comparable to the largest samples used in the Benedictin studies.
3. *Size of the writing sample:* The size of the writing sample was large in that we required each writer to write one page of text. However, our results were reported for varying sizes of data being available for handwriting matching. In fact we showed that when just one word (of eight letters) and two additional letters were used for matching, verification rates exceeded 91%. Identification rates were between 87% and 99% (depending on the number of possible writers—which were reduced from a maximum of 900 possible writers to only two writers).
4. *Number of intra-writer samples:* Since we required each writer to write three times, our intra-writer sample size was 3000. In order to make the sample more representative of the entire population we used subsets of samples, such as 500 or 750 in our experiments. These numbers are not too small to give a high degree of confidence. Much more than three samples are available for particular letters and letter combinations.
5. *Type of writing:* We did not specify as to whether the writers were to use hand-printing or cursive writing, thereby allowing them to write in their preferred form. About a third of the writers chose handprinting or a mixture of handprinting and cursive writing. In general, there is not a clear-cut separation between cursive and handprint there being a continuum between the two. While a separate study for printed writing could be performed from our samples in the future, our experience with designing postal systems shows that all forms of writing are encountered in the real world. Thus a measure over all writing forms is a useful one.
6. *Conjecture of human ability:* The issue of human versus machine ability was discussed earlier. The issue of information overload is not pertinent to handwriting examination as it is a cognitive skill involving examining a small amount of questioned document material. Humans, particularly expert document examiners, are not overwhelmed when looking at small amounts of handwritten text when there is no time constraint. Humans are able to pick up all sorts of contextual clues, such as

Copyright © 2003 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

a subtle flick of a serif or an artistic loop, which a computer program is currently incapable of. Limited verification tests with nonexperts in our laboratory indicates that their performance rates are comparable to that of the machine. Other studies (5) have shown the superiority of expert document examiners over nonexperts, thereby justifying our conjecture that experts would perform better than our machine.

Our machine reached reasonably high verification rates with simple features. Factoring in that the machine did not take into account many obvious contextual cues present in handwriting leads us to assert that it is highly probable that each writer is unique. Taking into account previous comparisons of machines and expert humans performing several cognitive skills, including reading handwriting, leads us to assert that when handwriting samples are representative and large enough the tasks of verification and identification can be performed by an expert with high accuracy. We appreciate Professor Saks's remarks, which allowed us to explain more fully the design of our experiments, the results, and our conclusions.

References

1. United States Supreme Court Ruling, "Daubert vs Merrell Dow Pharmaceuticals," 509, 1993.
2. Plamondon R, Srihari SN. On-line and off-line handwriting recognition: a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000;22(1):63–84.
3. Srihari SN, Cha S-H, Arora H, Lee S. Individuality of handwriting, J Forensic Sci 2002;47(4):856–72.
4. Foster KR, Huber PW. Judging science: scientific knowledge and the federal courts, MIT Press, 1999.
5. Kam M, Fielding G, Conn R. Writer identification by professional document examiners. J Forensic Sci, 1997;42(5):778–86.

Sargur N. Srihari, Ph.D.
University of Buffalo
Amherst, NY 14228-2583