

REPORT 2
09/15/2017



FORENSIC SCIENCE ASSESSMENTS

A Quality and Gap Analysis

Latent Fingerprint Examination

AUTHORS:

William Thompson, John Black, Anil Jain and Joseph Kadane

CONTRIBUTING AAAS STAFF:

Deborah Runkle, Michelle Barretta and Mark S. Frankel



The report *Latent Fingerprint Examination* is part of the project *Forensic Science Assessments: A Quality and Gap Analysis*. The opinions, findings, recommendations expressed in the report are those of the authors, and do not necessarily reflect the official positions or policies of the American Association for the Advancement of Science.

© 2017 AAAS

This material may be freely reproduced with attribution.

Published by the American Association for the Advancement of Science
Scientific Responsibility, Human Rights, and Law Program
1200 New York Ave NW, Washington DC 20005
United States

The American Association for the Advancement of Science (AAAS) is the world's largest general scientific society and publisher of the Science family of journals. Science has the largest paid circulation of any peer-reviewed general science journal in the world. AAAS was founded in 1848 and includes nearly 250 affiliated societies and academies of sciences, serving 10 million individuals. The non-profit AAAS is open to all and fulfills its mission to "advance science and serve society" through initiatives in science policy, international programs, science education, public engagement, and more.

Cite as: AAAS, *Forensic Science Assessments: A Quality and Gap Analysis- Latent Fingerprint Examination*, (Report prepared by William Thompson, John Black, Anil Jain, and Joseph Kadane), September 2017.
DOI: 10.1126/srhl.aag2874

ACKNOWLEDGMENTS

AAAS is especially grateful to the Latent Fingerprint Examination Working Group, Dr. William Thompson, Chair, Mr. John Black, Dr. Anil Jain, and Dr. Jay Kadane, for their expertise, tireless work, and dedication in the preparation of this report.

AAAS benefitted considerably from the advice and contributions of the Advisory Committee. See Appendix F for a list of its members.

A special acknowledgment is due to Jessica Wyndham, Interim Director of the AAAS Scientific Responsibility, Human Rights and Law Program, for her feedback on the report, and for seeing it to its completion.

AAAS appreciates the contributions of Drs. Simon Cole, Jason Tangen, and forensic scientists Heidi Eldridge, Carey Hall and others for their review and insightful comments on a previous draft of this report.

Carey Hall, from the Minnesota Bureau of Criminal Apprehension, contributed to the report primer prepared by Mr. Black.

AAAS acknowledges the *Journal of Forensic Identification*, Alan McRoberts, IAI Editor, and forensic scientist Alice Maceo, for granting permission to reproduce an image from *Qualitative Assessment of Skin Deformation: A Pilot Study* (2009).

During part of the period that he worked on this report, William Thompson was a visiting fellow at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge as part of its program Probability and Statistics in Forensic Science which was supported by EPSRC Grant Number EP/K032208/1.

AAAS Marketing staff member, Elaine Jarrett, in collaboration with designers at Eyedea Advertising & Design Studio, created the cover design for this report.

Appreciation is also expressed for the assistance of several other AAAS colleagues-Bethany Spencer, Nicolle Rutledge, Joshua Ettinger, Ellen Platts, and Joel Ericson -- who provided administrative and logistical support for the Working Group meeting, contributed to and edited content for the project website.

FORENSIC SCIENCE ASSESSMENTS: A QUALITY AND GAP ANALYSIS

LATENT FINGERPRINT EXAMINATION

Disclaimer.....	i
Acknowledgments.....	ii
Table of Contents	
List of Abbreviations.....	iv
Preface.....	1
Genesis of this Report.....	1
Conclusions and Recommendations.....	5
Scientific Foundation for Latent Fingerprint Examination.....	13
Question I: <i>Is there an adequate scientific foundation for understanding the degree of variability of fingerprints: (a) among unrelated individuals; and (b) among relatives?.....</i>	17
Question II: <i>Is there an adequate scientific foundation for understanding the degree of variability among prints made by the same finger: (a) on different surfaces, under different environmental conditions; and (b) over time as a person ages or is subject to injury?.....</i>	24
Question III: <i>Is there an adequate scientific foundation for understanding the accuracy of automated fingerprint identification systems (AFIS)?.....</i>	29
Question IV: <i>Is there an adequate scientific foundation for understanding the potential for contextual bias in latent print analysis and how might it be addressed?.....</i>	35
Question V: <i>Is there an adequate scientific foundation for understanding the accuracy of human fingerprint examiners and how their accuracy is affected by: (a) level of training and experience; (b) individual differences in perceptual ability; (c) analytic procedures and standards of practice; (d) quality control and quality assurance procedures; and (e) the quality of prints? If not, what kinds of research are needed to improve understanding of these issues?.....</i>	43
Question VI: <i>In light of the existing scientific literature, what kind of statements might fingerprint examiners reasonably make in reports and testimony in order to appropriately convey both the strength and uncertainty associated with fingerprint evidence?.....</i>	60
References.....	74
Appendices	
A. Methods.....	93
B. Latent Fingerprint Examination - A Primer.....	95
C. Bibliography and Questions.....	99
D. Working Group Roster.....	151
E. Working Group Bios.....	152
F. Project Advisory Committee and Staff.....	154

LIST OF ABBREVIATIONS

ACE- V	Analysis, Comparison, Evaluation and Verification
AFIS	Automatic Fingerprint Identification System
CNM	Close Nonmatch
CTS	Collaborative Testing Service
DFSC	Defense Forensic Science Center
DOJ	Department of Justice
ELFT-EFS	Evaluation of Latent Fingerprint Technologies- Extended Feature Sets
ENFSI	European Network of Forensic Science Institutes
FNIR	False Negative Identification Rate
FPIR	False Positive Identification Rate
FpVTE	Fingerprint Vendor Technology Evaluation
IAI	International Association for Identification
LQAS	Latent Quality Assessment Software
NIST	National Institute of Standards and Technology
NRC	National Research Council
NSTC SoFS	National Science and Technology Council Subcommittee on Forensic Science
OSAC	Organization of Scientific Area Committees
PCAST	President’s Council of Advisors on Science and Technology
PiAnoS	Picture Annotation System
PRC	Probability of Random Correspondence
RDTE-IWG	Research, Development, Testing and Evaluation – Interagency Working Group
ROI	Region of Interest
SIVV	Spectral Image Validation and Verification
SWGFAST	Scientific Working Group on Friction Ridge Analysis, Study and Technology

PREFACE

Forensic science is an essential tool for investigating crime and helping to determine guilt or innocence at trial but questions have been raised about the validity and reliability of many forensic science disciplines. In some cases, reports and testimony based on substandard science have contributed to the convictions of individuals later proven innocent

This report addresses latent fingerprint examination, setting forth what is known about fingerprint examination and what remains unknown. Following the section on the Genesis of this Report, is a summary of conclusions and recommendations. Each conclusion is accompanied by a page number(s) referring to the location in the body of the report where support for it can be found. The body of the report then provides a detailed analysis of the scientific foundations for latent print examination, identifying strengths and weaknesses, and making suggestions for further research. The body of the report is followed by appendices that: (1) describe the methods that were used to prepare this report; (2) provide a primer on latent print examination; (3) provide a bibliography and questions prepared by the National Science and Technology Council's Subcommittee on Forensic Science that the working group used as a starting point for its analysis; (4) provide background on the working group that prepared this report on behalf of AAAS; and (5) provide background on the project advisory committee and staff.

GENESIS OF THIS REPORT

In its 1993 opinion in *Daubert v. Merrell Dow Pharmaceuticals*, the Supreme Court set forth standards for the admission of scientific evidence in litigation. The Court ruled that, under Rule 702 of the Federal Rules of Evidence, a "trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant but reliable" (*Daubert v. Merrell Dow*, 1993). In short, federal judges have an obligation to be "gatekeepers," separating good science from dubious science.

Serious questions have been raised, however, about how well judges have performed this role. Attorneys representing defendants in civil cases have had considerable success in persuading courts to shut the gate to questionable scientific evidence offered by plaintiffs, but attorneys representing criminal defendants have been less successful in mounting such challenges against forensic science evidence offered by prosecutors (Giannelli, 2007; Neufeld, 2005; NRC Report, 2009 at pp. 3-17; Risinger, 2000). Consequently, forensic science of uncertain validity has continued to be offered in court.

Congress was aware of the criticisms of forensic science and, in 2005, in the Science, State, Justice, Commerce, and Related Agencies Appropriation of 2006, authorized the National Academy of Sciences/National Research Council (NRC) to report on the status of forensic science (H.R. Rep. No. 109-272, 2005). In 2009, the NRC issued its report, "Strengthening Forensic Science in the United States: A Path Forward." The report's most significant conclusion was that much of forensic science as currently practiced has "little rigorous systematic research to validate the discipline's basic premises and techniques" (NRC Report, 2009 at p. 22). The NRC Report also offered a scathing assessment of the performance of federal judges in their gatekeeping role: "Federal appellate courts have not with any consistency or clarity imposed standards ensuring the application of scientifically valid reasoning and reliable methodology in criminal cases involving *Daubert* questions" (NRC, 2009, p. 11).

The NRC Report also highlighted another important issue for forensic science. In many disciplines the conclusions reached by examiners depend heavily on subjective human judgment and interpretation, and may consequently be influenced by non-scientific factors. As the NRC Report notes, it is especially problematic that “forensic science experts are vulnerable to cognitive and contextual bias” and that “there is no good evidence to indicate that the forensic science community has made a sufficient effort to address the bias issue” (NRC Report, 2009, at n. 8).

The NRC Report provided a good start on improving forensic science and, coming from such a prestigious organization, has been widely quoted and influential. Nevertheless, its overall critique of forensic science was broad and general. While it devoted several pages to discussion of the strengths and limitations of the foundational literature underlying latent print examination, it did not assess the scientific foundations of the field at the level of specificity needed to guide policy and inform the courts. Furthermore, while the NRC Report called for research from the “scientific community,” it did not provide a detailed research agenda to aid that community, including funding agencies.

In response to the NRC Report, the White House National Science and Technology Council’s Committee on Science in 2009 established a Subcommittee on Forensic Science (NSTC SoFS). SoFS members were told to focus on “developing practical and timely approaches to enhancing the validity and reliability...in forensic science....” (Charter of the Subcommittee, 2009). Further, the reauthorization of the SoFS charter “encouraged” it to “create a prioritized forensic science research agenda through a comprehensive gap analysis or other appropriate means....” (Charter of the Subcommittee, 2012).

SoFS appointed five Interagency Working Groups (IWGs), one of which, the Research, Development, Testing, and Evaluation (RDTE) IWG, was tasked with fulfilling the charter’s mandate. The RDTE IWG soon realized that it was not possible to undertake a gap analysis unless they had a better grasp of the extant literature. Members drew up a set of questions for ten forensic fields: (1) bite marks, (2) bloodstain patterns, (3) digital evidence, (4) fiber evidence, (5) fire investigation, (6) firearms and tool marks, (7) footwear and tire tracks, (8) hair evidence, (9) latent fingerprints, and (10) paints and coatings. The answers to these questions were deemed critical in determining whether or not the scientific foundation for a particular discipline was valid and reliable. The members sent the questions to relevant forensic entities representing these fields, asking them to identify books, articles or other sources that addressed the questions. The questions, in turn, were submitted to groups of practitioners in the field who submitted articles they deemed “foundational” to the field and believed constituted a sound scientific basis for addressing the questions. The final product was, thus, a series of questions or propositions and articles, or other scientific sources, purported to address the question. The result of this effort, while not a gap analysis, was an important step on the way to that goal, providing future investigators with an annotated bibliography for each of the ten fields.

In May 2014, the NSTC’s SoFS issued a report – “Strengthening Forensic Science: A Progress Report” – based on the conclusions reached by each of the five IWGs that together constituted the SoFS. The report stated that the RDTE IWG had “pursued the identification of foundational research that can be mapped to specific principles across the various disciplines of forensic science” (NSTC, 2014).

Coincident with the SoFS’ activities, Congress took note of the problems with current forensic practice. Specifically, Senator Jay Rockefeller, Chair of the US Senate Committee on Commerce, Science, and Transportation, held three hearings on forensics (December 2011 (U.S. Senate Committee, 2011), May 2012 (U.S. Senate Committee, 2012), and June 2013 (U.S. Senate Committee, 2013)). At the first hearing, the Senator commented that most Americans think that forensic science is “nearly infallible, always

conclusive.” He went on to say “the reality is far from this depiction.” Senator Rockefeller introduced the Forensic Science and Standards Bill of 2012 that recommended the federal government establish a “national research agenda to improve, expand, and coordinate Federal research in the forensic sciences.” There was no companion bill from the House, and the bill never became law (Library of Congress, 2012). And in 2011, Senator Patrick Leahy introduced the Criminal Justice and Forensic Science Reform Bill, which called for the establishment of “oversight and advisory offices and committees” that would be facilitated by the Department of Justice (DOJ) and the National Institute of Standards and Technology (NIST) and would “ensure that basic research is conducted to establish the validity and reliability of key forensic science disciplines.” This legislation also did not pass.

The federal government signaled its continuing interest in forensic science by establishing the National Commission on Forensic Science (the Commission) by the DOJ and NIST in April 2013 (Charter of the NCFS, 2014). The Commission’s responsibilities include the “identif[ication] and assess[ment of] the current and future needs of the forensic sciences to strengthen their disciplines....” At the Commission’s first meeting, in February 2014, John Holdren, then Assistant to the President for Science and Technology and Director of the White House Office of Science and Technology Policy, noted that the Commission’s recommendations will “help ensure that the forensic sciences are supported by the most rigorous standards available – a foundational requirement in a nation built on the credo of ‘justice for all’” (Department of Justice, 2014). And then Under Secretary of Commerce for Standards and Technology Patrick Gallagher said the Commission’s purpose is to “help ensure that forensic science is supported by the strongest possible science-based evidence gathering, analysis and measurement” ([NIST News](#), 2014, accessed September 12, 2017).

During the February 2014 meeting, several members commented that the appointment of the Commission might be premature. That is, it is pointless to recommend requirements for accreditation, training, and certification given the uncertain status of the sciences themselves. Why train forensic scientists to competently perform unreliable and invalid tests? These Commissioners argued that an analysis of the scientific and technical foundations for the forensic sciences should have precedence. In addition to the Commission, the 2009 NRC Report also prompted the redesign of the standards-making processes in the forensic science community. A memorandum of understanding between the DOJ and NIST resulted in the formation of a Forensic Science Standards Board with oversight over an Organization for Scientific Area Committees (OSAC) and 25 subcommittees dealing with various forensic science disciplines. Included among these subcommittees is one dealing with friction ridge analysis. This current report can inform the work by these OSAC subcommittees.

The Goals of AAAS

This evaluation is part of a larger project intended to point out where forensic practice is well founded in science and where it is not, and to produce a research agenda to serve as the basis for arriving at forensic methods that will inspire greater confidence in our criminal justice system. Although this analysis has long been needed, and many in the public and private spheres have urged its implementation, this AAAS project is the first attempt to fill the void.

To oversee this effort, AAAS has appointed a distinguished Project Advisory Committee to advise on all aspects of the project. The Advisory Committee members appointed include a law enforcement official, a social scientist, a cognitive psychologist, a law professor, a judge, a biomedical researcher, a forensic scientist, and a statistician (see Appendix F). The two fields studied by AAAS (fire investigation and latent fingerprint examination) were determined by the Advisory Committee based on (1) how often they are

used in criminal investigations, and (2) the degree of current controversy and the extent to which the field's legitimacy is being challenged. This report and a previous report on [fire investigation](#) were accomplished by Working Groups consisting of up to five members.

The Advisory Committee and AAAS staff worked collaboratively to identify the types of scientists needed for each Working Group, as well as individuals to fill those positions. The forensic practitioners appointed to the Working Groups were important in assuring the forensic science community that it had a voice at the table. Additionally, the forensic scientists helped staff and other Working Group members understand the forensic practice being studied, and wrote a primer on that forensic practice that is a part of each report.

Other Working Group members were chosen from the academic science and technology disciplines and may or may not have had any familiarity or previous experience with forensic science. In its report, the NRC noted that because most forensic science was developed in crime labs, not academic laboratories, and its practitioners had little training in research and statistical methods, the forensic fields had "never been exposed to stringent scientific testing." Similarly, Senator Rockefeller, quoting a witness who testified at one of his hearings, lamented that most forensic fields lack a "culture of science." Based on these descriptions of the practice of forensics, scientists and engineers trained in research methods and embedded in a "culture of science" are included in each Working Group. Of course, the specific disciplines of science and engineering differ for each Working Group. For example, the latent fingerprint Working Group includes an academic statistician, biometric engineer, and psychologist who studies human judgment and decision making, in addition to a forensic practitioner.

CONCLUSIONS & RECOMMENDATIONS

LATENT FINGERPRINT EXAMINATION

The conclusions presented in this section of the report are drawn from the technical section that follows. Recommendations are tied to the conclusions. Each conclusion is accompanied by page numbers to guide the reader to the location in the technical sections that support it.

I. Is there an adequate scientific foundation for understanding the degree of variability of fingerprints: (a) among unrelated individuals; and (b) among relatives?

Conclusions

- Scientific research has convincingly established that the ridge patterns on human fingers vary greatly among individuals. These findings provide a scientific basis for the use of fingerprint comparison to distinguish individuals (p. 17).
- The scientific literature does not, however, provide an adequate basis for assessing the rarity of any particular feature, or set of features, that might be found in a fingerprint. Examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and there are no scientific criteria for determining when the pool of possible sources is limited to a single person (p. 21).
- While there is somewhat less variability among relatives than among unrelated individuals, the degree of variability is still sufficient to make fingerprint comparison a useful tool, even for distinguishing monozygotic twins (p. 17).
- Because the characteristics of fingerprints are unlikely to be statistically independent, it will be difficult to determine the frequency of any particular combinations of features. While research of this type is important, it is unlikely to yield quick answers (p. 22).
- There have been promising efforts to develop quantitative methods for estimating the probative value or weight of fingerprint evidence. While these methods are not yet ready for courtroom use, research of this type deserves further support and may well offer a long-term solution to the problem of evaluating the probative value of latent print evidence (p. 22).
- In the short run (i.e., over the next few years), questions about the probative value of latent print evidence will more easily be addressed by research on the performance of latent print examiners under realistic conditions, as discussed below in Section V (p. 22).

Recommendations

1. Resources should be devoted to further research on possible quantitative methods for estimating the probative value or weight of fingerprint evidence.
2. Resources should also be devoted to further research on the performance of latent fingerprint examiners under typical laboratory conditions (as discussed in Section V).

II. Is there an adequate scientific foundation for understanding the degree of variability among prints made by the same finger: (a) on different surfaces, under different environmental conditions; and (b) over time as a person ages or is subject to injury?

Conclusions

- There is a body of research that addresses the question of how much variation is possible in prints from the same finger under different circumstances. But there is also evidence that latent print examiners sometimes misjudge the degree of intra-finger variability that may occur, creating a risk of false exclusions. It is unclear whether this problem occurs because examiners fail to appreciate the implications of existing research on intra-finger variability, or because existing research is incomplete (pp. 24, 26).
- Existing research provides adequate support for the claim that, barring injury or purposeful obliteration of friction ridges, there is little variation in fingerprints over time. While there can be minor changes due to aging, these changes are unlikely to undermine the accuracy of fingerprint identifications to any significant degree (p. 27).

Recommendations

3. Research is needed on how accurately latent print examiners can assess intra-finger variability- that is, the degree to which prints may be changed due to distortion. To the extent their assessments are imperfect, researchers should endeavor to determine whether that problem arises from inadequate understanding of the existing scientific literature (in which case better training is warranted) or whether it results from deficiencies in the existing literature (in which case more research on intra-finger variability may be needed).
4. Research is also needed on ways to reduce the probability of false exclusions.

III. Is there an adequate scientific foundation for understanding the accuracy of automated fingerprint identification systems (AFIS)?

Conclusions

- AFIS are designed to rapidly search fingerprint databases in order to identify the reference prints with features most similar to an input print. Research has established that AFIS are highly accurate for “ten-print to ten-print identification” (i.e., comparison of rolled or slapped prints of all ten fingers in order to identify sets from the same person) (pp. 30, 33).
- While AFIS are considerably less accurate for latent print comparison than for exemplar ten-print comparison (and are thought to be less accurate than human examiners for that purpose) they are necessary and extremely useful for screening large numbers of reference prints to identify possible sources of a latent (which can then be evaluated by human examiners) (pp. 29- 30).
- The commercial firms that have developed and marketed AFIS have kept the details of how they work secret for proprietary reasons. Consequently, members of the larger scientific community have found it difficult to assess the probative value of specific AFIS results for identifying particular individuals and difficult to assess how and why these systems might fail (p. 29).

- The existence of competing commercial AFIS has also created problems of interoperability across law enforcement agencies that need to be addressed (p. 31).
- Latent AFIS are not, at present, designed to determine whether any particular reference print was, or was not, made by the same finger as an input print; nor are they designed to assess the weight (probative value) of a fingerprint comparison for proving that any particular pair of prints has a common source. But it is possible that AFIS could evolve over time to perform those functions (p. 33).

Recommendations

5. NIST should continue to evaluate the performance of commercial AFIS systems, particularly their performance in identifying latent prints. Open tests in which vendors are invited to participate are important for spurring competition in order to assure continuing improvement in AFIS technology. Continued testing will help law enforcement agencies choose systems best suited for their needs and provide information to the broader scientific community on how well those systems work.
6. Developing better quantitative measures of the quality of latent prints should be a research priority. Such measures will be helpful for assessing and improving AFIS as well as for evaluating the performance of human examiners.
7. Law enforcement agencies and vendors should work together, perhaps with guidance from NIST, to better assure interoperability of AFIS systems and avoid compatibility problems that may result in the loss of valuable evidence or investigative leads.

IV. Is there an adequate scientific foundation for understanding the potential for contextual bias in latent print analysis and how might it be addressed?

Conclusions

- Studies have shown that latent print examiners (like all human beings) are vulnerable to contextual bias. Their evaluations of the features of latent prints can be influenced inappropriately by premature exposure to reference prints; their conclusions can also be influenced inappropriately by information about the underlying criminal investigation. Contextual bias of this type is unlikely to change examiners' opinions in clear-cut cases, but may have stronger effects in ambiguous cases where the prints are more difficult to evaluate (p. 35).
- Contextual bias can happen without an examiner's conscious awareness, and, therefore, cannot be reliably suppressed or corrected by the individual (p. 35).
- Contextual bias can be mitigated through the use of context management procedures that avoid exposing examiners to contextual information that is unnecessary for a scientific assessment of the prints and delay exposing examiners to information that, while necessary, may create bias if presented prematurely. A number of laboratories have adopted context management procedures as a means of mitigating contextual bias in latent print analysis (p. 36).

Recommendation

8. Context management procedures should be adopted by all forensic laboratories in order to reduce the potential for contextual bias in latent print examination. Some examples of such procedures include blinding examiners to task-irrelevant information, using case managers, and sequencing workflow among analysts (i.e., using “sequential unmasking” or “linear ACE-V”). Laboratories that lack sufficient staff to administer context management procedures internally should deal with the problem through cooperative procedures with other laboratories.

V. Is there an adequate scientific foundation for understanding the accuracy of human fingerprint examiners and how their accuracy is affected by: (a) level of training and experience; (b) individual differences in perceptual ability; (c) analytic procedures and standards of practice; (d) quality control and quality assurance procedures; and (e) the quality of prints? If not, what kinds of research are needed to improve understanding of these issues?

Conclusions

- Studies of the accuracy of latent print examiners leave no doubt that trained fingerprint examiners have expertise that makes them much better at the task of fingerprint comparison than non-experts (pp. 44- 45).
- It is not clear, however, whether the error rates observed in existing studies reflect the error rates that occur in the actual practice of latent print analysis. Such studies can in principle determine the relative strength of different analysts and the relative difficulty of different comparisons, however, the relationship of such findings to the error rate in a specific case is problematic (p. 46).
- The probability of error in a particular case may vary considerably depending on the difficulty of the comparison. Factors such as the quality of the prints, the amount of detail present, and whether the known print was selected based on its similarity to the latent will all be important (p. 45).
- The probability of error will also depend on the examiner’s implicit thresholds for deciding what to report. These thresholds may be affected by the perceived consequences of different types of error, which could in turn be influenced by contextual information about the case. The thresholds may be different when examiners know they are being tested as compared to when examiners are doing routine casework (p. 46).
- The best way to study the performance of latent print examiners is to introduce known-source research samples into the routine flow of casework, so that examiners do not know their performance is being studied. Research of this type is much easier to do in laboratories that employ context management procedures because examiners in those laboratories are already blind to the source of the samples (which makes it easy for laboratory managers to give them research samples without their awareness that it is a research sample rather than an actual case) (p. 48).
- Research on examiner performance with various types of samples will help examiners realize their limitations and improve their skills by giving them feedback on their accuracy.

Once laboratories develop the capability of introducing research samples into the normal flow of casework, it will be possible to design and carry out studies on a variety of important issues, such as how the number and kinds of features discernable in the latent print affect the accuracy of examiners' conclusions, the extent to which examiner accuracy is impaired by distortion of the prints, the effects of examiner characteristics (e.g., visual acuity, workload) on accuracy, and the effect of training on accuracy (pp. 49- 50).

Recommendations

9. Forensic laboratories should undertake programs of research on factors affecting the performance of latent print examiners. The research should be done by introducing known-source prints into the flow of casework in a manner that makes test samples indistinguishable from casework samples.
10. Government funding agencies should facilitate research of this type by providing incentive funding to laboratories that undertake such programs and by funding the creation of research test sets- i.e., latent print specimens of known source that can be used for testing examiner performance. The research test sets should be designed with the help of practitioners, statisticians, and experts on human performance to ensure that the research is scientifically rigorous and that it addresses the issues most important to the field.
11. In research of this type, errors are to be expected and should be treated as opportunities for learning and improvement. It is not appropriate for examiners to be punished or to suffer other disadvantages if they make errors in research studies that are designed to test the limits of human performance. Nor is it appropriate for laboratories to suffer disadvantage as a consequence of engaging in such research. Accordingly, the criminal justice system should consider carefully whether the information about examiner performance in research studies should even be admissible as evidence. If the results of research studies are admitted as evidence in the courtroom, it should only be under narrow circumstances, and with careful explanation of the limitations of such data for establishing the probability of error in a given case.

VI. In light of the existing scientific literature, what kind of statements might fingerprint examiners reasonably make in reports and testimony in order to appropriately convey both the strength and uncertainty associated with fingerprint evidence?

Conclusions

- Latent print examiners traditionally claimed to be able to "identify" the source of a latent print with 100% accuracy. These claims were clearly overstated and are now widely recognized as indefensible. While latent print examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, there is no scientific basis for estimating the number of people who could not be excluded and, consequently, no scientific basis for determining when the pool of possible sources is limited to a single person. Moreover, research on examiners' accuracy when comparing known-source prints has provided ample evidence that false identifications can and do occur (p. 60).

- The Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST) and the DOJ published documents directing latent print examiners to continue to use the term “identification” in reports and testimony but add qualifying language that acknowledges an element of uncertainty. In our view, these proposals fail to deal forthrightly with the level of uncertainty that exists in latent print examination. In our view, the proposed reporting language allows examiners to make claims that cannot be justified scientifically (pp. 60- 61).
- Guidelines for evaluative reporting recommended by the European Network of Forensic Science Institutes (ENFSI) require examiners to estimate likelihood ratios. An advantage of this approach is that it avoids the need for an expert to decide whether the observed similarities and differences between prints are sufficient to justify a particular categorical conclusion. A disadvantage (shared with other subjective methods) is that it requires examiners to make subjective estimates of the probability of the observed data¹ under alternative hypotheses about the source of the prints. Whether examiners can make such judgments reliably and accurately is, at present, unknown (pp. 63- 65).
- The Defense Forensic Science Center (DFSC) of the Department of the Army has suggested a more moderate approach to reporting the strength of latent print evidence. Although we have some concerns about the specific language suggested by the DFSC, the approach taken by DFSC is preferable to those suggested by SWGFAST and the DOJ because the statements it allows are easier to defend scientifically. We propose some alternative language that we believe will be easier for lay people to understand (pp. 65- 67).
- The 2009 NRC Report called for the development of “quantifiable measures of the uncertainty in the conclusions of forensic analyses.” We agree that there is a need to quantify the strength of latent print evidence so that it is not left as a matter of subjective judgment, however, the existing scientific literature does not, at present, provide an adequate basis for quantitative estimates of the accuracy of the various determinations latent print examiners make on the basis of latent print evidence (p. 67).
- In the future, it might be possible to rely on AFIS systems to generate quantitative estimates of the probative value of a particular print comparison. At present, AFIS systems are not designed to provide such estimates. While it might be possible for AFIS systems of the future to generate meaningful statistical measures of evidentiary strength, additional research will be needed to determine whether this approach is viable (p. 67).
- A number of researchers have attempted to use mathematical models to describe the strength of fingerprint evidence for proving that prints have a common source. This research is promising and deserves further support, although the models are not yet ready for use in the courtroom (pp. 67- 68).
- In the short-term, studies of the type discussed previously in which research samples are introduced into the routine flow of casework could provide a valuable source of data on the strength of fingerprint evidence but this is problematic in agencies in which handling of physical evidence by examiners is necessary. The research could potentially allow examiners

¹ The data that a latent print examiner may observe when comparing two fingerprint impressions includes information at all levels of detail potentially observable in the impression. For discussion of the various features and levels of detail that latent print examiners may see in print impressions, see Appendix B.

to report the rates of false positives and false negatives that have occurred when latent print examiners made similar comparisons involving known-source prints. Relative error-rate estimates derived from such data will inevitably be approximate and imperfect due to uncertainty about whether the comparison being evaluated in a particular instance posed the same level of difficulty as the comparisons examined in the research. Such data could nevertheless provide important insights into how high, or how low, the rate of error might be in practice (p. 70).

- Members of the public are likely to hold misconceptions about latent print examination that have been shaped by decades of overstatement by latent print examiners. To combat these misperceptions, latent print examiners should include specific caveats in reports that acknowledge the limitations of the discipline. They should acknowledge: (1) that the conclusions being reported are opinions rather than facts (as in all pattern-matching disciplines), (2) that it is not possible for a latent print examiner to determine that two friction ridge impressions originated from the same source to the exclusion of all others; and (3) that errors have occurred in studies of the accuracy of latent print examination (p. 71).

Recommendations

- 12.** Examiners should be careful not to make statements in reports or testimony that exaggerate the certainty of their conclusions. They can indicate that the differences between a known and latent print are such that the donor of the known print can be excluded as the source of the latent print. They can also indicate that the similarity between a latent and a known print are such that the donor of the known print cannot be excluded as the source of the latent print. But they should avoid statements that claim or imply that the pool of possible sources is limited to a single person. Terms like “match,” “identification,” “individualization” and their synonyms, imply more than the science can sustain.

Because fingerprint impressions are known to be highly variable, examiners who observe a high degree of correspondence between a latent print and known print may be justified in making statements about the rarity of the shared features.

For example, examiners might say something like the following:

"The latent print on Exhibit ## and the record fingerprint bearing the name XXXX have a great deal of corresponding ridge detail with no differences that would indicate they were made by different fingers. There is no way to determine how many other people might have a finger with a corresponding set of ridge features, but this degree of similarity is far greater than I have ever seen in non-matched comparisons."

- 13.** When latent print examiners testify, they should be prepared to discuss forthrightly the results of research studies that tested the accuracy of latent print examiners on realistic known-source samples.

14. Further research is also needed on how lay people, such as police officers, lawyers, judges, and jurors evaluate and respond to fingerprint evidence. Research of this type would be helpful in evaluating how best to present fingerprint evidence in reports and expert testimony. It would help ensure that the statements made in reports and testimony will be interpreted in the intended manner.

SCIENTIFIC FOUNDATION FOR LATENT FINGERPRINT EXAMINATION

Introduction and Overview

This report has the following goals:

- To identify key scientific questions that must be addressed in order to assess the value of latent print evidence;
- To determine how well those questions have been answered by published scientific studies and identify gaps in the existing scientific knowledge;
- To prepare a research agenda needed to:
 - strengthen analytic methods;
 - reduce the potential for error; and
 - better evaluate the probative value of latent print evidence;
- To identify, in light of the above, changes that are warranted in
 - the way latent print examination is performed, and
 - the way latent print evidence is characterized in reports and testimony.

Latent print examiners have traditionally claimed that their conclusions rest on two fundamental premises: (1) ridge patterns from different fingers are different (*uniqueness*) and (2) fingerprint patterns do not change over time (*persistence*). In practice, these two premises can be restated as claims that there is a very high degree of variability among prints made by different fingers, but a low degree of variability within prints made by the same finger.

Accordingly, when assessing the scientific foundation of latent print examination, it is necessary first to consider what is known about: (1) the degree of variability among the prints of different fingers (inter-finger variability); and (2) the degree of variability within prints made by the same finger (intra-finger variability). We address these issues below in Sections I and II and conclude that there is indeed a high level of inter-finger variability, and a relatively low level of intra-finger variability, and hence that examination of fingerprints has great utility for distinguishing among individuals. We note some important weaknesses, however, in the scientific foundation for understanding both inter-print and intra-print variability and make recommendations for additional research to fill these gaps.

Whether fingerprints can, in principle, be distinguished is a different question from whether latent print examination is reliable and accurate in practice. Even if the ridge detail of every finger were unique and unchangeable, it does not follow that every impression made by every finger will always be distinguishable from every impression made by any other finger, particularly if the impressions are of poor quality (e.g., limited detail, smudged, distorted, overlaid on another impression). By analogy, it may be that every human face is unique, but we can still mistake one person for another, particularly when comparing poor-quality photos.

This is a limitation that has been noted by academic commentators (Cole, 2004) and acknowledged by fingerprint examiners. As a prominent latent print examiner recently explained:

When fingerprint comparisons are being made, they are not being made from friction ridge skin to friction ridge skin. They are being made from one imperfect, incomplete recording to another... [hence] correctly associating a degraded mark to its true source is by no means a certainty, even were one to presume absolute uniqueness of all friction ridge skin (Eldridge, 2017, p. 75).

Consequently, when examining the scientific foundation of latent print examination, it is necessary to consider more than the degree of inter-finger and intra-finger variability, one must also consider research on how well examiners can, in practice, distinguish the impressions of different fingers.

The President's Council of Advisors on Science and Technology (PCAST) recently issued an important report: "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods."² The PCAST report sets forth in considerable detail what is required to establish the validity of a method for assessing whether two items, such as fingerprints, have a common source based on a comparison of their features. The report distinguishes two types of validity: (1) *Foundational validity* means that empirical studies have shown that the method is "repeatable, reproducible, and accurate" when used in ways that are "appropriate for the intended application" in other words, that the method "can, in principle, be reliable" (p. 4-5); and (2) *Validity as applied* means the method has been applied in a manner that is reliable and appropriate in the case at hand. The PCAST report concludes, and we agree, that foundational validity can only be established by empirical research:

Scientific validity and reliability require that a method has been subjected to empirical testing, under conditions appropriate to its intended use, that provides valid estimates of how often the method reaches an incorrect conclusion.... Without appropriate estimates of accuracy, an examiner's statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy. (PCAST, 2016, p. 46)

Because we are concerned with the validity and probative value of latent print evidence, we have looked carefully at empirical research on the accuracy of latent print examination. We consider the accuracy of AFIS in Section III and the accuracy of human examiners in Section V. We conclude that both AFIS and human examiners have been shown to have high levels of accuracy in empirical studies. But we find important weaknesses in the existing scientific literature in both areas. Particularly with regard to the performance of human examiners we find that more and better research is needed. We recommend new programs of research that we believe will greatly benefit the field of latent print examination.

² This AAAS report had already been drafted when the PCAST report was issued. We were pleased to see that PCAST, in commenting on the scientific foundations of latent print examination, had independently reached many of the same conclusions that we had reached. There are, however, some points of difference between our conclusions and those of the PCAST report which we will note in the sections that follow.

Latent print examination relies heavily on the judgment and expertise of human examiners. While examiners often have valid, objective reasons for their judgments, the process of judgment itself depends on the knowledge, training, experience, memory and perceptiveness of the individual examiners and there is a possibility that different examiners will reach different conclusions. The PCAST report commented thoughtfully on the challenges of evaluating the validity of a forensic science method that relies, in part, on subjective judgment.

Subjective methods require particularly careful scrutiny because their heavy reliance on human judgment means they are especially vulnerable to human error, inconsistency across examiners, and cognitive bias. In the forensic feature-comparison disciplines, cognitive bias includes the phenomena that, in certain settings, humans may tend naturally to focus on similarities between samples and discount differences and may also be influenced by extraneous information and external pressures about a case. (PCAST, 2016, p. 5)

We agree with this assessment and, for that reason, have included a separate section (Section IV) that discusses what the scientific literature has revealed regarding the potential for cognitive and contextual bias in latent print examination and how this problem might be addressed. While we recommend additional research to assess the scope of the problem and how best to deal with it, we find (based on the existing scientific literature) that the risk of bias is strong enough to warrant taking protective steps. We outline measures that laboratories can and should take to mitigate this risk.

The ultimate goal of the discipline of latent print examination is to provide useful evidence to the legal system. In order to be useful, latent print evidence must be presented in an appropriate manner that avoids misstating its value; it must also be understood by its intended audience of investigators, lawyers, judges and jurors. Because we regard the communication of findings as an integral part of the discipline of latent print examination, our analysis included examination of the statements latent print examiners make about their conclusions in reports and testimony.

In Section VI of the report we survey various approaches latent print examiners have taken when communicating their findings as well as new approaches that have been suggested by commentators and advisory bodies. For each approach, we discuss: (1) whether the statements are warranted by the existing scientific literature on latent print examination; and (2) whether the statements are likely to be understood. We conclude that some statements that examiners made in the past, and some statements recommended for presentation in the future, are not well supported by the existing scientific literature on the accuracy of latent print examination. We also conclude that there are important gaps in the scientific literature on the communication of findings. While existing research has identified some statements that should not be presented because they are likely to be misunderstood, more research is needed to establish best practices for the communication of findings.

In summary, this report will address six foundational issues underlying latent print examination: inter-print variability of fingerprints (Section I); intra-print variability of fingerprints (Section II); accuracy of AFIS (Section III); cognitive bias and how to deal with it (Section IV); accuracy of human examiners (Section V); and communication of findings (Section VI).³

³ The six sections of this report each address a key question about latent print analysis. As discussed in Appendix A, the Working Group that prepared this report began by considering 15 questions about latent print analysis that had been identified as pertinent by the Interagency Working Group (IWG) of the Subcommittee on Forensic Science (SoFS) of the White House National Science and Technology Council. After careful consideration of the IWG efforts to define the issues, the Working Group concluded that the six questions around which this report is organized address the scientific validity of latent print analysis for use in criminal identification in a more concise and cogent manner. There is, however, general correspondence between the questions addressed in the six sections of this report (Sections 1-6) and the fifteen questions raised by the IWG (I – XV), as follows: Section 1 (I, II), Section 2 (III, IV), Section 3 (VI), Section 4 (V, VII, IX, X, XI, XII), Section 5 (XIII), Section 6 (not covered by IWG questions).

1. Is there an adequate scientific foundation for understanding the degree of variability of fingerprints: (a) among unrelated individuals; and (b) among relatives?

Existing scientific literature convincingly establishes that the friction ridge detail potentially observable in fingerprints varies greatly among individuals and hence that fingerprint comparison is a viable way to distinguish individuals. While there is somewhat less variability among relatives than among unrelated individuals, the degree of variability is still sufficient to make fingerprint comparison a useful tool, even for distinguishing monozygotic twins. As explained later in this section, the genetic component in fingerprint variability is not large enough to be an important factor in the utility of fingerprint identification (Jain, Prabhakar, and Pankanti, 2002).

The variability of friction ridge patterns in fingerprints arises from prenatal morphogenesis of dermatoglyphic (friction ridge) traits. Babler (1991) provided a comprehensive overview of the studies related to the embryologic development of friction ridge skin starting around 10 weeks' post-fertilization. He reported that the process of ridge formation is not an event occurring simultaneously across the whole finger surface, but initiates at several points on the surface and spreads out until these ridges finally meet. The factors that affect ridge configuration include growth stress, topography of the volar pads, the surface distribution of nerves, and bone development. Wertheim (2011) reviewed literature indicating that even though genetic information directs cellular function and determines a person's appearance, numerous steps and other factors are involved between the genesis of the DNA-encoded protein and the final development of the human body. So even identical twins would develop distinguishable friction ridge skin detail (Jain, Prabhakar and Pankanti 2000; Langenburg 2011; Tao et al., 2012).

Based on this literature, Wertheim and Maceo (2002) argued that (1) fingerprint ridge counts are predominantly affected by two temporal events, the onset of epidermal cellular proliferation and the timing of the regression of the volar pads and (2) fingerprint pattern types are predominantly affected by the symmetry of the volar pad. Although a number of plausible mechanisms for the formation of fingerprint patterns (epidermal ridge) have been proposed, Kücken (2007) argues that the pattern arises as a result of a buckling (folding) process in the cell layer of the epidermis. With this model, the well-known observation that the pattern type is related to the geometry of the embryonal fingertip can be explained. Kumbhani (2007) reviewed the literature published during the years 1891 to 2005, covering frequency of various fingerprint patterns in different communities and populations. These references explain the formation of fingerprints from a biological/dermatoglyphic basis. In addition, the book by Mavalwala (1977) provides an extensive bibliography (over 3,000 references) on the subject of dermatoglyphics. It is a good resource for research regarding the distribution of level 1 features (pattern type or class and ridge counts) on both fingers and palms in a variety of human populations (small tribes to large ethnic groups). Loesch (1983) covers the classification of dermatoglyphic patterns and their topology, the variability within and between human

populations of topologically significant pattern elements, their heritability, and diagnostic applications.

Since identical twins are supposed to have the same/similar genetic constitution, the distinctiveness of twins' fingerprints supports the theory that ridge development is strongly affected by random processes. The study by MacArthur (1938) used differences in terms of ridge count, finger pattern, palmar digital and axial triradii, and palm patterns for the dermatoglyphic similarity analysis. Difference comparisons between parents and children, siblings, mono- and dizygotic twins, and left and right hands of individuals suggested that fingerprints have the potential to distinguish even monozygotic twins. Colletto and Zimmerknopf (1987) proposed the use of dermatoglyphic analysis to determine twin zygosity using 120 same sex twin pairs, of which 66 were monozygous and 54 were dizygous. A discriminant analysis was applied to within-pair differences of 35 dermatoglyphic variables; a subset of 15 variables was selected that resulted in a perfect classification of the twin pairs in their study.

Because the presence of an arch pattern on at least one fingertip has previously been suggested as an autosomal dominant trait with reduced penetrance, Reed et al. (2006) used fingerprints from 2,484 twin-pairs to estimate the heritability for the presence of at least one fingertip arch pattern. The best fitting model showed a heritability of 91%, indicating that it is justified to conclude that specific genes influence the occurrence of fingertip arch patterns. Identical twins have the closest genetics-based relationship and, therefore, the maximum similarity between fingerprints is expected to be found among identical twins. Nevertheless, a study conducted by Jain et al. (2002) showed that a state-of-the-art AFIS could successfully distinguish identical twins, though with a slightly lower accuracy than non-twins. Another observation made by Jain et al. was that the fingerprints of two identical twins will have the same class label with much higher probability than the general population. Similar observations were made by Liu and Srihari (2009), and Srihari et al. (2008). Instead of fingerprints, Okajima and Usukara (1984) collected palmprints from 20 pairs of Japanese twins and examined the total minutiae count and fork index (the proportion of forks in total minutiae) on this database. Both traits presented relatively high intra-class correlation coefficients between monozygotic twins. Nevertheless, the ability of AFIS to distinguish even monozygotic twins with relatively high accuracy suggests that the genetic component in fingerprint variability is not large enough to be an important factor in the utility of fingerprint identification (Jain, Prabhakar, and Pankanti, 2002).

Fingerprints vary with respect to three levels of features: (1) level 1 features are macro details such as ridge flow and pattern type; (2) level 2 features include minutiae points, such as bifurcations and ridge endings; and (3) level 3 features include all dimensional attributes of a ridge, such as ridge path deviation, width, shape, pores, edge contour, incipient ridges, breaks, creases, scars, and other permanent details (Champod et al., 2016).

Determining just how distinctive fingerprints might be has been a major focus of research. In an important review article, Stoney (2001) discussed ten different models for measuring the

rarity of fingerprint features, published from 1982 to 1999. He pointed out that the models were (with one exception) based on very limited data and argued persuasively that none of the statistical models had been validated.

Pankanti, Prabhakar and Jain (2002) derived a theoretical formulation for estimating the probability of falsely associating minutiae-based representations from two arbitrary fingerprints belonging to different fingers. They compared these probability estimates with typical AFIS (fingerprint matcher) accuracy results. This study has important limitations, however, in that it makes assumptions about the distribution and independence of fingerprint features that appear not to be true. It assumed a uniform distribution on minutiae locations and directions, which is not true for fingerprint images; it also assumed (despite evidence to the contrary) that minutiae locations are distributed independently of the minutiae direction.

In order to address these weaknesses in Pankanti et al.'s model, Zhu et al. (2007) and Dass et al. (2009) developed a family of finite mixture models to represent the distribution of minutiae in fingerprint images, including minutiae clustering tendencies and dependencies in different regions of the fingerprint image domain. Consider two fingerprint images each containing m and n minutiae points, respectively. A mathematical model that computes the probability of random correspondence (PRC) with k minutiae matches was derived based on the mixture models. A PRC of 2.25×10^{-6} corresponding to $k = 12^4$ minutiae matches was computed for the NIST4 Special Database (see www.nist.gov/srd/nistsd4.cfm), when the numbers of query and template minutiae both equal 46 ($m=n=46$).⁵ The PRC values declined as the models required larger numbers of minutiae matches and as additional matching features beyond minutiae were specified. For example, Chen and Jain (2009) showed that for $m=n=52$ and $k=12$, the $PRC=5.77 \times 10^{-7}$. But, when ridge period and curvature are incorporated in addition to minutiae, the PRC value is reduced to 1.86×10^{-10} . When pore spacing is incorporated, the PRC value drops further to 6.93×10^{-11} . Since the distinctiveness of prints is highly affected by fingerprint quality, Dass (2010) quantitatively studied the effect of noise in minutiae detection and localization, resulting from varying image quality, on fingerprint individuality. Empirical results on two fingerprint databases demonstrated how the PRC values increase as the fingerprint image quality becomes poor. The PRC value corresponding to the "12-point match" with 26 observed minutiae in the query and template fingerprints increased by several orders of magnitude when the fingerprint quality degraded from "best" to "poor."

In addition to level 2 features (e.g., minutiae), statistical models for fingerprint individuality based on level 1 features (e.g., class type, ridge flow, etc.) and level 3 features (e.g., pores) have been developed. Su and Srihari (2009) develop three generative models to represent the distribution of fingerprint features: ridge flow, minutiae, and minutiae together with ridge points, respectively. Three metrics were discussed in this paper: (1) PRC of two samples, (2) PRC

⁴ Many papers on fingerprint individuality provide the probability of random correspondence for a match consisting of 12 minutiae points (12-point guideline).

⁵ To put these numbers in perspective, consider that there are more than 2.25×10^9 Americans, and hence that, for any given set of features, the model suggests more than 1000 people in the country would have corresponding features.

among a random set of n samples (n PRC); and (3) PRC between a specific sample and a random set of n samples (specific n PRC). Experimental results showed that the theoretical estimates of fingerprint individuality using this model consistently followed the empirical values based on the NIST4 database.

There have been noteworthy efforts to use information about pores to improve AFIS accuracy. Roddy and Stosz (1997) developed a model to predict the performance of the pore-based fingerprint matching. Their paper provides statistics on fingerprint pores and discusses the efficacy of using pores in addition to the traditionally used minutiae to improve AFIS performance. Parsons et al. (2008) attempted to answer the following question: Can pore information supplement expert judgments with a more quantitative measure of the strength of evidence? They used an automatic method for pore point extraction and statistical methods using spatial point processes for evidential value analysis. The results of this preliminary analysis give a positive answer to the above question. Anthonioz and Champod (2014a) proposed to assess the strength of evidence by integrating pores in an evaluative framework including level 1 and level 2 details using a likelihood ratio (LR) based approach. Their approach provides LRs that reflect the relative probability of the observed features under alternative propositions about the source of the prints (i.e., same-source or different-source). As will be discussed further in Section VI of this report, one of the main advantages of such an approach is that it provides transparency regarding the weights to assign to each feature type. Chen and Jain (2009) exploited minutiae feature as well as non-minutiae features used for matching fingerprints by incorporating all three levels of fingerprint features: pattern or class type (level 1); minutiae and ridges (level 2); and pores (level 3). Correlations among these features and their distributions were also taken into account in their model. Experimental results show that the theoretical estimates of the rarity of fingerprint patterns using this model consistently follow the empirical values on the public domain NIST4 database.

In sum, there is tremendous variability among prints made by different fingers. This variability clearly provides a scientific basis for using fingerprints to distinguish individuals. It is possible to draw histograms showing the probability of finding a given degree of similarity between two fingerprints. Because inter-finger variability is very high, and intra-finger variability relatively low (as will be discussed in the following section), the histogram for same-source prints (sometimes called the genuine score distribution), is markedly higher than the histogram for different source prints (sometimes called the imposter score distribution). Nevertheless, the genuine and imposter score distributions have finite variance and, in practice, they almost always overlap (see Figure 1). One of the major challenges facing fingerprint recognition system designers is to develop fingerprint representation schemes and similarity measures to minimize these variances and the degree of overlap between the genuine and imposter distributions. The challenge for latent print examiners is to do the same thing case-by-case, using their individual powers of observation, memory and judgment.

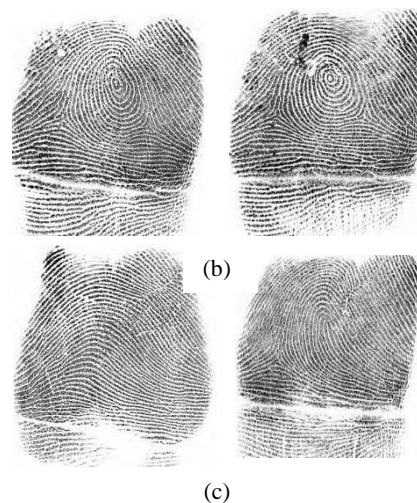
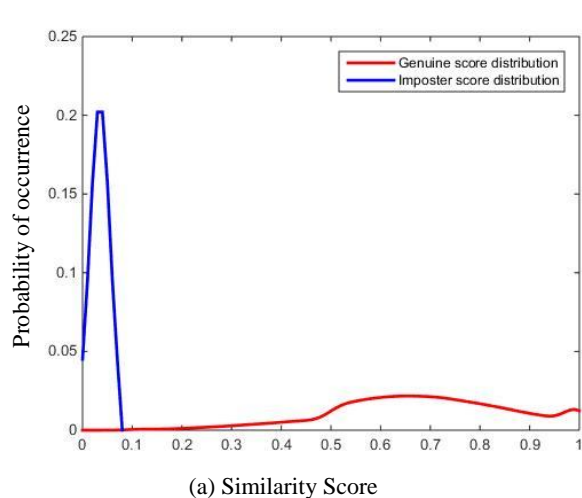


Figure 1. (a) Genuine and imposter score distributions from a state-of-the-art commercial off-the-shelf (COTS) matcher, and (b) and (c) are genuine and imposter fingerprint pairs, respectively. The scores in (a) were generated from 2,000 rolled fingerprint paired in NIST SD4 using a state-of-the-art fingerprint matcher (images provided by Anil Jain).

While the existing scientific literature indicates a low likelihood that prints from different individuals share a large number of common features, the literature does not provide an adequate basis for assessing the rarity of any particular feature, or set of features, that might be found in a print (Nagar, Choi, and Jain, 2012; Su and Srihari, 2010). While some research has been conducted on the frequency of fingerprint characteristics in various human populations (e.g., Loesch, 1983; Mavalwala, 1977; Sarkar, 2004; for review, see Kumbhani, 2007), this line of research is in its infancy. Consequently, there is uncertainty about how many matching features and what types of matching features are necessary to reduce the potential donor pool to a single source. Examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and there are no scientific criteria for determining that the pool of possible sources is limited to a single person.

Uncertainty in this area means there is an inadequate scientific foundation for determining how many features, of what types, are needed in order for an examiner to draw definitive conclusions about whether a latent print was made by a given individual (and hence, as discussed further in Section VI, that examiners should not draw such conclusions). It also means that there is an inadequate scientific foundation for evaluating the probative value of the limited feature sets observed in many lower-quality latent prints.

Additional research on the frequency of individual fingerprint characteristics in various human populations would be one possible way to reduce uncertainty in this area. Similar research on the frequency of the genetic characteristics examined in forensic DNA analysis has provided a solid quantitative basis for assessing the probative value of DNA evidence. But progress on this issue is likely to be more difficult for fingerprints than it was for DNA evidence. The probability of finding a set of genetic features in a DNA test is relatively easy to estimate because the

features occur at rates that are statistically independent of one another.⁶ Statistical dependencies are more likely for fingerprint features and will make it far more difficult to estimate the frequency of combinations of features for fingerprints than for DNA profiles (Chen and Jain, 2009). Consequently, research of this type, while important, is unlikely to yield quick answers.

Examples of the kinds of dependencies that can occur among fingerprint features can be found in studies of the variability in fingerprint minutiae (e.g., Gutiérrez et al., 2007; Gutiérrez-Redomero et al. 2011). The research by Gutiérrez and colleagues found statistically significant differences in a Spanish population in the distributions of minutiae types between the genders and among different types of fingerprint patterns (arch, loops, and whorls). The distribution of minutiae types also varied among fingers—for example, ridge endings were more common on thumbs and index fingers while bifurcations and convergences were more common on the middle, ring, and little fingers. As discussed earlier, there are also dependencies between minutiae location and direction (Pankanti et al., 2002). The distribution and frequency of fingerprint features may also vary among human populations (Kumbhani, 2007; Sarkar, 2004; Seweryn, 2005). Hence, models for estimating the frequency of specific combinations of features are likely to be complex and will require extensive data from multiple populations.

Despite these challenges, there have been some promising efforts to develop quantitative methods for estimating the probative value or weight of fingerprint evidence (Neumann, Evett and Skerrett, 2012;⁷ Abraham et al., 2013; Anthonioz and Champod, 2014b). These methods estimate the relative probability of observing specific similarities between two fingerprints under two alternative hypotheses about the source of the prints: (1) that the prints are from the same finger; and (2) that the prints are from different fingers. They derive statistical estimates in part by searching databases of print images to assess the rarity of specific combinations of features. For reasons discussed further in Section VI, we do not think these models are ready to be used in the courtroom at this time. However, research of this type deserves further support and may well offer a long-term solution to the problem of evaluating the probative value of latent print evidence. In the short run (i.e., over the next few years), questions about the probative value of latent print evidence will more easily be addressed by research on the performance of latent print examiners under realistic conditions, as discussed below in Section V.

⁶ Whether the genetic markers used in forensic DNA analysis are actually independent was an issue that was hotly debated in the early 1990s (Thompson, 1993). A number of prominent scientists suggested that dependencies among the markers might arise from non-random mating patterns—i.e., people’s tendency to mate with other members within subpopulations (Aronson, 2007). The issue was eventually resolved after an extensive program of research revealed relatively little variation in the frequency of the markers across various population groups, suggesting that any dependencies arising from population structuring are likely to have only a modest effect on the frequency of DNA profiles (Kaye, 2010).

⁷ We discuss the strengths and limitations of this very important study in more detail in Section VI.

Conclusions

- Scientific research has convincingly established that the ridge patterns on human fingers vary greatly among individuals. These findings provide a scientific basis for the use of fingerprint comparison to distinguish individuals.
- The scientific literature does not, however, provide an adequate basis for assessing the rarity of any particular feature, or set of features, that might be found in a fingerprint. Examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and there are no scientific criteria for determining when the pool of possible sources is limited to a single person.
- While there is somewhat less variability among relatives than among unrelated individuals, the degree of variability is still sufficient to make fingerprint comparison a useful tool, even for distinguishing monozygotic twins.
- Because the characteristics of fingerprints are unlikely to be statistically independent, it will be difficult to determine the frequency of any particular combinations of features. While research of this type is important, it is unlikely to yield quick answers.
- There have been promising efforts to develop quantitative methods for estimating the probative value or weight of fingerprint evidence. While these methods are not yet ready for courtroom use, research of this type deserves further support and may well offer a long-term solution to the problem of evaluating the probative value of latent print evidence.
- In the short run (i.e., over the next few years), questions about the probative value of latent print evidence will more easily be addressed by research on the performance of latent print examiners under realistic conditions, as discussed below in Section V.

II. Is there an adequate scientific foundation for understanding the degree of variability among prints made by the same finger: (a) on different surfaces, under different environmental conditions; and (b) over time as a person ages or is subject to injury?

When comparing prints, the examiner must consider dissimilarities as well as shared features. That requires the examiner to evaluate whether discrepancies between the prints might have arisen from stretching or contortion of the skin, the condition of the surface on which the print was found, presence of an underlying print impression, or other types of distortion. In other words, the examiner must consider not only the degree of variability that exists across prints from different fingers (inter-finger variability), but also the variability that may exist among prints made by the same finger (intra-finger variability). An examiner who underestimates intra-finger variability may mistakenly exclude an individual as the source of a latent print based on a difference arising from distortion.

There is a body of research that addresses the question of how much variation is possible in prints from the same finger under different circumstances (Cappelli, Maio and Maltoni, 2001; Fagert and Morris, 2015; Kalka and Hicklin, 2014; Maceo, 2009; Sheets et al., 2014). An important source of variability among prints from the same finger is differences in the surface on which the prints are deposited. Distortion can be introduced by curvature, wrinkling, and unevenness of stretching of the item on which the prints are found. However, even when prints are recorded on the same surface under controlled conditions, such as would occur when individuals are fingerprinted for employment or during booking after arrest, there can be significant variability as “...variable factors, such as the elasticity of skin and uneven pressure, mean that there will never be perfect congruence between two prints, even if they originate from the same source” (Expert Working Group, 2012, p. 8).

Maceo (2009) reported that increasing amounts of deposition pressure resulted in more of a given finger contacting the surface, thus increasing both width and length of the resulting latent fingerprint. This is demonstrated in the images below from her research.

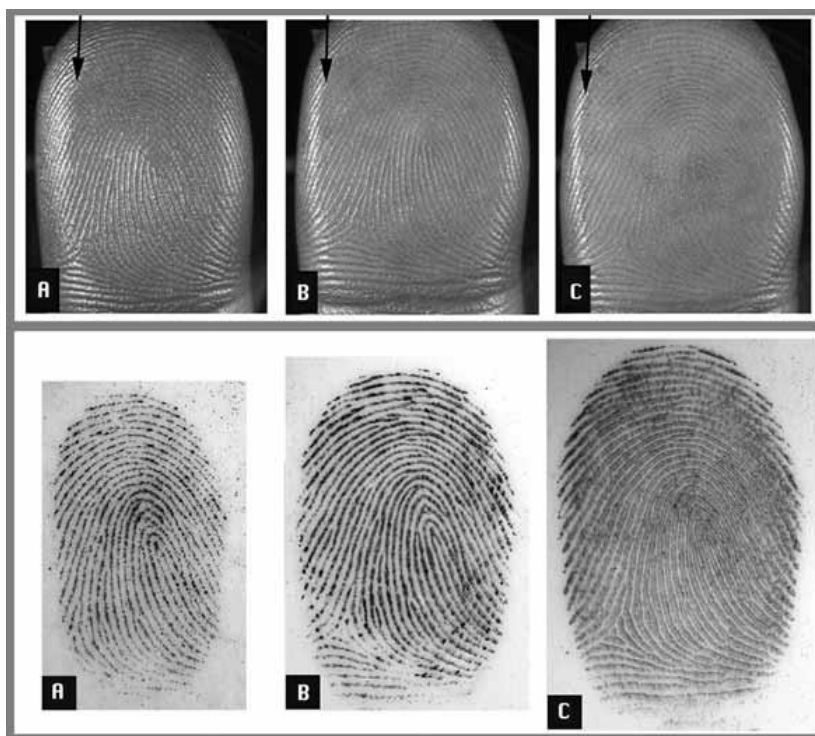


Figure 2. *Finger contact area and resultant latent prints under low (A), medium (B), and high (C) deposition pressure.* (Reprinted with permission from Maceo, A. (2009). *Qualitative Assessment of Skin Deformation: A Pilot Study. Journal of Forensic Identification*, Vol. 59, No. 4, pp 390-440. Figure 13, page 409 in the original article.)

Li and Jain state “...there can be variability due to the placement of the finger on the sensing plane, smudges and partial prints in the latent that is lifted from the crime scene, nonlinear distortion due to the finger skin elasticity, poor quality image due to dryness of the skin, and many other factors” (Li and Jain, 2009). Jain further reports “There are several sources of variability in multiple impressions of fingerprint that lead to detection of spurious minutiae or missing genuine minutiae and deformation of genuine minutiae” (Jain, 2005). Steele suggests that repeated recordings of the same person’s skin under controlled conditions will result in minor differences in appearance due to the amount of deposition pressure, the shape of the surface, and other factors (Steele, 2004). Egli et al. (2007, p. 189) conclude that “[W]ithin finger variability is affected by the visualisation technique used on the mark, the number of minutiae and the minutiae configuration.”

Do latent print examiners understand intra-print variability well enough to determine accurately whether a given difference between print patterns arose from distortion or reflects a difference in the underlying ridge patterns? The existing literature suggests that improvements are needed in this area.

One indication of problems can be seen in “black-box” studies designed to test examiner accuracy (Pacheco, Cerchiai and Stoiloff, 2014; Thompson, Tangen and McCarthy, 2013, 2014; Ulery et al., 2011, 2012). Participants in these studies examined pairs of prints. The researchers knew whether each pair of prints was made by the same finger or different fingers,

but the examiners did not. That allowed the researchers to determine how frequently examiners were mistaken in their source determinations. A key finding of this research was that false exclusions are far more common than false identifications. In other words, examiners are more likely to mistakenly conclude that same-source prints were made by different fingers than to mistakenly conclude that different-source prints were made by the same finger. While there may be several explanations for the higher rate of false exclusions, one possible explanation is that examiners make this error due to a tendency to underestimate the degree of intra-finger variability (Ray and Dechant, 2013). Another finding of the black-box studies is that examiners frequently report that comparisons are “inconclusive”—that is, they decline to venture an opinion on whether the prints have the same source or not. This finding may also arise, in part, from uncertainty about the degree of intra-finger variability.

The most extensive study of this type was conducted by Ulery et al. (2011). They asked 169 experienced latent print examiners to compare latent and exemplar prints that were known to be either from the same finger or different fingers. The test conditions were “designed to correspond to that part of casework in which a single latent is compared to a single exemplar print” and “a large majority of respondents agreed that the fingerprints were representative of (or similar to) casework, and that the overall difficulty was similar to casework” (Ulery et al., 2016, p. 67). The overall false positive rate was low (this aspect of the study will be discussed in more detail in Section V), but the false negative rate was much higher. Eighty-five percent of the examiners made at least one false negative error (mistakenly “excluding” prints known to be from the same finger). These errors occurred after examiners excluded a significant portion of all same-source comparisons because they found the latent print had “no value” for comparison (29.3%) or because they found the results “inconclusive” after comparing the latent with the exemplar (33.5%). In other words, examiners were not required to offer a definitive conclusion and they offered such a conclusion on less than half of the same-source print-pairs they were asked to compare. Only 86% of these definitive conclusions were correct (i.e., the examiner reported that the same-source prints were from the same finger); 14% were incorrect (i.e., the examiner mistakenly “excluded” the latent as having been made by the same finger as the exemplar). In a follow-up study conducted seven months later, 72 of the original examiners were asked to repeat their assessment of the print-pairs. No false positive errors were repeated, but 30% of the false negative errors were repeated. The authors suggest that the errors arose in part from the difficulty of these comparisons and from examiners’ attempts to make categorical decisions in borderline cases (Ulery et al., 2012). It is nevertheless telling that so many errors occurred.

Another indication of problems in examiners’ understanding of intra-finger variability is the existence of debate over what is called the “one-dissimilarity doctrine” (Thornton, 1977; Expert Working Group on Human Factors in Latent Print Analysis, 2012). As explained by the NIST Expert Working Group:

Some examiners discount seemingly different details as long as there are enough similarities between the two prints. Other examiners practice the one-dissimilarity rule,

excluding a print if a single dissimilarity not attributable to perceptible distortion exists.
(p. 7)

That practitioners disagree about such a basic question suggests that the degree of intra-finger variability in prints is not adequately understood.

The existing literature has begun to examine how often fingerprint examiners err, but has not adequately examined the question of why they err (Langenburg, 2009). Langenburg (2009) outlined several reasons why false exclusions may occur: (1) the examiner made an incorrect assessment of anatomical location; (2) the examiner simply did not see the corresponding features for some reason (e.g., distortion); and (3) the examiner gave undue weight to differences between the prints that arose from distortion. Another reason could possibly be: (4) the examiner was misinformed about the circumstances under which the latent was collected. Ulery and colleagues have published a series of additional studies (sometimes called “white-box” studies) to further explore factors that influence examiners’ decisions and the causes of error (Hicklin et al., 2011; Hicklin et al. 2013; Ulery et al., 2013, 2014, 2015, 2016). These studies are a valuable contribution to the literature, but more research is needed.

When erroneous exclusions occur, it would be useful to know the reasons. If examiners underestimate the degree to which prints can vary due to distortion, it would be useful to know whether this occurs because they are inadequately informed of the existing literature on fingerprint distortion or because the literature itself fails to address adequately the kinds of distortion that led to the error, or because of some other reason. Ascertaining the reasons for errors is an essential step toward developing stronger, more accurate methods. We discuss methods for conducting such research in Section V.

Ulery et al. (2016) note that in light of research showing that false exclusions are common “some agencies have begun to change the criteria for an exclusion.” Whether such changes successfully reduce the error rate in practice is another important issue that warrants further research.

Another important question about within-finger variability is the persistence of fingerprints—whether they change over time or as a result of injuries in ways that reduce their value. The existing literature is sufficient to support the claim that fingerprints are persistent over the lifetime of an individual (Barnes, 2011; Maceo, 2005, 2011; Yoon and Jain, 2015). While there can be changes in friction ridge detail due to aging (see e.g., Barros, Faria and Kuckelhaus, 2013; Modi et al., 2007; Stucker et al., 2001), or occupation (see Cutro, 2011), these changes are minor. While there has been no research on whether such changes degrade the accuracy of latent print examinations, it seems unlikely that accuracy would be affected to any significant degree.

Injuries may cause more significant alteration (or obliteration) of friction ridge skin, although alterations of this type are often detectable (Haylock, 1987; Olsen, 1978). Indeed, there are automated systems that can detect these alterations (Yoon, Feng, and Jain, 2012). These kinds

of changes are unlikely to produce erroneous identifications, but may reduce the amount of information available from a print impression and thus reduce the chances of identifying an individual. Erroneous exclusions are possible because some injuries can significantly affect the overall ridge flow, in extreme cases producing the appearance of a different pattern entirely, as indicated in Figure 3 below. These impressions were made by the same source, but the damage evident in the impression on the right has adversely affected the general appearance of the original friction ridge information.



Figure 3. Source: FBI Law Enforcement Bulletin; accessed through clpex.com

Conclusions

- There is a body of research that addresses the question of how much variation is possible in prints from the same finger under different circumstances. But there is also evidence that latent print examiners sometimes misjudge the degree of intra-finger variability that may occur, creating a risk of false exclusions. It is unclear whether this problem occurs because examiners fail to appreciate the implications of existing research on intra-finger variability, or because existing research is incomplete.
- Existing research provides adequate support for the claim that, barring injury or purposeful obliteration of friction ridges, there is little variation in fingerprints over time. While there can be minor changes due to aging or occupation, these changes are unlikely to undermine the accuracy of fingerprint identifications to any significant degree.

III. Is there an adequate scientific foundation for understanding the accuracy of automated fingerprint identification systems (AFIS)?

AFIS were developed in the late 1970s to automate fingerprint searches, whether for background checks (ten-print to ten-print comparison) or for identifying possible suspects based on crime scene evidence (latent print to ten-print comparison). The growth of databases of known prints (ranging from a few million subjects for state law enforcement agencies to over a hundred million subjects for national agencies) has made AFIS an essential law enforcement tool.

AFIS are available from several competing commercial vendors and have been purchased by nearly every law enforcement agency and many crime laboratories. State-of-the art AFIS can quickly compare an input print to millions of reference prints in a database in order to identify the reference prints with the most similar ridge patterns. When AFIS are used to search sets of 10 rolled or plain fingerprints against a reference database, virtually all comparisons are fully automatic and identifications can be made with little or no human intervention. By contrast, when AFIS are used to search a latent print against a database, the system typically generates a list of possible candidates, ranked by similarity to the latent, which then must be carefully reviewed by a human examiner. A second examiner is typically asked to verify identifications made by the first examiner.

The major AFIS programs were developed by commercial firms and the details of how these systems work have been kept secret for proprietary reasons. While commercial incentives undoubtedly spurred development of these systems, the resulting secrecy, as well as lack of access to AFIS, makes it difficult for the larger scientific community to evaluate their performance. NIST plays an important role in understanding of AFIS performance by conducting periodic evaluations of these systems (<http://www.nist.gov/itl/iad/ig/fingerprint.cfm>; accessed September 12, 2017). The NIST studies assessed their performance as “black box” evaluations, without knowing the technical details of how each AFIS works. Without knowing the details of the feature extraction and comparison algorithms the AFIS deploys, it is not possible to explain why an AFIS fails to find the source of a latent. Further, it is difficult to determine when and why they fail. Moreover, the secrecy surrounding the internal details of AFIS makes it difficult for the larger scientific community to play a role in system improvement.

Some of the NIST evaluations examined AFIS accuracy for ten-print to ten-print identification; other evaluations examined AFIS accuracy for latent print identification. In order to evaluate AFIS performance when comparing images on ten-print cards to the ten-print images in a database, NIST has conducted Fingerprint Vendor Technology Evaluations (FpVTE). At the first FpVTE in 2003 a total of 18 different companies submitted 34 different systems, allowing the most comprehensive evaluation of AFIS to that point (Wilson et al., 2004). NIST conducted another FpVTE in 2012, which evaluated AFIS performance using operational datasets containing several million subjects. In the 2012 NIST evaluation, state-of-the-art AFIS achieved false negative identification rates (FNIR) of 1.9% for single index fingers, and 0.27% for two

index fingers, at a false positive identification rate (FPIR) of one in 1000 (10^{-3}) (Watson, et al. 2014). (Note that the FPIR of 10^{-3} was chosen to allow a comparison of different vendors at a common threshold. The operational systems base their comparisons on all ten fingers and hence are expected to perform at much lower FPIR in practice). The NIST research has established that AFIS are highly accurate for ten-print to ten-print comparison. However, there is general agreement among practitioners that, at present for latent fingerprint comparisons, “automatic fingerprint-matching algorithms are significantly less accurate than a well-trained forensic expert” (Moses, 2011, at 6.4.5), although there has been no direct empirical test of this assumption.

NIST held a workshop (see Dvornychenko and Garris, 2006) to lay the foundations for testing AFIS performance on the more difficult task of comparing latent prints to the ten-print images in databases. Before testing could commence, suitable test sets had to be identified and prepared, agreements had to be reached on programming interfaces, and methods had to be developed for scoring performance. In 2010, NIST conducted its first evaluation of AFIS accuracy for latent prints, called Evaluation of Latent Fingerprint Technologies-Extended Feature Sets (ELFT-EFS Evaluation #1) (Indovina et al., 2011). It examined the accuracy of AFIS when experienced latent fingerprint examiners manually marked “extended features” of the latent prints in addition to the features that were identified automatically by the system. The extended features included region of interest (ROI), minutiae, ridge count, quality map, pattern class, and skeleton. Five participants (Sagem, NEC, Cogent, Sonda and Warwick) submitted their systems. The highest rank-1 identification accuracy (i.e., proportion of latent queries for which the corresponding mated fingerprint in the reference database had the highest comparison score) was only 66.7% in matching 1,114 latent fingerprint images against 100,000 exemplar prints. A second evaluation (ELFT-EFS Evaluation #2; Indovina et al., 2012) repeated the same tests using updated matchers provided by the five vendors mentioned above. The evaluation results showed measurable performance improvements, particularly when the input to the matchers included both latent print images and manually marked Extended Features (EFS).

The NIST studies indicate that the major factor affecting the accuracy of AFIS is fingerprint quality (<http://www.nist.gov/itl/iad/ig/fingerprint.cfm>; accessed September 12, 2017). Hence, the quality of the prints in question must be taken into account when assessing the accuracy of a particular AFIS output. NIST has developed a quantitative measure of quality called NFIQ for ten-prints of a known source. This tool has proven useful in predicting/assessing AFIS accuracy (http://www.nist.gov/itl/iad/ig/bio_quality.cfm; accessed September 12, 2017), but there is not, at present, an adequate quantitative measure for the quality of latent prints. While some attempts have been made to develop such a measure (Hicklin, Buscaglia and Roberts, 2013; Kellman et al., 2014; Yoon et al., 2013;), more work is needed in this area.⁸ As discussed in the

⁸ As the PCAST report noted (PCAST, 2016, p. 97, n. 288), promising work in this area has been done by Hicklin et al. (2013), who developed what they call the Latent Quality Assessment Software (LQAS), a tool for evaluating the clarity of prints; and by researchers at the University of Lausanne, who are developing a quality metric and statistical assessment tool for latent prints that they call the Picture Annotation System (PiAnoS)(see, <https://ips-labs.unil.ch/pianos/>, last visited September 12, 2017).

following sections, a quantitative measure of latent print quality could be helpful for evaluating the accuracy of human examiners as well as AFIS. Hence, additional research in this area is clearly warranted.

The three best performing latent-print AFIS in the NIST evaluations of latent print technologies are marketed by Cogent, NEC and Morpho. The 2013 NIST evaluations further showed that a score-level fusion of the two best performing AFIS, with comparable accuracies, leads to an overall performance improvement. This indicates that different AFIS tend to extract somewhat complementary information from fingerprint images, which suggests that each system has room for improvement. Note that different AFIS may fail due to different reasons. Because the internal representations (feature extraction) and matching strategies used by these systems are not in the public domain for proprietary reasons, however, it is difficult to determine why an AFIS fails, and difficult for the broader scientific community to play a role in system improvement.

The use of different latent AFIS by different agencies may also create interoperability issues.⁹ Jain and Feng (2011) note that matching accuracy can decline when examiners extract features (encode prints) using one AFIS and then submit them to another AFIS for search against a database. The search algorithms of a given AFIS often make use of additional features, beyond the standard minutiae, that are encoded only by that AFIS and are not available in data encoded by other systems. Jain and Feng (2011) suggest that interoperability could be improved by further efforts to standardize the templates used for extracting and encoding information from latent prints.

The performance of AFIS could undoubtedly be improved through further research. Topics worth investigating include new methods for enhancing latent images or generating novel templates from latent prints, and robust matching techniques in the presence of large distortion. Promising efforts to improve AFIS performance through a synergy between forensic examiners (who perform multiple manual markups of the latents) and automatic AFIS (which perform automatic markup and comparison) have recently been reported (Arora et al., 2015). Additional research on the accuracy of automated systems for latent print attributions (e.g., Cole et al., 2008) would also be useful.

In the academic community, Jain and Feng (2011) proposed to use manually marked features, including minutiae, singularity, ridge quality map, ridge flow map, ridge wavelength map, and skeleton for latent matching. The experimental results by matching 258 latent prints in the NIST SD27 database against 29,257 rolled prints show that a minutiae-based baseline for the rank-1 identification accuracy of 34.9% improved to 74% when extended features were used. These experiments investigated the accuracy levels that could be achieved on a publicly available latent database using extended features. Jain (2011) conducted additional experiments on two

⁹ https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/afis_10-20-2014_draftforcomment.pdf

latent databases: NIST SD27 and ELFT-EFS-PC. The following findings were reported: (1) almost all the extended features led to some improvement in latent matching accuracy; (2) level 1 features (including ridge flow, and ridge wavelength) are more effective in improving latent matching accuracy than level 2 features (minutiae) and level 3 features (including pores and ridge shape, etc.); and (3) high image resolution (at least 1000 ppi) images are necessary but not sufficient for reliably extracting level 3 features.

There is general agreement that latent matching accuracy can be improved using extended features, which are typically manually marked for latents. However, marking extended features (orientation field, ridge skeleton, etc.) in poor quality latents is very time-consuming and might be feasible only in rare cases. Paulino et al. (2013) propose using minimal manual input in the form of manually marked minutiae. Experimental results on two different latent databases (NIST SD27 and WVU latent databases) show that Paulino's algorithm outperformed two commercial fingerprint matchers. Further, a fusion of their proposed algorithm and commercial fingerprint matchers leads to further improvements in matching accuracy.

The features marked by examiners are not always compatible with those automatically extracted by an AFIS. In an effort to improve the AFIS accuracy, Paulino et al. (2010) combined manually marked (ground truth) minutiae with automatically extracted minutiae from an AFIS. Experimental results on NIST SD27 database demonstrate the effectiveness of the proposed fusion strategy. Based on the observation that features marked by different examiners may be different, Arora et al. (2015) propose a synergistic crowd powered latent identification framework, where multiple latent examiners and an AFIS work in conjunction with each other to boost the identification accuracy of AFIS. Experimental results show that the fusion of an AFIS with examiner markups improves the rank-1 identification accuracy of the AFIS by 7.75% (using six markups) on the 500 PPI NIST SD27, 11.37% (using two markups) on the 1000 PPI ELFT-EFS public challenge database, and by 2.5% (using a single markup) on the 1000 PPI RS&A database against 250,000 rolled prints in the reference database.

While latent matching has gained wide attention, latent image preprocessing is still a relatively new domain in need of further scientific study and development of best practice guidance. Guan et al. (2014) propose to extend Spectral Image Validation and Verification (SIVV) to serve as a metric for latent fingerprint image quality measurement. Results in this paper show that the new metric can provide positive indications of both latent fingerprint image quality and the performance of fingerprint preprocessing.

One limitation faced by researchers in this area is the relatively small size of available datasets containing multiple prints of varying quality from the same individuals. The PCAST report noted this problem and proposed a solution:

The most important resource to propel the development of objective methods would be the creation of huge databases containing known prints, each with many corresponding "simulated" latent prints of varying qualities and completeness, which would be made available to scientifically-trained researchers in academia and industry. The simulated

latent prints could be created by “morphing” the known prints, based on transformations derived from collections of actual latent print-record print pairs. (PCAST, 2016, p. 103).

While the idea of developing a large database of known prints has merit, the task is likely to be challenging. Past attempts to synthesize latent fingerprint images have not been successful (Feng et al., 2012; Hong et al., 2015) and efforts to collect real latent prints of varying quality from many known individuals are likely to be time-consuming and expensive. We agree, however, that a database of this kind would be an important resource for research and deserves to be funded.

The PCAST report called for continuing efforts to improve AFIS, with the goal of converting latent print analysis “from a subjective method to an objective method” (p. 103). It pointed to a number of advantages of using automated systems rather than human examiners to compare and draw conclusions from fingerprints: automated systems apply decision criteria consistently and rapidly; they are not subject to cognitive bias; and because they operate so rapidly, it is possible to test their performance using larger datasets than would be possible with human examiners.

We agree that efforts to automate latent print examination have the potential to yield major benefits, but we caution that these systems are not, at present, designed to accomplish the task that PCAST contemplates. At present, latent AFIS are designed to rapidly search a database in order to identify the reference prints with features most similar to an input print. They are not designed to determine whether any particular reference was, or was not, made by the same finger as the input print, nor are they designed to assess the weight (probative value) of a fingerprint comparison for proving that any particular pair of prints has a common source.

It is certainly possible, however, that latent AFIS could evolve over time in a manner that would allow them to be tested in the manner PCAST contemplates. For example, latent AFIS might be programmed to produce a decision about each comparison made between two prints—e.g., that the prints have the same source, a different source, or that the comparison is inconclusive. Research could then assess the true and false positive rate of the systems when evaluating prints of known source under various conditions (e.g., with varying print quality). Alternatively, the systems could be programmed to produce a quantitative estimate of the weight-of-evidence, such as a likelihood ratio, to indicate the degree of support that the comparison provides for the conclusion that the prints have the same (or different) source.¹⁰ Methods for

¹⁰ A number of scholars have argued that forensic scientists should strive to characterize their findings using quantitative strength-of-evidence statements (e.g., likelihood ratios) rather than categorical conclusions (e.g., “match,” inclusion,” “identification,” “exclusion” (Morrison and Stoehl, 2014; Robertson, Viagnaux and Berger, 2016). The PCAST report has been criticized for assuming that forensic scientists will report categorical conclusions rather than evolving toward strength-of-evidence statements based on statistical models (Morrison et al., 2016). In our view, the best way for latent print examiners to report their findings will depend partly on which type of reports will be most readily and accurately understood by the intended audience. As discussed in Section VI, how

assessing the validity of such quantitative statements have been discussed in the forensic science literature (Morrison, 2011) and could be applied to the assessment of quantitative systems for latent print identification. It will, of course, be important that the systems be validated under conditions comparable to those used in forensic practice and for prints comparable in quality to those compared in forensic cases (Morrison and Stoehl, 2014).

Conclusions

- AFIS are designed to rapidly search fingerprint databases in order to identify the reference prints with features most similar to an input print. Research has established that AFIS are highly accurate for “ten-print to ten-print identification” (i.e., comparison of rolled or slapped prints of all ten fingers in order to identify sets from the same person).
- While AFIS are considerably less accurate for latent print comparison than for exemplar ten-print comparison (and are thought to be less accurate than human examiners for that purpose) they are necessary and extremely useful for screening large numbers of reference prints to identify possible sources of a latent (which can then be evaluated by human examiners).
- The commercial firms that have developed and marketed AFIS have kept the details of how they work secret for proprietary reasons. Consequently, members of the larger scientific community have found it difficult to assess the probative value of specific AFIS results for identifying particular individuals; and difficult to assess how and why these systems might fail.
- The existence of competing commercial AFIS has also created problems of interoperability across law enforcement agencies that need to be addressed.
- Latent AFIS are not, at present, designed to determine whether any particular reference print was, or was not, made by the same finger as an input print; nor are they designed to assess the weight (probative value) of a fingerprint comparison for proving that any particular pair of prints has a common source. But it is possible that AFIS could evolve over time to perform those functions.

lay people, such as lawyers and jurors, understand various reporting statements is a question that warrants further research.

IV. Is there an adequate scientific foundation for understanding the potential for contextual bias in latent print analysis and how might it be addressed?

Contextual bias is said to occur when a judgment or decision is influenced by irrelevant or inappropriate information that the decision maker gleans from the surrounding context. The 2009 NRC Report concluded that “forensic science experts are vulnerable to cognitive and contextual bias” (p. 4, note 8). This concern arose in part from empirical studies showing that forensic scientists are sometimes influenced by information that is irrelevant to their scientific assessments. For example, latent print examiners were less likely to report a match between a latent print from a crime scene and a suspect when they were told the suspect had a solid alibi (Dror and Charlton, 2006; Dror, Charlton and Peron, 2006; Dror and Rosenthal, 2008; see generally, Expert Working Group on Human Factors in Latent Print Analysis, 2012). Furthermore, the high-profile error in the Mayfield case¹¹ has been attributed, in part, to a form of contextual bias (Office of Inspector General, 2006; Stacey, 2004).

Contextual bias is not limited to forensic scientists. It is a universal phenomenon that affects decision making by people from all walks of life and in all professional settings, including science (Kassin, Dror and Kukucka, 2013; Risinger et al., 2002). People are particularly vulnerable to contextual bias when making judgments in the absence of objective standards based on data that may be somewhat ambiguous and subject to differing interpretations. Contextual bias occurs without conscious awareness; it does not indicate misconduct or bad intent. Rather, exposure to contextual information can bias the conclusions of experts who perform their jobs with utmost honesty and professional commitment.

Latent print examiners often need to evaluate data that are somewhat ambiguous and subject to differing interpretations. There are no objective standards for them to apply when making critical decisions about what conclusion to report. Under these circumstances, there is clearly a potential for examiners’ judgments to be influenced by contextual bias (Dror, 2011; Expert Working Group on Human Factors in Latent Print Analysis, 2012).

One kind of contextual bias may arise from examiners’ exposure to investigative facts from the underlying criminal case. A second kind of bias may arise from examiners’ being exposed to a suspect’s fingerprints before they analyze a latent (Dror et al., 2015). Knowledge of the features of the suspect’s prints may unconsciously influence the way examiners interpret the latent. The error in the Mayfield case has been attributed, in part, to this second kind of bias: “Having found as many as 10 points of unusual similarity, the FBI examiners began to ‘find’

¹¹ Brandon Mayfield, a lawyer from Portland Oregon, was mistakenly linked to a terrorist incident (the 2004 Madrid train bombing) when three latent print examiners employed by the Federal Bureau of Investigation, and an independent examiner hired by Mayfield himself, all concluded that Mayfield was the source of a latent print found on a plastic bag containing detonators used by the terrorist bombers. Latent print examiners were improperly influenced by the fact that he was a Muslim. Mayfield was arrested and held for two weeks before investigators discovered that a man named Ouhmane Daoud, who was already suspected of involvement with the incident, matched the latent print better than Mayfield. Mayfield received an apology and a financial settlement from the government (Office of Inspector General, 2006).

additional features in LFP 17 [the latent print] that were not really there, but rather suggested to the examiners by features in the Mayfield prints” (Office of Inspector General, 2006, p. 6).

The most effective way to minimize contextual bias is to avoid exposing examiners to contextual information that is unnecessary for a scientific assessment of the prints and to delay exposing examiners to information that, while necessary, may create bias if presented prematurely. *Context management procedures* (sometimes called blinding procedures) are used in many areas of science in order to prevent bias. Commentators have made a number of excellent suggestions for how to implement context management procedures in forensic science generally (Cole, 2013a; Dror and Cole, 2010; Found and Ganas, 2013; Stoel et al., 2014; Thompson, 2011), and in latent print analysis specifically (Cole, 2013b; Dror et al., 2015; Expert Working Group on Human Factors in Latent Print Analysis, 2012). A number of forensic laboratories have already adopted such procedures for latent print analysis, including the FBI laboratory (Office of the Inspector General, Department of Justice, 2011).

However, these procedures have not yet been widely adopted by forensic laboratories (NRC, 2009). At present, examiners in many forensic laboratories are part of a law enforcement team and communicate directly with detectives or other investigators; they may also have access to police reports, suspects’ rap sheets, and other information about the case. Information of this type is not needed to perform a fingerprint comparison and has the potential to create bias. Consequently, forensic laboratories should take steps to avoid exposing examiners to such information, at least until after the examinations are complete and the conclusions are recorded.

The National Commission on Forensic Science recently adopted a “views document” that expressed the Commission’s view that:

1. Forensic science service providers should rely solely on task-relevant information when performing forensic analyses.
2. The standards and guidelines for forensic practice being developed by the Organization of Scientific Area Committees (OSAC) should specify what types of information are task-relevant and task-irrelevant for common forensic tasks.
3. Forensic laboratories should take appropriate steps to avoid exposing analysts to task-irrelevant information through the use of context management procedures detailed in written policies and protocols.¹²

These suggestions should be followed by latent print examiners. Context management procedures should be adopted by all forensic laboratories, including those laboratories that are part of law enforcement agencies. These procedures are necessary to protect the scientific integrity of latent print evidence. Where these procedures are not adopted, there is a risk that

¹² National Commission on Forensic Science. Ensuring That Forensic Analysis is Based Upon Task-Relevant Information (Views document adopted December 8, 2015) available at <https://www.justice.gov/ncfs/file/818196/download> (accessed September 12, 2017).

latent print evidence will be influenced (tainted) by other evidence or information that is irrelevant to a scientific assessment of the prints, thereby rendering latent print evidence less valuable for fact-finders in the criminal justice system (Thompson, 2016). Context management procedures should be a key element of the standards for latent print analysis developed by standards-setting organizations like the Organization of Scientific Area Committees (OSAC). Failure to control for contextual bias (when it is possible to do so) is unacceptable in the broader scientific community. It should be unacceptable in forensic science as well.

Scientists have long recognized that the results of an observation can be affected by the state of the observer. As early as 1795, astronomers recognized that “observer effects” can distort and undermine the accuracy of experts’ observations (Risinger et al., 2002). Where the observers’ preconceptions or motives influence the interpretation of data, the phenomenon is sometimes called examiner bias or confirmation bias, although it is important to note that the “bias” entailed in the phenomenon may occur without the observer intending or even being aware of it (Thompson, 2009a).

The term “context effect”, sometimes used synonymously with “observer effect”, originated in psychology. It was initially used to describe circumstances in which the perception of a stimulus is affected by the surrounding context, as where a gray object looks lighter against a dark than a light background (Dresp-Langley and Reeves, 2011). In forensic science, however, the term “context effect” has been used more broadly to describe situations in which the results of a forensic analysis are affected by the context in which it is performed, and particularly by the contextual information available to the analyst. In one study, for example, latent print examiners become less likely to identify a latent print as that of the suspect when told that the suspect had a strong alibi (Dror and Charlton, 2006; Dror, Charlton and Peron, 2006). A context effect becomes contextual bias when the influence of context is deemed improper or inappropriate, as when scientific conclusions are affected by contextual information that is irrelevant to the scientific judgment (e.g., the suspect’s alibi).

There is growing evidence that forensic scientists are susceptible to contextual bias (see, Kassin et al., 2013 for a review). Evidence of contextual bias has been found in latent print analysis (Dror, Charlton and Peron, 2006; Dror and Rosenthal, 2008), document examination (Miller, 1984), bite mark analysis (Osborne et al., 2014), bloodstain pattern analysis (Laber et al., 2014), forensic anthropology (Nakhaeizadeh, Dror and Morgan, 2014); crime scene investigation (Helsloot and Groenendaal, 2011), and DNA analysis (Dror and Hampikian, 2011; Thompson, 2009b). The concerns have also been reinforced by the discovery of errors that have been attributed, at least in part, to contextual bias in latent print analysis (Office of Inspector General, 2006; Stacey, 2004), bite mark analysis (Pretty and Sweet, 2010) and DNA testing (Thompson, 2008, 2013). In its 2009 report on forensic science, the National Research Council acknowledged these concerns and agreed that they are a problem for the field, declaring unequivocally that “forensic science experts are vulnerable to cognitive and contextual bias” and that this bias “renders experts vulnerable to making erroneous identifications” (NRC, 2009, p. 4, note 8).

These conclusions are consistent with a large psychological literature showing that human beings are susceptible to cognitive and contextual bias, that people can be biased without being aware of it, and that even well-trained experts are susceptible to bias (for reviews of this literature, see Nickerson, 1998; Risinger et al., 2002; Saks et al., 2003; Thompson, 2009a; Kassin et al., 2013). In its 2012 report, the NIST Expert Working Group on Human Factors in Latent Print Analysis offered extensive discussion of this literature, concluding that “[c]ognitive scientists and psychologists have observed cognitive bias in hundreds of scientific studies across dozens of domains” and that “research has also demonstrated its existence in latent print examination” (Expert Working Group, 2012, p. 11, Box 1.2). As the NIST report explained, contextual bias can arise without conscious awareness and can influence “people acting in good faith and attempting to be fair interpreters of the evidence” (p. 10). Nor does scientific training make one immune to bias. Contextual bias has been found among highly trained experts, such as radiographers (Schreiber, 1963; Tape and Panzer, 1986). Indeed, psychologists have argued that the nature of expertise may render experts *particularly* vulnerable to bias (Dror, 2011). Thus, “[t]o recognize that latent print examiners are potentially subject to bias is not to single them out but rather to suggest that they are not exempt from those cognitive biases that all interpreters of data and information face” (Expert Working Group, 2012, p. 40).

Concerns about contextual bias have been raised in a variety of scientific fields (Risinger et al. 2002). The most common way to address such concerns is to adopt blind or double-blind methods that shield the person interpreting critical data from extraneous information that might improperly influence the interpretation (Levine, 1986; Risinger et al., 2002; Shultz and Grimes, 2002; Thompson, 2011; 2009a) or (when impractical to shield the examiners) by conducting blind verifications after the initial examination using different examiners. Blind procedures are most common in fields where practitioners must rely on subjective judgment to interpret data, such as medicine and psychology; they are seen less often in research in the physical sciences, perhaps because data in those areas is viewed as more objective and less subject to human interpretation (Sheldrake, 1999). But blind procedures are widely used in all fields for peer-review of scientific articles, for grading of written examinations, and for other functions for which contextual bias is a concern. These procedures avoid contextual bias by the straightforward expedient of preventing exposure to potentially biasing information.

Forensic scientists have lagged behind other scientific fields in addressing contextual bias. In its 2009 report, the NRC noted that “[t]he forensic science disciplines are just beginning to become aware of contextual bias and the dangers it poses” (p. 185). The NIST Expert Working Group attributed the problem partly to ignorance and partly to the incorrect belief that conscientious experts can correct for their own biases:

Within the forensic science community, some people still lack an understanding of what bias is and how best to address it. Often, cognitive bias is treated as an ethical issue or as an issue that will resolve once someone is aware of the problem. However, the cognitive process used when gaining experience (e.g., using schemas, chunking information, automaticity, and more reliance on top-down information) in itself opens the practitioner to vulnerabilities, including bias, tunnel vision, lack of flexibility, and

selective attention. Cognitive bias results from computational trade-offs carried out in the brain and is not a conscious act or an act that can be avoided at will (Expert Working Group, p. 11, Box 1.2).

A few commentators have argued that forensic scientists have the capability of overcoming any biases they might have through professionalism, good will, and honest intentions (Leadbetter, 2007; Thornton, 2010). While these claims are undoubtedly sincere, they are not credible in light of the literature reviewed here. Psychologists have shown that people have a “blind spot” when it comes to recognizing their own biases (Pronin and Kugler, 2007; Pronin, Gilovich and Ross, 2004; Pronin, Yin and Ross, 2002), and that very blind spot may explain these claims. We all have difficulty identifying and correcting for bias because of a basic limitation of the human mental process: we have little insight into whether and how much our judgments have been influenced by particular facts or information to which we are exposed. We cannot rely on introspection to tell us whether—or how much—we have been influenced by any particular fact or factor. Hence, we cannot trust anyone’s claim that a particular fact or factor had no influence on their judgment, at least not when the claim is based solely on introspection (Nisbett and Wilson, 1977; Wilson and Brekke, 1994).

Because people cannot recognize the influence of contextual information, they cannot be counted on to correct for it, no matter how ethical or professional they are. Exposing a forensic examiner to information irrelevant to the scientific task creates the potential for what Wilson and Brekke (1994) have called *mental contamination*. They explain their use of the metaphor of contamination as follows:

Something that is contaminated is not easily made pure again, which we believe is an apt metaphor for many mental biases. We argue that, because of a lack of awareness of mental processes, the limitations of mental control, and the difficulty of detecting bias, it is often very difficult to avoid or undo mental contamination (p. 117).

Wilson and Brekke identified four conditions that must be satisfied in order for a person to avoid biasing effects of mental contamination. First, the person must be *aware* of any unwanted mental processes invoked by the contaminating information. Second, the person must be *motivated* to correct for the effects of the mental contamination. Third, the person must be aware of the *direction and magnitude* of the bias created by the contaminating information. And fourth, the person must have *sufficient control* over their responses to the contaminating information to be able to make the correction. Because these conditions are virtually impossible to satisfy, Wilson and Brekke are “rather pessimistic about people’s ability to avoid or correct for mental contamination” (p. 120). The best way to avoid bias is to prevent mental contamination in the first place by avoiding exposure to potentially contaminating information through blinding procedures.

Blinding procedures pose some practical difficulties in forensic science because contextual information about the underlying case is often necessary to allow the forensic laboratory to function effectively (see e.g., Budowle et.al., 2009; Butt, 2013; Charlton, 2013; Ostrum, 2009;

Wells, 2009). For example, in order to make efficient use of laboratory resources, when investigators collect a large number of prints at a crime scene, latent print examiners may need to consult with them about which prints are most likely to be pertinent to the investigation so that those prints can be examined first. After laboratory analyses are completed, laboratory personnel may need information about the underlying case to help police and lawyers make sense of various laboratory findings. How can forensic scientists be “blind” to context when they need contextual information to do their jobs?

As a number of commentators have noted, this dilemma can be resolved by separating functions in the laboratory, allowing some individuals to be aware of context while others are “blind.” In the “case manager model” (Found and Ganas, 2013; Stoel et al., 2014; Thompson, 2011), forensic scientists can serve either as case managers or analysts. The role of the case manager is to communicate with investigators, participate in decisions about what specimens to collect at crime scenes and what examinations are needed, and to manage the flow of work in the laboratory. In contrast, the role of the analyst is to perform analytic tests and comparisons on specimens submitted to the laboratory in accordance with the instructions of the case managers. This separation of functions allows case managers to be fully informed of the investigative context while analysts remain blind to investigative facts that are unnecessary for the analyses they are asked to perform. The risk of the case manager approach is that the case manager may not always recognize why a given piece of information is relevant, thereby affecting the outcome.

In comparative disciplines like latent print analysis, there is also potential for a form of confirmation bias that occurs when interpretation of an evidentiary sample (e.g., a latent print from a crime scene) is inadvertently influenced by knowing the characteristics of reference samples to which it is compared (Dror et al., 2015). To minimize this kind of bias, commentators have proposed that workflow in the laboratory be sequenced, so that the more difficult evidentiary traces are examined (and results recorded) before the reference samples are examined. This sequential procedure (sometimes called “sequential unmasking”) was initially proposed for comparison of DNA profiles (Krane et al., 2008; Stoel et al., 2014), but the FBI laboratory has adopted a similar procedure for latent print examination. Called “linear ACE-V,” the FBI’s procedure involves temporary masking of reference prints while analysts make and record their initial assessments of the evidentiary prints (Office of the Inspector General, Department of Justice, 2011; but see Cole, 2013, who notes that details of the FBI’s protocol are not yet public).

The NIST Expert Working Group endorsed blinding procedures for latent print examination, declaring that “[P]rocedures should be implemented to protect examiners from exposure to extraneous (domain-irrelevant) information in a case” (Recommendation 3.3, p. 44). It also endorsed the linear ACE-V procedure in which latent prints are examined and key features

identified before the examiner is exposed to any reference prints.¹³ The Working Group explained this recommendation as follows:

This recommendation is not suggesting that examiners should be denied access to information that is legitimately relevant to their substantive analysis. Examiners must have the information that is necessary to do their jobs effectively. Some information will be both potentially biasing and domain relevant—for example, information about the substrate from which a print was lifted, or the fact that the source print is the result of an AFIS search.... However, given the genuine dangers of cognitive bias, the better practice is to protect examiners from inadvertent bias by shielding them from information that is clearly unnecessary and not relevant to their assessment (p. 44).

The National Commission on Forensic Sciences has endorsed blinding procedures for all forensic comparative disciplines, saying “[t]he most effective way to address the challenge of contextual bias is to avoid exposing analysts to task- irrelevant information.” The Commission called on forensic laboratories “to avoid exposing analysts to task-irrelevant information through the use of context management procedures detailed in written policies and protocols (Commission Views document adopted December 8, 2015, *Ensuring That Forensic Analysis is Based Upon Task-Relevant Information*, p. 1).

According to the National Commission, information is *task-relevant* for analytic tasks such as latent print comparison if it is necessary for drawing conclusions:

- (i.) about the propositions in question,
- (ii.) from the physical evidence that has been designated for examination,
- (iii.) through the correct application of an accepted analytic method by a competent analyst.

The Commission gave the following examples of information that would be “task-irrelevant” (i.e., not task-relevant) for a latent print examiner: information about the suspect’s criminal history; information that the suspect confessed to the crime; information that the suspect was implicated by other physical evidence at the crime scene (e.g., DNA evidence); and information

¹³ As a further safeguard against confirmatory bias arising from the examiner’s exposure to reference prints, the Expert Working Group also made the following recommendation:

Recommendation 3.2: Modifications to the results at any stage of latent print analysis (e.g., feature selection, utility assessment, discrepancy interpretation) after seeing a known exemplar should be viewed with caution. Such modifications should be specifically documented as having occurred after comparison had begun (p.43).

The PCAST report makes similar recommendations (PCAST, 2016, p. 10):

Work by FBI scientists has shown that examiners often alter the features that they initially mark in a latent print based on comparison with an apparently matching exemplar. Such circular reasoning introduces a serious risk of confirmation bias. Examiners should be required to complete and document their analysis of a latent fingerprint *before* looking at any known fingerprint and should separately document any additional data used during their comparison and evaluation.

that another latent print examiner identified the suspect as the source of a print found on a different item at the same crime scene (National Commission on Forensic Science, *Ensuring That Forensic Analysis is Based Upon Task-Relevant Information*, Dec 8, 2015).

The literature reviewed in connection with the current report provides strong support for the recommendations of the NIST Working Group and the National Commission on Forensic Science with regard managing contextual bias. These proposals are sound and should be implemented. Context management procedures should be adopted by all forensic laboratories, including those laboratories that are part of law enforcement agencies. These procedures are necessary to protect the scientific integrity of latent print evidence. Where these procedures are not adopted, there is a risk that latent print evidence will be influenced (tainted) by other evidence or information that is irrelevant to a scientific assessment of the prints, thereby rendering latent print evidence less valuable for fact-finders in the criminal justice system (Thompson, 2016). Context management procedures should be a key element of the standards for latent print examination developed by standards-setting organizations like the OSACs. Failure to control for contextual bias (when it is possible to do so) is unacceptable in the broader scientific community. It should be unacceptable in forensic science as well.

Forensic laboratories with very limited staffing (e.g., a single latent print examiner) might find it difficult to adopt context management procedures. For small laboratories with limited resources, it might nevertheless be possible to mitigate contextual bias through cooperative arrangements with other laboratories, perhaps on a regional basis. For example, examiners in one laboratory might handle case-management functions or perform verifications for examiners in the other laboratory. Laboratories that lack the resources to adopt reasonable context management procedures will not be able to perform at the same level of scientific rigor as other laboratories, and this limitation should be acknowledged.

Conclusions

- Studies have shown that latent print examiners (like all human beings) are vulnerable to contextual bias. Their evaluations of the features of latent prints can be influenced inappropriately by premature exposure to reference prints; their conclusions can also be influenced inappropriately by information about the underlying criminal investigation. Contextual bias of this type is unlikely to change examiners' opinions in clear-cut cases, but may have stronger effects in ambiguous cases where the prints are more difficult to evaluate.
- Contextual bias can happen without an examiner's conscious awareness and, therefore, cannot be reliably suppressed or corrected by the individual.
- Contextual bias can be mitigated through the use of context management procedures that avoid exposing examiners to contextual information that is unnecessary for a scientific assessment of the prints and delay exposing examiners to information that, while necessary, may create bias if presented prematurely. A number of laboratories have adopted context management procedures as a means of mitigating contextual bias in latent print analysis.

V. Is there an adequate scientific foundation for understanding the accuracy of human fingerprint examiners and how their accuracy is affected by: (a) level of training and experience; (b) individual differences in perceptual ability; (c) analytic procedures and standards of practice; (d) quality control and quality assurance procedures; and (e) the quality of prints? If not, what kinds of research are needed to improve understanding of these issues?

Although latent print examination has been used in the legal system for over 100 years, studies assessing the accuracy of latent print identifications have appeared only recently. In the key studies, fingerprint examiners were asked to compare prints of known origin, allowing examiners' accuracy to be measured against ground truth (see e.g., Kellman et al., 2014; Langenburg, Champod and Genessay, 2012; Pacheco et al., 2014; Tangen et al., 2011; Thompson, Tangen and McCarthy, 2014; Ulery et al., 2011, 2012). The studies show low rates of false identifications, ranging from 0 to 2.6%; but somewhat higher rates of false exclusions, ranging from 2.9% to 28%. (A false identification is the mistaken conclusion that prints made by different people have the same source; a false exclusion is the mistaken conclusion that prints made by the same person have different sources.)

The PCAST report argued emphatically and persuasively that studies of this type are essential for establishing the validity of latent print identification by human examiners.

For subjective methods, foundational validity can be established *only* through black-box studies that measure how often many examiners reach accurate conclusions across many feature-comparison problems involving samples representative of the intended use. In the absence of such studies, a subjective feature-comparison method cannot be considered scientifically valid.

Foundational validity is a *sine qua non*, which can only be shown through empirical studies. Importantly, good professional practices—such as the existence of professional societies, certification programs, accreditation programs, peer-reviewed articles, standardized protocols, proficiency testing, and codes of ethics—cannot substitute for empirical evidence of scientific validity and reliability (PCAST, 2016, p. 66).

The PCAST report also sets forth a number of criteria that scientific validation studies should satisfy:

(a) they should be based on sufficiently large collections of known and representative samples from relevant populations; (b) they should be conducted so that the examinees have no information about the correct answer; (c) the study design and analysis plan should be specified in advance and not modified afterwards based on the results; (d) the study should be conducted or overseen by individuals or organizations with no stake in the outcome; (e) data, software and results should be available to allow other scientists to review the conclusions; and (f) to ensure that the results are robust and reproducible,

there should be multiple independent studies by separate groups reaching similar conclusions (PCAST, 2016, p. 66).

After reviewing research on the accuracy of human latent print examiners, the PCAST report concluded that “[r]emarkably, there have been only two black-box studies that were intentionally and appropriately designed to assess validity and reliability” of latent print analysis (PCAST, 2016, p 91). One of the studies was published by FBI employees and contractors (Ulery et al., 2011); the other study was completed in 2014 at the Miami-Dade Crime Laboratory, but has not yet been published in a peer-reviewed journal (Pacheco et al., 2014). Nevertheless, PCAST concluded that these studies have established the “foundational validity” of latent print analysis, which means that latent print analysis has been shown to be “repeatable, reproducible, and accurate, at levels that have been measured and are appropriate for the intended application” (PCAST, 2016, p. 4).¹⁴ Consequently, according to the PCAST report, courts (and the general public) should accept that latent print analysis can, in principle, be reliable. When evaluating latent fingerprint evidence, the focus should be on “validity as applied”—i.e., whether the expert has applied the method properly in the case at hand.¹⁵

We have reviewed the same literature and our conclusions largely align with those of the PCAST report, although we do not attempt to distinguish “foundational validity” from “validity as applied” and we consider all studies that examine the accuracy of latent print examiners, rather than focusing just on those that are “intentionally and appropriately designed” for a particular purpose. Our goal is to draw conclusions from the literature as a whole, recognizing (consistent with the concept of convergent validity) that studies will have different strengths and limitations, and that the literature as a whole will have strengths and limitations.

In our view, three important conclusions can be drawn from these accuracy studies:

- (1) *Performance improves with training*—In these studies, trained examiners performed significantly better (lower error rates) than trainees; trainees performed much better than novices (Langenburg, Champod and Genessay, 2012; Tangen et al., 2011; Thompson et al., 2013, 2014). However, among trained experts there were no significant differences between relatively new and more experienced examiners (Arora et al., 2015; Thompson et al., 2014). These studies indicate that trained fingerprint examiners have expertise that makes them much better at the task of fingerprint

¹⁴ Questions might be raised about whether the two studies PCAST considered relevant to the question of foundational validity actually meet the criteria (quoted above) that PCAST established for such validation research. We suspect there might be some debate about whether these studies were “conducted or overseen by individuals or organizations with no stake in the outcome” and whether two studies (one unpublished) constitute “multiple independent studies by separate groups.” Nevertheless, we agree that these are important studies that speak directly to the question of foundational validity of latent print analysis.

¹⁵ The PCAST report suggests that its concept of “foundational validity” corresponds to the legal requirement in Rule 702(c) of the Federal Rules of Evidence that the expert testimony be the “product of reliable principles and methods,” while the concept of “validity as applied” corresponds to the legal requirement in Rule 702(d) that “the expert has reliably applied the principles and methods to the facts of the case.”

comparison than non-experts. However, the studies raise questions about how much, if anything, trained examiners gain through additional years of experience. In the United States, examiners with extensive experience often seek certification from the International Association for Identification (IAI).¹⁶ However, those who were certified performed no better than those who were not certified (Pacheco et al., 2014; Ulery et al., 2011).

- (2) *Performance varies depending on the difficulty of the comparison*—Not surprisingly, error rates were higher in studies in which the comparisons were more difficult. For example, Langenburg, Champod and Genessay (2012) claimed that 12 of the comparisons they asked examiners to make were exceptionally difficult because the prints were of low quality with limited detail. This high level of difficulty undoubtedly explains why this study showed the highest rate of false positives (2.6%). The second highest false positive rate (1.6%) was found by Thompson, Tangen and McCarthy (2014), who selected imposter (non-matching) prints for some of their comparisons through an AFIS search of a database for prints with high similarity scores. These findings indicate that the probability of error in a particular case may vary considerably depending on the difficulty of the comparison. Factors such as the quality of the prints, the amount of detail present, and whether the print was pre-selected based on its similarity to the latent will all be important. The reliability of examiners' judgments is also affected by the difficulty of the comparisons. When Ulery et al. (2012, p. 1) asked 72 examiners to repeat their assessments (made seven months earlier) of latent-exemplar pairs, they found that "repeatability and reproducibility [of examiner's decisions] were notably lower for comparisons assessed by the examiners as 'difficult' than for 'easy' or 'moderate' comparisons..." In light of this variation, it is unreasonable to think that the "error rate" of latent fingerprint examination can meaningfully be reduced to a single number or even a single set of numbers (Kellman et al., 2014). At best, it might be possible to describe, in broad terms, the rates of false identifications and false exclusions likely to arise for comparisons of a given level of difficulty
- (3) *Performance varies across examiners*—There is considerable variation among examiners in the features they identify as important during examinations (Ulery et al., 2014, 2016; Hicklin et al., 2011, 2013).¹⁷ There is also variation among examiners in whether the features observed in particular comparisons are deemed sufficient for identification or for exclusion (rather than deeming the evidence inconclusive). These differences suggest that there is room for further improvement in examiner performance through better assessment of their performance and better training.

¹⁶ The IAI is a major professional association for pattern matching experts, including latent print examiners. Its voluntary certification program for latent print examiners is reportedly rigorous and challenging (Grieve, 1990), with passing rates around 70% (https://www.theiai.org/certifications/latent_print/intro.php (accessed September 12, 2017)).

¹⁷ In one study, for example, six well-trained examiners were asked to mark-up 258 latent prints for submission to an AFIS search. There was considerable variation in the features that the examiners marked as important which, in turn, cause variation in the resulting accuracy of the AFIS (Arora et al., 2015).

Studies of the accuracy of examiners' performance are extremely important for evaluating factors that may influence the rate of error, and for comparing the performance of trained examiners against the performance of other groups, but the existing studies generally do not fully replicate the conditions that examiners face when performing casework. Consequently, the error rates observed in these studies do not necessarily reflect the rate of error in actual practice (Haber and Haber, 2014; Koehler, 2017; Thompson et al., 2014).

A key limitation of these studies is that the examiners knew they were being tested, which may have affected their performance.¹⁸ In particular, it may have affected their threshold of decision—that is, how readily they reached a definitive conclusion based on a given set of data (Biedermann, Bozza and Taroni, 2016). Psychological research has shown that subjective decisions are often influenced by the decision makers' impression of the consequences of different types of errors (Kassin, Dror and Kukucka, 2013). The participants in these fingerprint studies might have worried that a false identification would be used to impugn their profession, and so they might have set an unusually high threshold for reporting an identification; alternatively, since examiners in many of these tests were anonymous, and knew that any errors would not have personal repercussions, they may have lowered their thresholds. There is no assurance, however, that examiners will set the same sufficiency threshold for decision-making in actual practice. Perceptions of the consequences of possible errors may well vary from setting to setting and from case to case in ways that greatly affect willingness to reach a definitive conclusion, and hence the probability of a false identification or false exclusion.

Ulery et al. (2017) noted the importance of varying decision thresholds in explaining errors and disagreements among examiners over "exclusions":

In making an exclusion decision, the examiner considers his/her assessment of similarities and dissimilarities, along with his/her level of uncertainty in this assessment, and then determines if the information is sufficient to render an exclusion. The sufficiency threshold is based on an implicit utility function, in which the examiner considers the relative benefits of making a correct exclusion versus the costs of making a mistake. Errors and disagreements among examiners may be due in part to lack of guidance on the relative costs and benefits of each decision, or systematic pressures encouraging some decisions more than others. These pressures will vary by agency or among cases, and examiners' responses to these pressures will vary (p 66).

Of course, the same variation in decision thresholds may also affect the decision to report an "identification." This consideration provides further support for the conclusion that the error rates in black-box studies may not reflect error rates in casework. Consequently, as Koehler notes,

¹⁸ Informing someone that they are being tested can create what psychologists call "demand characteristics" that change the person's responses (Orne, 1962). Individuals who know they are being tested may shift their threshold of decision in ways designed to make them look good (Paulhus, 1991). Hence, performance testing provides a more realistic picture of human performance if the participants do not know they are being tested.

Nobody knows the rate at which forensic science examiners produce false match reports or otherwise reach the wrong conclusion. Unlike in medicine, there is no systematic research program in the forensic sciences that seeks to identify the rates at which errors are made in practice (Koehler, 2017, p. 153).

In our view, the absence of such a systematic research program is a significant problem. While the black-box studies conducted to date show that latent print examiners are capable of evaluating prints with a high level of accuracy, we do not know how accurate examiners are in practice. Nor is there an adequate foundation for understanding how their accuracy is affected by: (a) level of training and experience; (b) individual differences in perceptual ability; (c) analytic procedures and standards of practice; (d) quality control and quality assurance procedures; or (e) the quality of prints. To answer these questions, additional research is required. This research should be conducted in a “test-blind” manner, with samples incorporated into examiners’ routine casework, so that they do not know when they are analyzing an actual case and when they are participating in research.

Research in which participants are “blind” to the existence of the study poses a number of challenges, which may explain why no such studies have been conducted in the field of latent print examination. Blind research trials are particularly difficult to carry out if examiners communicate directly with detectives and have access to police reports and other information about a case. To conduct such a study, laboratory managers would need to enlist the support of police in preparing simulated case materials. The materials would need to be sufficiently realistic to pass as a real case. Furthermore, the police might need to provide other false information to the examiner to prevent the examiner from discovering that the case was simulated. Elaborate simulations of this type are feasible and have been conducted successfully to test the accuracy of DNA analysis (Peterson et al., 2003), but they are burdensome and expensive.

Fortunately, research of this type is much easier to conduct in laboratories that employ a context management system to shield examiners from irrelevant contextual information. If the examiner typically is exposed only to the prints themselves (along with task-relevant information about how the prints were collected and the surface from which they were lifted) the case manager could insert a test comparison occasionally without the analyst being able to distinguish the test from routine casework. As noted earlier, blinding procedures (to reduce contextual bias) are a key element of good scientific practice. That these procedures facilitate blind testing of examiners is another reason they should be implemented.¹⁹ In an ideal research program, a certain percentage of each examiner’s casework could be test comparisons (without the examiner’s knowledge of which cases they were). Perhaps one case in 20 that an examiner worked would be a constructed test rather than an actual case. Because examiners in some agencies often work from digital images of prints, it would be possible for a government

¹⁹ This same point has been made by the National Commission on Forensic Science, which noted: “The ability to implement blind research ... is a secondary benefit that arises when laboratories adopt context management systems, as recommended by this Commission.” National Commission, 2016, p. 3.

agency or a commercial vendor to prepare a large number of test comparisons that could be distributed to laboratories nationwide to facilitate research.

The PCAST report has also called for “test-blind” research on the accuracy of forensic examiners, calling it the “ideal” approach to proficiency testing. The PCAST report uses the term “proficiency testing” to refer to “ongoing empirical tests to evaluate the capability and performance of analysts” and declares that proficiency tests “should be performed under conditions that are representative of casework and on samples, for which the true answer is known, that are representative of the full range of sample types and quality likely to be encountered in casework in the intended application” (PCAST, 2016, p. 57). While we agree that empirical testing of this type would be helpful in assessing the competence of particular examiners, we believe it will serve a broader purpose. This kind of “empirical testing” is the best way to address basic questions about the validity of latent print examination. It is not merely a test of individual proficiency (although it surely is that), it is also research that is needed for purposes of validation and quality assurance.

The PCAST report recognizes that challenges will arise in implementing “test-blind” research/proficiency testing in forensic laboratories. The main problem is that “laboratories vary considerably as to ...what information is provided to an analyst about the evidence or the case in question.” (p. 59). Consequently, it recommends that laboratories be given some time to implement it:

PCAST believes that test-blind proficiency testing of forensic examiners should be vigorously pursued, with the expectation that it should be in wide use, at least in large laboratories, within the next five years. However, PCAST believes that it is not yet realistic to require test-blind proficiency testing because the procedures for test-blind proficiency tests have not yet been designed and evaluated (p. 59).

The National Commission on Forensic Science has also called for “test-blind” research in forensic laboratories. In September 2016, the Commission unanimously adopted a “views document” that declared:

It is the view of the National Commission on Forensic Science that:

1. Additional research is needed to assess the performance of forensic science laboratories on routine analytic tasks such as comparison of samples to determine whether they have a common source.
2. Studies should be conducted by introducing known-source samples into the routine flow of casework in a blinded manner, so that examiners do not know their performance is being studied.
3. Government agencies should facilitate research of this type by funding pilot research programs.
4. Government agencies should facilitate research of this type by developing (or funding the development of) sets of test samples that can be used to carry out

research. The research test sets should be designed with the assistance of practitioners, statisticians and experts on research methodology in order to facilitate studies that address important scientific questions.

5. Government agencies should identify and revise any regulations or memoranda of understanding regarding access to databases that prohibit, or appear to prohibit, access to these databases for the purpose of conducting research on laboratory performance.
6. In order to avoid unfairly impugning examiners and laboratories who participate in research on laboratory performance, judges should consider carefully whether to admit evidence regarding the occurrence or rate of error in research studies. If such evidence is admitted, it should only be under narrow circumstances, and with careful explanation of the limitations of such data for establishing the probability of error in a given case.

Once laboratories develop the capability of introducing known-source samples into the normal flow of casework, for the purpose of research, it will be possible to design and carry out studies on a variety of important issues. The PCAST report focused on the importance of research to establish error rates in routine testing of typical samples under standard laboratory conditions. We agree that such studies would be the best way to estimate error rates and believe such studies would provide a more accurate picture of the accuracy of latent print examination in practice than existing black box studies. But research of this type should not be limited to error rate estimation. Research of this type could and should address a variety of other important issues.

For example, studies could be designed to examine the number and kind of features that must be present in a latent print in order for an examiner to make an accurate report. Research using this methodology could also explore the extent to which accuracy is impaired by systematically introducing various kinds and degrees of distortion into latent prints. The effect of examiner characteristics (e.g., visual acuity) and of different training programs on accuracy could also be explored. Those designing such studies would ideally consult with practitioners to assure that the test samples are realistic and that the research addresses meaningful questions. They would also consult with experts on research design and statistics to ensure that study methodology is sound.

The test samples chosen for such studies will vary depending on the purpose of the research. For research designed to estimate error rates in practice, it will be important that researchers present test comparisons that replicate the full range and distribution of difficulty normally faced by examiners. The study by Pacheco, Cerchiai and Stoiloff (2014) illustrates the way in which researchers might work with practitioners to create appropriate test samples for use in such a research program. Although participants in the Pacheco et al. study were not blind to the fact that they were being tested, the sample preparation strategy used for this research could also be used to create a set of test samples for use in blinded research. Pacheco and colleagues asked three IAI certified latent print examiners to evaluate 320 latent prints based on difficulty of comparison. Each latent print was rated on a scale from 0-21 points and divided into three

groups (insufficient to difficult, difficult to moderate, and moderate to easy) in order to present the study's participants with a broad range of latents that were representative of actual casework.

As noted above, however, blinded research will be useful for purposes beyond error rate estimation, and this may require different and less representative sets of test samples. If researchers are interested in studying the ability of examiners to recognize and deal with distortion, for example, the study could include latent prints with extreme amounts of distortion. If researchers were interested in testing the limits of examiners' expertise in dealing with extremely difficult comparisons, then the test samples would include comparisons chosen precisely because of their challenging nature.²⁰ In addition to their value for understanding the limits of examiners' expertise, studies with challenging samples could be helpful to examiners. The researchers could easily provide feedback to examiners on their performance on difficult cases, thereby helping them identify weaknesses in their evaluations and improve their skills.

Research of this type could also be helpful for improving personnel selection. At present, we know little or nothing about the perceptual and cognitive abilities needed to excel at latent print examination. Anecdotal evidence suggests that some examiners are more gifted than others. Some examiners reportedly develop reputations for special talent with particular tasks, such as comparing degraded or distorted latent prints. Research of examiners' performance on difficult cases, if combined with measurement of perceptual and cognitive ability, might make it possible to identify specific abilities associated with a high level of performance. That, in turn, might make it possible to identify such individuals in advance—to determine which candidates for the job of latent print analyst are most promising. For example, it would be useful to determine whether individual differences in visual acuity affect examiners' performance on the most challenging cases.²¹ If so, visual acuity could become a useful factor in personnel selection.

²⁰ In a testing program of this type, errors are to be expected. Indeed, frequent errors will be the hallmark of an effective testing program. But forensic scientists may hesitate to engage in the challenging kind of testing called for here if the results for highly challenging cases are unfairly used to impugn their competence or question their performance on more routine cases. Consequently, we join the National Commission on Forensic Science in urging state and federal judges to consider carefully whether the rate of error in challenging performance tests should be admissible in the courtroom:

While the results of such research will be valuable and enlightening on a number of important issues, it would be misleading to equate the rate of error on research samples designed to be highly challenging with the rate of error for cases in general or with the probability of error in a specific case, particularly if the case involved relatively easy or straightforward analysis. Consequently, if the results of performance testing are admitted as evidence in the courtroom, it should only be under narrow circumstances, and with careful explanation of the limitations of such data for establishing the probability of error in a given case (National Commission, 2016, p. 10).

²¹ Here again it is important that the research focus on challenging comparisons. If the comparisons are easy, and the rate of error is very low, it will be difficult to distinguish outstanding performers from those who are merely average because all of them would get the right answer most of the time.

Most laboratories require examiners to take periodic proficiency tests, but current proficiency testing programs, although vital for assuring that examiners possess basic competency, do not achieve many of the important benefits that could be achieved with the kind of research program outlined here (Koehler, 2017; forthcoming). Most current proficiency tests have several limitations: analysts know they are being tested (which may cause them to perform differently during proficiency tests than when performing casework); the tests involve relatively few samples; and the tests are typically designed to assess basic competency rather than performance on highly challenging cases.²² Hence, as Koehler notes, “the tests that examiners take are generally so easy, unrealistic, and otherwise unlike case-work, that even the test manufacturers have said that error rates on these ‘proficiency tests’ should not be used to estimate casework error rates” (Koehler, 2017, p. 154).

In their current form, proficiency tests have limited value for establishing the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary. These tests provide little useful feedback to forensic analysts on the limits of their expertise when dealing with difficult cases or marginal evidence and hence have little value for helping experienced analysts hone and improve their skills. They have little value for evaluating the effectiveness of training programs, for evaluating the relative strengths and weaknesses of various analysts, or other such functions.

The 2009 NRC report called for the development of “quantifiable measures of the reliability and accuracy of forensic analyses” that reflect “actual practices on realistic case scenarios...” It also declared that these studies should establish “the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary.” (NRC, 2009, p. 23, Recommendation 3(b)). The best way to realize these important goals in the domain of latent fingerprint examination is to fund and conduct the kind of research program described here.

Studies Testing the Performance of Human Examiners

In this section, we discuss details of the key studies that we relied upon when evaluating the literature on the accuracy of latent print examiners. Most of these studies were also analyzed and discussed in the recent PCAST report. Our assessment of the strengths and weaknesses of each study largely aligns with that of PCAST. When evaluating these studies, it is important to consider several features: (1) the background and training of the participants; (2) the source and nature of the prints that participants were asked to compare; (3) the number of same-source prints and different-source prints that participants were asked to compare; (4) the number of comparisons on which participants offered a conclusion; and the number they

²² The president of Collaborative Testing Services, an organization that provides test samples that are widely used for proficiency testing in forensic laboratories, told the National Commission on Forensic Science during its seventh meeting (August 10, 2015) that he is under constant commercial pressure to make proficiency tests easier (National Commission, 2016).

deemed inconclusive; (5) the number of correct and incorrect conclusions that participants reached when comparing same-source and different-source prints.

The key metrics of examiner performance are the false positive and false negative rates—i.e., the percentage of different-source comparisons that are incorrectly associated (false positives); and the percentage of same-source comparisons that are incorrectly distinguished (false negatives). According to the PCAST report, the most scientifically relevant false positive and false negative rates are those for comparisons in which participants ventured a conclusion “because fingerprint evidence used against a defendant in court will typically be the result of a conclusive examination” (see PCAST, 2016, p. 92). Consequently, PCAST ignores “inconclusive” determinations when computing error rates.²³ Whether or not they are included in calculating error rates, the rate of “inconclusive” should be monitored and reported by researchers. It is likely to be an important metric when assessing examiner performance in test-blind research.²⁴

Evetts and Williams (1995) tested 130 fingerprint experts from England and Wales, each with ten or more years of experience. Participants were asked to compare ten latent prints to known impressions. Nine of the comparisons were from past casework and were presumed to be same-source pairs; the tenth comparison involved prints from different people. The authors report no erroneous identifications and ten erroneous exclusions, but failed to report the number of comparisons on which participants offered a conclusion (rather than calling it inconclusive), which makes the study difficult to evaluate.

Wertheim, Langenburg, and Moenssens (2006) published data on the rate of false identifications that occurred during training exercises they conducted for latent print examiners. Of their 108 participants, 92 reported more than one year of experience as examiners and the other 16 reported less than one year or no experience/training at all. Participants were given exercise packets consisting of 10 latent prints and ten-print cards of eight possible source individuals. An important weakness of the study is that participants were told that one of the eight individuals was the source of each latent; the exercise required them to “find the match” knowing that it was one of the 80 (10 prints x 8 individuals) reference prints in the packet. Participants completed multiple packets. The difficulty of the packets had been rated and participants were given easier or harder packets in successive rounds based on their performance in early rounds. Overall, participants made 7492 “individualizations” and 7373 of these (98.5%) were correct; indicating a false positive rate of 1.5%. However, the authors of

²³ Examiners who know they are being tested may be more cautious about reaching conclusions than they would ordinarily be, resulting in a high rate of inconclusive findings. In such circumstances, it could be misleading to include inconclusive results when computing error rates. As an extreme example, suppose an examiner compared 100 different-source prints, reached a conclusion on only 10 of the comparisons (those where the conclusion could be reached with the highest level of confidence), and falsely concluded that one of the 10 pairs was from the same finger. The false positive rate for conclusive results (1 in 10) would arguably be a better indication of the examiner’s capability than the false positive rate for all comparisons (1 in 100).

²⁴ Indeed, test-blind research in forensic laboratories may allow laboratory managers to test the effects of changes in training or procedure in order to find ways to decrease the rate of “inconclusives” without unacceptable increases in the rate of error.

the study determined that 77 of the 118 false identifications had been caused by “clerical error” in recording the results rather than an actual false association,²⁵ which suggests a false positive rate of only 0.5%. The authors noted further that the rate of false identifications was higher for participants with less than a year of experience (2%) than participants with more than a year of experience (0.2%). The authors also broke down the rate of error according to participants rating (on a scale of 1-3) of their confidence in their conclusion. Among experienced examiners who expressed the highest level of confidence, only 2 of 5861 reported individualizations were to the wrong finger, a false positive rate of only 0.034% (however, this was after the authors removed 59 false individualizations that they attributed to “clerical error”). Given the obvious design flaws of this study, we take the authors’ conclusions with a grain of salt, but we note that the findings are generally consistent with those of better-designed studies.

Gutowski (2006) presented data on the performance of Australian latent print examiners, and of all test takers, on the Collaborative Testing Service (CTS) Fingerprint Proficiency Tests for the years 2000-2005. CTS is a commercial service that provides proficiency tests for forensic scientists worldwide. The tests require examinees to compare sets of print impressions and to determine whether specified pairs have, or do not have, a common source. Examinees knew that they were being tested, but did not know the correct answers. Considering only instances in which examinees reached a conclusive result, there were 4156 different-source comparisons resulting in 100 false identifications and 4056 correct exclusions—a false positive rate of $100/4156=0.024$, or 2.4 percent.²⁶ The false positive rate for the Australian participants, who were all certified, experienced examiners employed by the Victoria Police Forensic Science Department, was $2/105=0.019$, or 1.9 percent.²⁷

²⁵ False identifications were attributed to clerical error when the participant had identified the correct individual (of the eight possibilities) but mistakenly recorded a match to the wrong finger; or when participants identified the correct finger (e.g., right index finger) but mistakenly recorded a match to the wrong individual. While we trust that the researchers made a good faith effort to ascertain the truth, it is conceivable that some of these “clerical errors” were actual errors (false identifications or false exclusions) that were mistakenly attributed to clerical mishap due to the coincidence of matching the correct finger (a 1 in 10 chance) or the correct individual (a 1 in 8 chance). It is also conceivable that such error might occur in practice, creating false matches or false exclusions. Another critique of the Wertheim et al. study, offering a reanalysis of the error rate data, is offered by Cole (2006, p. 84).

²⁶ Gutowski computed what he called “the overall error rate for false positives” by dividing the number of false positive by overall number of comparisons called for by the tests, and concluded that it was $100/30,642=0.0033$, or 0.33 percent. In our view, this is a meaningless and misleading calculation because the great majority of comparisons in the CTS proficiency tests involved latent prints from the same finger. For these comparisons, it is not possible to have a false positive result. As noted above, the false positive rate is the percentage of different-source comparisons that are incorrectly associated.

²⁷ Collaborative Testing Service has warned that its proficiency tests were not designed and should not be used for error rate estimation (Collaborative Testing Services, Inc. CTS Statement on the Use of Proficiency Testing Data for Error Rate Determinations, March 30, 2010, pg. 3). While we agree that these proficiency tests are not well-designed for error-rate estimation, we believe the results of these tests nevertheless provide useful information about the potential for error. If nothing else, they refute claims that errors are extraordinarily unlikely events. For example, there were a surprising number of errors on the most recent CTS proficiency test (Collaborative Testing Services, Inc. Forensic Testing Program, Latent Print Examination Test No. 16-515/516 Summary Report. May 11, 2016). Four hundred thirty-one latent print examiners took this test, which required that they compare

Langenburg, Champod, and Wertheim (2009) reported data from an experiment involving 43 latent print examiners who were attending a professional conference. Participants were divided into three groups. For two of the groups, the researchers attempted (rather unsuccessfully, it appears) to introduce contextual bias by telling them, before they compared prints, what conclusion another examiner had reached when comparing the same prints. The authors report that participants were skeptical of this attempted manipulation and suspected that it might be an attempt to bias them toward an incorrect conclusion. The third group was a control-group in which participants were simply asked to compare prints and were told nothing about what other examiners had concluded. Each participant compared six pairs of prints, three of which were from the same source and three of which were from different sources. There were no false positives or false negatives in the two experimental groups, but participants in these groups were approximately three times more likely than participants in the control group to report comparisons as “inconclusive,” which suggests they may have been unusually cautious about reporting a conclusive finding. In the control group, the 15 participants made 45 comparisons of same-source prints and reported 39 individualizations, 3 exclusions, and 3 “inconclusive.” The false negative rate (computed as PCAST recommends) was thus 3/42 or 7%. Participants also made 45 comparisons involving different source prints and reported 42 exclusions, one false identification, and two “inconclusive,” for a false positive rate of 1/43 or 2.3%. The PCAST report computed the false positive rate the same way and noted that the upper 95% confidence bound on this false positive rate is 11%.

Langenburg (2009) published data from a pilot study of examiner performance. The study was designed to examine the accuracy, precision, reproducibility, repeatability, and “biasability” of the ACE-V process. A total of six experienced examiners from the same agency laboratory (with 6 – 35 years’ experience) performed comparisons of 120 latent impressions to eight known standards. The study was separated into three phases and resulted in 60 ACE and 60 ACE-V trials per participant. The study reported no erroneous identifications; however, as the PCAST report notes, there were very few conclusive examinations involving comparison of different-source prints.²⁸ Because there were so few opportunities for a false identification to occur, the failure to observe a false identification tells us little about what the true rate of false

12 latent prints with reference prints from 4 individuals. Thirty-seven of the participants incorrectly identified at least one of the latent prints as coming from the wrong finger (there were 51 false identifications in all); 11 participants mistakenly excluded a finger that was the true source of a latent print, and 4 of those also mistakenly identified it as coming from a different finger.

²⁸ The published report of this study is a bit unclear with regard to the total number of same-source and different source comparisons that participants were asked to make. The PCAST report puts the number of different-source comparisons on which examiners reached a conclusive result at 15. Based on an observed false identification rate of 0/15, the PCAST report puts the upper 95 percent confidence bound for the false positive rate at 19%. Our reading of the article suggests that the number of different-source comparisons that led to a conclusive report may have been only 12, which would make the upper bound of the confidence interval even higher. In either case, the number of different-source comparison is far too small to draw reliable conclusions about the rate of false positives. The author of the study repeatedly emphasized that the results showed “100% accuracy for all trials where an opinion of identification was reported,” but this finding means little given the small number of opportunities that existed for a false identification to occur.

identifications might be. It appears that the rate of false exclusions was 2/270 (0.7%) for the ACE phase of the study and 6/277 (2.2%) for the ACE-V phase of the study.

Tangen, Thompson and McCarthy (2011) reported an Australian study in which 37 experienced latent print examiners and 37 novices (college students with no training in fingerprint examination) were asked to compare latent-exemplar pairs and rate the likelihood they came from the same source on a 12-point scale. In order to compute error rates, the authors chose to label scores of 1-6 on the scale as identifications and scores of 7-12 as exclusions. Of course, this is not the procedure actually used in forensic laboratories, and for that reason alone the results of this study may not accurately reflect error rates in practice. Nevertheless, the results reveal some interesting patterns.

There were 12 same-source pairs; 12 different-source pairs for which the exemplar was chosen at random from a database; and 12 different-source pairs where the exemplar was the print that an AFIS ranked most similar to the latent among the prints in a government database. Not surprisingly, the trained examiners vastly outperformed novices, especially on the more challenging different-source pairs that were selected from a database to be similar. For the randomly chosen different-source pairs, the examiners had no false identifications in 444 comparisons; for the different source pairs chosen to be similar, the examiners made 3 false identifications in 444 comparisons. PCAST estimated the upper bound of the 95% confidence interval for these comparisons to be 1.7%.

Ulery et al. (2011) published a large-scale study on the accuracy and reliability of forensic latent fingerprint decisions known as the FBI Black-box study. It is clearly the best designed and most extensive of the studies on the accuracy of latent print examiners. The participants were 169 experienced latent print examiners who varied with respect to their organization, training history, and other demographics. The median number of years of experience was ten, and 83% of the participants were certified latent print examiners. Each participant compared approximately 100 pairs of latent and exemplar prints from a pool of 744 pairs that were assigned at random. The pool included 520 same-source pairs and 244 different-source pairs. The exemplars for the different source pairs were selected from among the closest matches found when searching the latent against 58 million prints in an AFIS database. The goal of this selection procedure was to make the different-source pairs representative of the kind of comparisons that examiners must make when candidate prints are selected from a database.

- Participants made a total of 17,121 comparison, 23% of which resulted in a conclusion of “no value.”
- A total of 4083 different-source pairs were deemed of value for identification, and examiners were able to make conclusive calls on 3638 of those pairs. Six of those calls were erroneous identifications (0.17%). (The PCAST report puts the upper 95% confidence bound at 0.33%, which corresponds to 1 error in 306 cases). The six errors were committed by five examiners, three of whom were certified, one who was not certified, and the certification status of the other was not known (one certified examiner made two erroneous identifications).

- In addition, participants reported 450 erroneous exclusions (7.5%) among 5,969 latents that were deemed to be of value for identification.
- 85% of the participants made at least one false exclusion.

In a follow-up to their initial study, Ulery et al. (2012) reported results on the repeatability and reproducibility of decisions by latent print examiners. Of the 169 examiners who participated in their initial study, 72 examiners were presented with the same prints after a seven-month interval to determine if one examiner would consistently reach the same decision on the same fingerprints (repeatability); they were not told they had previously seen these prints. Data were also reported from the same 72 examiners' initial test results to determine whether different examiners had reached the same decision on the same fingerprints (reproducibility).

- Each examiner in the repeatability retest was assigned 25 comparisons from a pool of 744 image pairs.
- Latent print examiners repeated 89.1% of their individualizations and 90.1% of their exclusions.
- No false positive errors were repeated, and 30% of false negative errors were repeated. Most of the changed decisions resulted in inconclusive decisions.
- Repeatability of all comparison decisions combined was 90.0% for mated pairs; 85.9% for non-mated pairs.

Langenburg, Champod and Genessay (2012) reported research on how examiners' performance was affected by using novel tools designed to assist them in latent print comparisons. These tools included: (1) a quality tool to assess the clarity of the friction ridge details; (2) a statistical tool to provide likelihood ratios representing the strength of the corresponding features in a comparison; and (3) consensus information from a group of trained fingerprint experts. Participants were 159 trained fingerprint examiners as well as some novices and trainees. They were asked to compare seven same-source and five different-source latent-exemplar pairs of prints. These comparisons were designed to be difficult. The latent prints were of marginal quality for comparison and some had distortion and artifacts. The different-source exemplars were chosen from a database to be similar to the latents with which they were paired.

PCAST obtained from the researchers' data for the trained examiners alone (excluding the novices and trainees) and reported the following results:

For the non-mated pairs, there were 17 false positive matches among 711 conclusive examinations by the experts. The false positive rate was 2.4 percent (upper 95 percent confidence bound of 3.5 percent). The estimated error rate corresponds to 1 error in 42 cases, with an upper bound corresponding to 1 error in 28 cases (PCAST, 2016, p. 93; footnotes omitted).

Given the difficulty of the task, it is not surprising that this study had the highest false positive rate (of studies of this type).

Kellman et al. (2014) reported an important study designed to determine what features of fingerprints make latent print examinations easy or difficult. They asked 56 fingerprint examiners to evaluate latent-exemplar pairs. For half of the comparisons the exemplar was known to be from the same finger as the latent; for the other half, the exemplar was a print from a different finger selected from a database due to its similarity with the latent. The false-positive rate was 3% and the false-negative rate was 14%. However, the conditions of the study did not match typical conditions of forensic practice. For example, examiners were given only three minutes to determine whether the prints were from the same or different fingers and responses of inconclusive or “no value” were not allowed. Hence, the error rates in the study probably overstate the error rates that would occur in actual practice. The primary value of the study was its identification of fingerprint features that make a comparison difficult and thereby increase the risk of error. This study points the way toward the development of objective metrics for the difficulty of fingerprint comparison that could be valuable in assessing examiner performance and evaluating the risk of error.

The study by Pacheco et al. (2014) has not been published but is described in a report to the US DOJ that has been posted online (<https://www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf>, accessed September 12, 2017). The researchers asked 109 latent print examiners with at least one year of experience each to evaluate 40 latent prints. Participants were required to compare each latent print to ten-print reference standards from three individuals. Unlike some of the previous studies, the reference standards were not chosen from a database based on their similarity to the latent. They were simply taken from ten individuals who volunteered to provide prints for the study. Examiners knew they were participating in a study, but compared the prints in the manner they normally would in casework. Among the comparisons on which examiners made a conclusive determination, the false positive rate was 4.2% and the false negative rate was 8.7%. The PCAST report notes that the upper-bound of the 95% confidence interval on the false positive rate would be 5.4%, which corresponds to 1 false match in 18 cases (PCAST, 2016, p. 95). The researchers suggested that some of the false positives may have arisen from “clerical error” in recording conclusions rather than from mistakes in reaching conclusions, but acknowledge that they “could not determine this with certainty” (Pacheco et al., 2014, p. 64). The researchers conducted a second and third phase of the study in which some examiners were asked to verify the conclusions of others. All of the false positives and over half of the false negatives were “detected” during the verification phase, which led the researchers to report that the false positive rate for the entire ACE-V process (including verification) was 0% and the false negative rate was 3%. This finding suggests that verification is an important process in catching errors. Because the verification process used in the study differed in important ways from actual practice, however, it is not clear whether verification would be as effective in practice as observed in this study.

Liu, Champod, Wu and Luo (2015) reported a study in which 40 Chinese latent print examiners each evaluated five latent-exemplar pairs of prints for which ground-truth (same-source or different-source) was known. The prints were selected to represent “difficult cases,” where the experts would be operating “at the boundaries of the decision limits” (p. 34). The examiners used an online platform to annotate and compare the prints, which allowed the researchers to

study not only examiners' overall accuracy but also some of the steps in their analysis of the prints. The study found significant variability among the examiners in the features they marked for comparison and in their assessments of factors that affected the appearance of the prints. Examiners also varied in their assessment of whether some of the prints had value for comparison. While examiners made accurate source determinations for four of the five pairs of prints, one pair, identified as a "close nonmatch" (CNM) was mistakenly identified as being from the same source by three of 27 examiners who found the prints to be of value for identification (p. 35). The researchers note that CNM prints have "similar ridge flow, ridge path, ridge sequence and minutia [sic] type, and configuration" (p. 35). They also note that the use of AFIS to identify potential matches has increased the risk that examiners will be faced with CNM prints. They suggest that "[a]s database size is increasing, examiners' experience alone is no longer sufficient to deal with CNM prints" (p. 35). They also suggest that more attention should be given to the sources of within-print variability in order to help examiners evaluate the significance of "dissimilarity" they observe during the print comparison process, a suggestion consistent with our recommendations in Section II (p. 35).

When considering the literature as a whole, it is important to note that the print comparisons in many of these studies were not performed under laboratory conditions and did not include the verification step of the ACE-V process, which might have caught some of the errors. On the other hand, the examiners knew they were being tested, which might have made them more careful. The difficulty of generalizing the results of such studies to actual laboratory practice is one of the reasons we favor test-blind research in forensic laboratories. Research performed on working analysts who do not know they are being tested while working under standard laboratory conditions would be the best way to gain further insight into the potential for error in latent print analysis, and how to reduce the rate of error.

Conclusions

- Studies of the accuracy of latent print examiners leave no doubt that trained fingerprint examiners have expertise that makes them much better at the task of fingerprint comparison than non-experts.
- It is not clear, however, whether the error rates observed in existing studies reflect the error rates that occur in the actual practice of latent print analysis. Such studies can in principle determine the relative strength of different analysts and the relative difficulty of different comparisons, however the relationship of such findings to the error rate in a specific case is problematic.
- The probability of error in a particular case may vary considerably depending on the difficulty of the comparison. Factors such as the quality of the prints, the amount of detail present, and whether the known print was selected based on its similarity to the latent will all be important.
- The probability of error will also depend on the examiner's implicit thresholds for deciding what to report. These thresholds may be affected by the perceived consequences of different types of error, which could in turn be influenced by contextual information about

the case. The thresholds may be different when examiners know they are being tested than when examiners are doing routine casework.

- The best way to study the performance of latent print examiners is to introduce known-source research samples into the routine flow of casework, so that examiners do not know their performance is being studied. Research of this type is much easier to do in laboratories that employ context management procedures because examiners in those laboratories are already blind to the source of the samples (which makes it easy for laboratory managers to give them research samples without their awareness that it is a research sample rather than an actual case).
- Research on examiner performance with various types of samples will help examiners realize their limitations and improve their skills by giving them feedback on their accuracy. Once laboratories develop the capability of introducing research samples into the normal flow of casework, it will be possible to design and carry out studies of a variety of important issues, such as how the number and kinds of features discernable in the latent print affect the accuracy of examiners' conclusions, the extent to which examiner accuracy is impaired by distortion of the prints, the effects of examiner characteristics (e.g., visual acuity, workload) on accuracy, and the effect of training on accuracy.

VI. In light of the existing scientific literature, what kind of statements might fingerprint examiners reasonably make in reports and testimony in order to appropriately convey both the strength and uncertainty associated with fingerprint evidence?

Latent print analysts traditionally claimed the ability to determine with certainty whether a latent print of sufficient quality was made by a particular finger. In reports and testimony, they routinely claimed to have “identified” the source of a print. “Identification” was a claim that the latent print and exemplar have ridge features that would be found in only one donor—i.e., “that the latent matches the exemplar and that it would not match exemplars from anyone else in the world” (Expert Working Group, 2012). Latent print analysts often claimed that their “identifications” were “100% accurate” with an error rate of zero (Cole, 2014, 2009, 2005).

Academic commentators have long derided such claims as scientifically indefensible (Cole, 2005). Saks and Koehler (2005) declared it a “myth” that examiners can discern by looking whether a particular fingerprint is unique. The claim that examiners can narrow the potential donors of a particular latent print to a single individual and thus infallibly identify the source of a print has also been rejected as indefensible by the NRC (NRC, 2009), the NIST Expert Working Group on Human Factors in Latent Print Analysis (Expert Working Group, 2012), and PCAST (PCAST, 2016), which noted “such statements are not scientifically defensible: all laboratory tests and feature comparison analyses have non-zero error rates.” (PCAST, 2016, p. 3).

Based on our review of the literature, we agree with these assessments. As noted earlier in this report, latent print examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but there is no scientific basis for estimating the number of people who could not be excluded and, consequently, no scientific basis for determining when the pool of possible sources is limited to a single person. Moreover, research on examiners’ accuracy when comparing known-source prints has provided ample evidence that false identifications can and do occur. Consequently, we have concluded that latent print examiners should avoid claiming that they can associate a latent print with a single source and should particularly avoid claiming or implying that they can do so infallibly, with 100% accuracy. Our understanding is that many if not most latent print examiners in the United States have already ceased making such claims.

It appears that latent print examiners are gradually coming to accept the need to moderate the claims that they make in reports and testimony, but many seem reluctant to abandon the claim that they can “identify” the source of a latent print (Cole, 2014). In order to rescue the use of this term, some have suggested that “identification” be re-defined so that it no longer constitutes a claim that the examiner can (and has) determined “that the latent matches the exemplar and that it would not match exemplars from anyone else in the world” (Expert Working Group, 2012, p. 67). Under the new definition, “identification” will instead mean merely that the examiner has decided that it is practically certain that the latent and exemplar have the same source. For example, in 2013 SWGFAST re-defined “identification” as:

...the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility (SWGFAST Document #10. 2013).

In our view, this approach fails to deal forthrightly with the uncertainty that currently exists about the rarity of any given friction ridge impression. As we have repeatedly mentioned, latent print examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, but *there is no scientific basis for determining how many people would not be excluded and no way to determine when the pool of possible sources is limited to a single person*. The SWGFAST approach avoids having the examiner make the unjustifiable claim that the pool of possible sources is limited to a single person by allowing examiners to claim that it is *practically certain* that the pool of possible sources is limited to a single person. In reality, there is not, at present, an adequate scientific basis for either claim. There is no basis for estimating the number of individuals who might be the source of a particular latent print. Hence, a latent print examiner has no more basis for concluding that the pool of possible sources is *probably* limited to a single person than for concluding it is *certainly* limited to a single person. The claim that “the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility” comes awfully close to the indefensible claim that the source of the latent print can be determined with certainty.

The DOJ recently made a similar attempt to mince the term “identification” in its “Proposed Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline” (available at: <https://www.justice.gov/archives/olp/file/861911/download> -- accessed September 12, 2017)²⁹ which states:

The examiner may state or imply that an identification is the determination that two friction ridge prints originated from the same source because there is sufficient quality and quantity of corresponding information such that the examiner would not expect to see that same arrangement of features repeated in another source. While an identification to the absolute exclusion of all others is not supported by research, studies have shown that as more reliable features are found in agreement, it becomes less likely to find that same arrangement of features in a print from another source.

The DOJ’s proposal also listed the following as statements “not approved for the use in Latent Print Examination Testimony and/or laboratory reports”:

²⁹ This document was proposed by the DOJ for the purpose of discussion and public comment. It has not been adopted as DOJ policy. Our understanding is that the DOJ is still considering whether to adopt a formal policy for reporting and testimony and what that policy might be.

1. An examiner may not state or imply that two friction ridge prints originated from the same source to the absolute exclusion of all other sources.
2. An examiner may not state or imply a level of certainty in his/her conclusion that is absolute or numerically calculated.
3. An examiner may not state or imply that the method used in performing a friction ridge print comparison has a zero error rate or is infallible.

We agree that the statements listed as “not approved” in the DOJ proposal are scientifically unwarranted and should not be made in reports or testimony. We disagree, however, with the language that the DOJ suggests that examiners use. In our view, the DOJ’s proposed language is also scientifically unwarranted.

Under the DOJ proposal examiners are forbidden from saying that the arrangement of features found in a latent print and in a matching exemplar would not be repeated in another source, as that statement would constitute the forbidden and unsupportable claim that the two friction ridge impressions originated from the same source to the exclusion of all others. But examiners are allowed to say that they “would not *expect* to see that same arrangement of features repeated in another source.” Examiners are thus allowed to make an assertion that is deemed scientifically unsupportable and improper so long as they hedge by saying that they “*expect*” that the assertion is true rather than saying outright that it is true.

In our view the DOJ proposal, like the SWGFAST approach, fails to acknowledge the extent of scientific uncertainty that exists regarding the rarity of particular friction ridge impressions. There is no scientific basis for estimating the number of individuals who might have a particular arrangement of ridge features. Consequently, there is no scientific basis on which latent print examiners might form expectations as to whether a particular set of features is likely or unlikely to be repeated. Any expectations latent print examiners may have on this matter rest on speculation and guesswork, rather than empirical evidence.

Commentators writing about forensic DNA evidence (e.g., Balding, 1985) have discussed how rare a DNA profile must be in order for one to be confident that it would not be repeated in the human population. Clearly the frequency must be less than $1/N$, where N is a number much larger than the human population. If, for example, the expected frequency of a DNA profile were 1 in 100 billion, then it *probably* would not be repeated in the human population (which is approximately 7.4 billion). But if the expected frequency of the profile were as high as 1 in 10 billion, then it *probably* would be repeated.³⁰

³⁰ With binomial data and a hypothesized match probability of 1 in 10 billion, the mean number of coincidental matches (duplications) in the human population would be 0.74. Using the Poisson approximation to the Binomial distribution, the expected probability of no duplications is $\exp(-\text{mean})$, which works out to 48%. Hence, the probability of at least one duplication would be 52%, making a duplication more probable than not. With a hypothesized match probability of 1 in 100 billion, the mean number of duplications would be 0.074, and the expected probability of no duplications would be 93%. The expected probability of one or more duplications would therefore be only 7%, which would justify the conclusion that a duplication *probably* would not occur.

The same issue arises in latent print analysis. The determination that the observable details of a fingerprint are “unlikely to be repeated” rests on the ability of latent print examiners to make extraordinarily precise estimates of the frequency of those details in the human population. Latent print examiners would need, for example, to be able to distinguish a set of details that occurs with a frequency of 1 in 100 billion or less from a set that occurs with a frequency of 1 in 10 billion or more. If latent print examiners cannot make such distinctions accurately, then they cannot determine whether a particular set of details is likely or unlikely to be repeated, and therefore have no basis for making the claim entailed by the DOJ’s proposed uniform reporting language (Zabell, 2005).

The problem with the DOJ’s proposed reporting language, then, is that there is no scientific evidence—none whatsoever—that latent print examiners have the ability to estimate with the required level of precision the frequency of the feature sets observable in latent prints in the human population. Because there is no proof that they can make such precise judgments accurately, there is no reason to believe that conclusions resting on their ability to make such precise judgments will be reliable or valid. Consequently, assertions about the probability that a feature set observed in a latent print will be repeated in another source are unwarranted. In our view, latent print examiners should not be making such statements.

Because there is no scientific basis for estimating the number of people who might be the source of a particular friction ridge print, we recommend that latent print examiners stop using the terms “identification” and “individualization.” These terms clearly imply that latent print examiners have the ability to single out the source of a print—to link it to a particular individual to the exclusion of any others. Attempts to rescue such terms by adding caveats, or by re-defining them to mean something other than what people ordinarily think they mean, are likely to be misleading. In our view these terms are beyond rescue and should be abandoned.

The PCAST report suggested that forensic scientists instead use the term “proposed identification” in order to “appropriately convey the examiner’s conclusion, along with the possibility that it might be wrong” (PCAST 2016, p. 45). While we agree that examiners should acknowledge the possibility of error, we do not think the problem with the term “identification” is solved by modifying it with the term “proposed.” The term *identification*, proposed or not, implies an ability to limit the source of a friction ridge print to a single individual. That is an ability that latent print examiners cannot justifiably claim to have.

Latent print examiners in Europe often take an entirely different approach to reporting the results of a fingerprint comparison. Many European forensic scientists follow a guideline for evaluative reporting issued by the European Network of Forensic Science Institutes (ENFSI). This guideline requires forensic scientists to indicate the strength of evaluative findings by assigning a likelihood ratio which “measures the strength of support the findings provide to discriminate between propositions of interest” (ENFSI, 2015, at 2.4). For a latent print examiner, there are typically two propositions of interest: (1) that the prints came from the same finger; and (2) that the prints came from different fingers. To assign a likelihood ratio, the latent print examiner estimates the probability that the observed degree of correspondence

between the prints would be present under each proposition; the ratio of those two probabilities is the likelihood ratio (see box below, p. 69).

For example, suppose the examiner sees a high degree of correspondence between a latent print and an exemplar and believes that this level of correspondence is nearly certain to be present if the prints are from the same finger but very unlikely to be present if the prints are from different fingers. To make the example concrete, suppose the examiner thinks the probability of the observed results is: 1.0 (certain) under the proposition that the print came from the same finger; and 0.000001 (1 chance in 1 million) under the proposition that the prints are from different fingers. The examiner would assign a likelihood ratio of 1 million. In reports and testimony, the examiner might then report this likelihood ratio directly by saying something like: “I believe the observed results are approximately 1 million times more probable if the prints I compared were made by the same finger than if they were made by two different fingers.” The examiner might also (or instead) make a non-numerical statement about the strength of the evidence that is based on the likelihood ratio. Several guidelines exist for stating the strength of a likelihood ratio in non-quantitative terms. Under Swedish guidelines, for example, a forensic scientist who assigns a likelihood ratio of 1 million or higher is allowed to say that the findings “...provide extremely strong support” for the favored proposition (that the prints have the same source) over the alternative proposition (that the prints have a different source) (Nordgaard et al., 2012). Alternatively, the expert can say the observed results “...are exceedingly more probable given” that the prints have the same source than if they have a different source. Under the guidelines issued by the United Kingdom-based Association of Forensic Service Providers (AFSP, 2009; see also Evett et al., 2000) an expert who assigns a likelihood ratio of 1 million would be allowed to say the results provided “extremely strong support” for the favored proposition.

One advantage of the ENFSI approach is that it avoids putting the latent print examiner in the position of having to decide whether the latent print evidence is strong enough to reach a particular conclusion, such as “identification” or “exclusion.” Under the SWGFAST and DOJ approaches, by contrast, the examiner must decide whether the evidence is strong enough to support the conclusion of individualization (Biedermann, Bozza and Taroni, 2016).³¹ This presumably requires the examiner to set a decision threshold, which may be influenced by the examiner’s assessment of the utility of the alternative decisions and the disutility of possible errors. An examiner who is worried about making a false identification, for example, would require a higher degree of similarity between the prints than an examiner who is worried about failure to make an identification (e.g., because it might allow a criminal to escape punishment). The ENFSI approach avoids the need to set a decision threshold because the examiner is not required to make any decisions. The examiner simply makes a statement about the strength of the evidence based on the estimated likelihood of the observed results under the alternative

³¹ The PCAST report appears to accept without discussion that forensic scientists should be reporting decisions about whether the evidence is strong enough to justify a categorical conclusion (e.g., “proposed identification;” “proposed exclusion”) rather than making statements about the strength of the evidence for distinguishing relevant hypotheses. In our view, it is important to consider the merits of a wider range of reporting options.

propositions. Whether the evidence is strong enough to justify any particular conclusion is a matter left to those who will rely on the latent print evidence, such as police, lawyers and jurors (Biedermann, Bozza and Taroni, 2008).

The disadvantage of the ENFSI approach is that it requires examiners to make subjective judgments regarding the probability of the observed data under the alternative propositions. As already discussed, there is no scientific evidence on whether latent print examiners can make such judgments accurately. How do we know, for example, that examiners can distinguish a set of friction ridge features that would be found in one person in 100,000 from a feature set that would be found in one person in 10 million? Because we have no scientific basis for determining the rarity of particular feature sets, there is no way to determine whether or not the subjective estimates of probability that this approach demands correspond with reality.

Of course, uncertainty about whether examiners' subjective probability estimates are warranted also affects the viability of other reporting methods, such as the SWGFAST and DOJ approaches. While these approaches do not require examiners to make their probability estimates explicit, they require that examiners be able to determine whether particular feature sets are rare enough that they are likely to be found in but one person in the world. That means that examiners must be able to distinguish feature sets that would occur so rarely that they are likely to be found once in the human population (e.g., a frequency of 1 in 100 billion or less) from feature sets that occur frequently enough that they are likely to be duplicated (e.g., a frequency of 1 in 10 billion or more). If examiners cannot make such distinctions accurately (and there is no scientific evidence that they can) then all of the approaches discussed thus far are unsupportable.

A more moderate approach to reporting the strength of latent print evidence was adopted by the DFSC of the Department of the Army in 2015. The DFSC approach required the examiner to make a subjective judgment about the rarity of a set of friction ridge features, but allows the examiner to report that judgment in a vague, imprecise manner. The DFSC reporting statement (initially proposed by Swofford, 2015) is as follows:

"The latent print on Exhibit ## and the record finger/palm prints bearing the name XXXX have corresponding ridge detail. The likelihood of observing this amount of correspondence when two impressions are made by different sources is considered extremely low" (Department of the Army, 2015).

The expert is essentially making a statement about the probability of finding the observed degree of correspondence under the proposition that the prints have a different source, which is similar to what experts do under the ENFSI approach.³² Rather than stating the estimate as a

³² The examiner is, in effect, making a statement about the denominator of the likelihood ratio.

number, however, the examiner simply states that the likelihood “is considered extremely low.”³³

Given the current uncertainty about the rarity of particular sets of ridge features, vague statements (e.g., “extremely low”) are easier to defend than precise estimates of probability (e.g., “one in 10 million”). For the same reason, non-numerical statements about strength-of-evidence, such as those allowed under the ENFSI guidelines are easier to defend than precise estimates of likelihood ratios. Because ridge features have been demonstrated to be highly variable, an examiner may well be justified in asserting that a particular feature set is rare, even though there is no basis for determining exactly how rare. And an examiner may well be justified in saying that a comparison provides “strong evidence” that the prints have a common source, even though there is no basis for determining exactly how strong. Whether the specific language proposed by the DFSC or the AFSP is ideal is open to debate, but statements of this type are preferable to claims about whether a particular feature set is “likely to be repeated.”

People often become confused about the meaning of forensic scientists’ statements about conditional probabilities. For example, forensic DNA analysts sometimes testify about the probability of observing a particular DNA profile if the source was a random person from some reference population, rather than the suspect. Unfortunately, people often misunderstand such testimony to be a statement about the probability, in light of the DNA evidence, that someone other than the suspect was the source of the DNA profile (Koehler, Chia and Lindsey, 1993; Thompson and Newman, 2015).³⁴ To reduce the likelihood of such misunderstandings, we recommend that latent print examiners avoid reporting estimates of the conditional probability of their observations “when two impressions are made by different sources.” We suggest that they instead frame their estimates as statements about how rare or common the observed profile might be in the human population.

As an alternative to the DFSC language, we suggest that examiners say something like the following:

³³ The phrase “is considered extremely low” invites the question—considered low by whom? The use of the passive voice in this phrase seems designed to suggest that the opinion offered is a consensus view, rather than the opinion of a single examiner. Yet it is likely that only one examiner, or perhaps one examiner with a second verifier, actually examined the ridge pattern in detail. We think it would be better to use language that makes it clear that the estimate of likelihood is the examiner’s opinion.

³⁴ The error is most easily explained formally. Let $H1$ be the hypothesis that two items have the same source, $H2$ be the alternative hypothesis that the two items have a different source, and E be evidence (observed by the forensic scientist) of the similarity of the two items. Forensic scientists often testify about $p(E|H2)$. For example, a latent print examiner might say, as in the DFSC reporting language, that the likelihood of the observed amount of correspondence is extremely low under the hypothesis that the prints have a different source. But lay people often misunderstand such testimony to be a statement about $p(H2|E)$. Hence, they might mistakenly interpret the DFSC language as a claim, in light of the fingerprint evidence, that there is an extremely low likelihood that the prints have a different source. We think the risk of such misunderstandings will be lower if latent print examiners avoid commenting on the conditional probability of the observed data under the hypotheses and instead comment about how rare or unusual the observed patterns might be.

"The latent print on Exhibit ## and the record fingerprint bearing the name XXXX have a great deal of corresponding ridge detail with no differences that would indicate they were made by different fingers. There is no way to determine how many other people might have a finger with a corresponding set of ridge features, but it is my opinion that this set of features would be unusual."

When presenting such statements, examiners should also acknowledge that errors are possible and have occurred in studies of latent print accuracy. We agree with the PCAST report that latent print examiners should be prepared to discuss forthrightly the results of research studies (such as Ulery et al., 2011, and Pacheco et al., 2014) that tested the accuracy of latent print examiners on realistic known-source samples.

Quantitative Approaches

The 2009 NRC report called for the development of "quantifiable measures" of the strength of forensic evidence. This call for quantification based on empirical research was echoed by the PCAST report, which declared: "The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of 'judgment.' It is an empirical matter for which only empirical evidence is relevant" (PCAST, 2016, p. 6). We agree that there is a need to quantify the strength of latent print evidence so that it is not left a matter of vague subjective judgment. If there were an adequate quantitative measure of the strength of fingerprint evidence, it could be presented in reports and testimony in place of statements based on subjective judgment.

Given that AFIS use quantitative algorithms for comparing prints, it might seem that they could be used to generate quantitative estimates of the strength of fingerprint evidence. At present, however, AFIS are not set up to provide such estimates. They provide similarity scores that are not easily translatable to a statistical estimate of probative value (strength) of fingerprint evidence. Details of how these similarity scores are calculated are generally treated as trade secrets by AFIS vendors, making the meaning and value of the scores difficult to evaluate. It might be possible for AFIS systems of the future to generate meaningful statistical measures of evidentiary strength called likelihood ratios, but additional research will be needed to determine whether this approach to estimating the value of print comparisons will be viable.

Researchers have made a number of efforts to quantify the weight of fingerprint evidence using models that are open for public examination (Anthonioz and Champod, 2014b; Abraham et al., 2013; Lennard, 2014; Neumann, Evett and Skerrett, 2012; Stoney, 2001). So far, these studies have concentrated solely on minutiae, which are one of several sources of information used by latent print examiners. Recent efforts by Neumann and his colleagues to explore this approach have been particularly noteworthy (Neumann, Evett and Skerrett, 2012; Neumann et al., 2015). The goal of this work is to describe the strength of fingerprint evidence for proving prints have a common source, using a statistical measure of evidentiary strength called a likelihood ratio (see box on the following page). The researchers hope that forensic scientists might evidentially use

these models to support claims about the value of particular print comparisons, or even to generate quantitative estimates of weight that could be presented to juries.

In order for quantitative estimates of this kind to be meaningful, however, the statistical models on which they are based must accurately represent the degree of inter-finger and intra-finger variability in the prints that examiners encounter in casework, taking account of the range of features, quality, and degree of similarity. Because databases suitable for such research have not yet been developed, researchers have not been able to make empirically-based estimates of variability and have relied instead on ad hoc models based on estimates that are essentially guesses about the degree of variability. Critics have noted (Kadane, 2012; Stern, 2012), and the researchers have conceded, that these models are “only a very ad hoc approximation of the likelihood ratio” (Neumann et al., 2015) While this research certainly supports the conclusion that fingerprint evidence can, in general, have substantial probative value, the research has not reached a point at which reliable quantitative estimates can be generated for the weight of particular fingerprint comparisons. It may eventually be possible to convert latent print analysis from a discipline that relies primarily on subjective analysis of human examiners to one based on more objective quantitative measurements and statistical modeling, but that is a long-term project. The article of Anthonioz and Champod (2014b) suggests a possible way forward toward this goal.

Likelihood ratios are statements about the relative probability of observed data under alternative propositions about the source of the data (Robertson and Vignaux, 1995; Schum, 1994). A likelihood ratio for a fingerprint comparison will typically state how much more (or less) probable the observed similarities are if the latent print in question was made by the finger of a specified suspect than if it was made by a different person drawn at random from some reference population. Research on lay perceptions of statistical evidence suggests that people often have difficulty understanding and using likelihood ratios appropriately (Martire et al., 2013, 2014; Thompson and Newman, 2015). Nevertheless, likelihood ratios have been used to characterize other types of forensic evidence (Aitken and Taroni, 2004). They are often used by statisticians to characterize the strength of evidence for proving a proposition. And there may be viable ways to explain them to police lawyers and jurors.

One approach to explaining likelihood ratios is the use of verbal statements of their strength. In Europe, the Association of Forensic Service Providers (AFSP) has proposed a set of statements that can be used to characterize likelihood ratios of various magnitudes (AFSP, 2009; see also Evett et al., 2000). For example, if the likelihood ratio exceeds 10,000 it is suggested that the forensic scientist say that the evidence provides “very strong evidence to support” for the supported proposition; if the likelihood ratio is less than 10, it is suggested that the forensic scientist say it provides “limited support” for the propositions. Another option for explaining likelihood ratios might be to restate them in a manner that makes them comparable to match probabilities or “random match equivalents” (Thompson, 2012). Because likelihood ratios can be used in Bayesian analysis to specify the degree to which an initial or “prior” estimate of the odds of some proposition should be updated in light of new evidence, another approach to explaining them would be for the examiner to demonstrate how much various estimates of “prior” odds should be updated in light of the fingerprint evidence. (Robertson and Vignaux, 1995). For example, the expert might say “if you previously thought the odds of the defendant being the source of the print were X , then in light of this likelihood ratio you should now think the odds are Y ”, for various values of X . A final alternative would require examiners to make and present their own calculation of the posterior odds that the prints have the same source by combining the likelihood ratio with their personal estimates of (or assumption about) the prior odds. A disadvantage of this last option is that it requires examiners to consider (or make assumptions about) matters beyond their scientific expertise—matters that arguably should be left to police, lawyers and jurors, rather than forensic scientists (Champod, 2009; Thompson, 2016; Thompson et al., 2013). Additional empirical work on how lay people respond to likelihood ratios, when explained in various ways, would be helpful for evaluating the advantages and disadvantages of the possible ways likelihood ratios might be explained in reports and testimony.

In the short-term, the best option for quantifying the weight of fingerprint evidence is to rely on data generated by research on examiners' performance when comparing known-source prints. Examiners could simply report the rate of false identifications and false exclusions that have occurred in such studies. This is the approach to reporting recommended by the PCAST report:

The forensic examiner should report the overall false positive rate and sensitivity for the method established in the studies of foundational validity and should demonstrate that the samples used in the foundational studies are relevant to the facts of the case (p. 56).

The PCAST report made two specific recommendations regarding the way in which false positive rates should be calculated and reported. First, to take account of the imprecision of such estimates due to sample size limitations in the underlying studies, PCAST recommended that examiners report the "upper 95% one-sided confidence bound"³⁵ (PCAST, 2016, p. 51):

For example, if a study found no errors in 100 tests, it would be misleading to tell a jury that the error rate was 0 percent. In fact, if the tests are independent, the upper 95 percent confidence bound for the true error rate is 3.0 percent. Accordingly, a jury should be told that the error rate could be as high as 3.0 percent (that is, 1 in 33) (p. 51, note 116).

Second, PCAST recommended that "inconclusive" findings be ignored when calculating the rate of true positive and false positive results in research studies, noting that this approach is appropriate "because evidence used against a defendant will typically be based on conclusive rather than inconclusive examinations" (PCAST p. 51).

Until further research on examiner performance is completed, PCAST recommended that error rate estimates be based on the two studies it recognized as properly designed to address the foundational validity of latent prints (Pacheco et al., 2014; Ulery et al., 2011):

Overall, it would be appropriate to inform jurors that (1) only two properly designed studies of the accuracy of latent fingerprint analysis have been conducted and (2) these studies found false positive rates that could be as high as 1 in 306 in one study and 1 in 18 in the other study (PCAST, p. 96).

PCAST also recommended that examiners disclose whether the underlying studies were conducted in a test-blind manner, as "non-blind proficiency tests are likely to overestimate the accuracy because the examiners knew they were being tested" (PCAST, p. 59).

Deriving statistical conclusions of this kind from the testing program recommended above may well be the most practical option, at least for the near future, for quantifying the value of fingerprint evidence. Such estimates will be most meaningful if we can reliably measure the

³⁵ Appendix A of the PCAST report discusses several alternative ways to compute a one-sided upper confidence bound and finds all of these alternative acceptable.

quality of latent prints and the difficulty of comparing them. As noted earlier, additional research is needed before we can hope to make reliable assessment of these important variables. But the results of existing studies are helpful for at least giving a broad indication of the accuracy of latent print analysis. Latent print examiners should be careful to acknowledge the limitations of these studies, but we agree with the PCAST report that latent print examiners should be prepared to describe and present the results of these studies to juries.

Caveats

Public perceptions of latent print examination have undoubtedly been shaped by decades of overstatement. One of the problems that examiners now face when attempting to convey a more realistic and appropriate sense of the value of latent print evidence is that people generally think a reported association between a latent and reference print constitutes a virtually infallible identification. In our view, latent print examiners should take affirmative steps, when reporting their findings, to address these common misconceptions. We recommend that latent print examiners include in their reports a series of specific caveats about the limitations of the discipline. Statements like the following would be appropriate:

1. Latent print examination allows examiners to draw conclusions about whether two friction ridge impressions could have originated from the same source. These conclusions are opinions, they are not facts.
2. It is not possible for a latent print examiner to determine that two friction ridge prints originated from the same source to the absolute exclusion of all other sources. A latent print examiner may be able to exclude a substantial proportion of the human population as the source of a latent print, but it is not possible to determine how many people would not be excluded, nor is it possible to determine when the pool of possible sources is limited to a single person.
3. Studies have shown that latent print examiners are highly accurate in associating latent prints with reference prints known to be from the same source, and in excluding reference prints known to be from a different source. But latent print examination is not infallible. Both false associations and false exclusions have occurred in studies of examiner performance and in actual cases.

Conclusions

- Latent print examiners traditionally claimed to be able to “identify” the source of a latent print with 100% accuracy. These claims were clearly overstated and are now widely recognized as indefensible. While latent print examiners may well be able to exclude the preponderance of the human population as possible sources of a latent print, there is no scientific basis for estimating the number of people who could not be excluded and, consequently, no scientific basis for determining when the pool of possible sources is limited to a single person. Moreover, research on examiners’ accuracy when comparing

known-source prints has provided ample evidence that false identifications can and do occur.

- SWGFAST and the DOJ published documents directing latent print examiners to continue to use the term “identification” in reports and testimony but add qualifying language that acknowledges an element of uncertainty. In our view, these proposals fail to deal forthrightly with the level of uncertainty that exists in latent print examination. In our view, the proposed reporting language allows examiners to make claims that cannot be justified scientifically.
- Guidelines for evaluative reporting recommended by the European Network of Forensic Science Institutes (ENFSI) require examiners to estimate likelihood ratios. An advantage of this approach is that it avoids the need for an expert to decide whether the observed similarities and differences between prints are sufficient to justify a particular categorical conclusion. A disadvantage (shared with other subjective methods) is that it requires examiners to make subjective estimates of the probability of the observed data³⁶ under alternative hypotheses about the source of the prints. Whether examiners can make such judgments reliably and accurately is, at present, unknown.
- The DFSC of the Department of the Army has suggested a more moderate approach to reporting the strength of latent print evidence. Although we have some concerns about the specific language suggested by the DFSC, the approach taken by DFSC is preferable to those suggested by SWGFAST and DOJ because the statements it allows are easier to defend scientifically. We propose some alternative language that we believe will be easier for lay people to understand.
- The 2009 NRC report called for the development of “quantifiable measures of the uncertainty in the conclusions of forensic analyses.” We agree that there is a need to quantify the strength of latent print evidence so that it is not left a matter of subjective judgment, however, the existing scientific literature does not, at present, provide an adequate basis for quantitative estimates of the accuracy of the various determinations latent print examiners make on the basis of latent print evidence.
- In the future, it might be possible to rely on AFIS to generate quantitative estimates of the probative value of a particular print comparison. At present, AFIS systems are not designed to provide such estimates. While it might be possible for AFIS systems of the future to generate meaningful statistical measures of evidentiary strength, additional research will be needed to determine whether this approach is viable.
- A number of researchers have attempted to use mathematical models to describe the strength of fingerprint evidence for proving that prints have a common source. This research is promising and deserves further support, although the models are not yet ready for use in the courtroom.
- In the short-term, studies of the type discussed previously in which research samples are introduced into the routine flow of casework could provide a valuable source of data on the strength of fingerprint evidence but this is problematic in agencies in which handling of

³⁶ The data that a latent print examiner may observe when comparing two fingerprint impressions includes information at all levels of detail potentially observable in the impression. For discussion of the various features and levels of detail that latent print examiners may see in print impressions, see Appendix B.

physical evidence by examiners is necessary. The research could potentially allow examiners to report the rates of false positives and false negatives that have occurred when latent print examiners made similar comparisons involving known-source prints. Relative error-rate estimates derived from such data will inevitably be approximate and imperfect due to uncertainty about whether the comparison being evaluated in a particular instance posed the same level of difficulty as the comparisons examined in the research. Such data could nevertheless provide important insights into how high, or how low, the rate of error might be in practice.

- Members of the public are likely to hold misconceptions about latent print examination that have been shaped by decades of overstatement by latent print examiners. To combat these misperceptions, latent print examiners should include specific caveats in reports that acknowledge the limitations of the discipline. They should acknowledge: (1) that the conclusion being reported are opinions rather than facts (as in all pattern-matching disciplines), (2) that it is not possible for a latent print examiner to determine that two friction ridge impressions originated from the same source to the exclusion of all others; and (3) that errors have occurred in studies of the accuracy of latent print examination.

REFERENCES

1. Abraham J., Champod C., Lennard C. and Roux C. (2013). Modern Statistical Models for Forensic Fingerprint Examinations: A Critical Review. *Forensic Science International*, Vol. 232, pp. 131-150.
2. Abraham J., Champod C., Lennard C. and Roux C. (2013). Spatial Analysis of Corresponding Fingerprint Features from Match and Close Non-Match Populations. *Forensic Science International*, Vol. 230, pp. 87-98.
3. Aitken C. and Taroni F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd Ed. Chichester: John Wiley & Sons, 540 pages.
4. American Academy of Forensic Sciences (AAFS). (2013). www.aafs.org/sites/default/files/pdf/National%20Commission%20on%20Forensic%20Science.pdf.
5. Anthonioz A. and Champod C. (2014a). Integration of Pore Features into the Evaluation of Fingerprint Evidence. *Journal of Forensic Sciences*, Vol. 59, No. 1, pp. 82-93.
6. Anthonioz A. and Champod C. (2014b). Evidence Evaluation in Fingerprint Comparison and Automated Fingerprint Identification Systems—Modeling Between Finger Variability. *Forensic Science International*, Vol. 235, pp. 86-101.
7. Aronson J.D. (2007). *Genetic Witness: Science, Law and Controversy in the Making of DNA Profiling*. Rutgers University Press.
8. Arora S., Kai C. and Jain A.K. (2015). Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups. *Proceedings of the 8th International Conference on Biometrics (ICB)*, Phuket, Thailand, May 2015.
9. Association of Forensic Science Providers (2009). Standards for the Formulation of Evaluative Forensic Science Expert Opinion. *Science & Justice*, Vol. 49, No. 3, pp. 161-164.
10. Babler W.J. (1991). Embryologic Development of Epidermal Ridges and Their Configurations. *Dermatoglyphics- Science in Transition*, Vol. 27, No. 2, pp. 95-112.
11. Barnes J.G. (2011). Chapter 1: History, in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs, NCJ 225320.
12. Barros R.M., Faria B.E.F. and Kuckelhaus S.A.S. (2013). Morphometry of Latent Palmprints as a Function of Time. *Science & Justice*, Vol. 53, No. 1, pp. 402-208.

13. Biedermann A., Bozza S. and Taroni F. (2008). Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications. *Forensic Science International* Vol. 177, pp. 120-132.
14. Biedermann A., Bozza S. and Taroni F. (2016). The Decisionalization of Individualization. *Forensic Science International*, Vol. 266, pp. 29-38.
15. Budowle B., Bottrell M.C., Bunch S.G., et al. (2009). A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement. *Journal of Forensic Sciences*, Vol. 54, No. 4, pp. 798-809.
16. Butt L. (2013). The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions- Commentary by a Forensic Examiner. *Journal of Applied Research in Memory and Cognition*, Vol. 2, No. 2, pp. 59-60.
17. Cappelli R., Maio D. and Maltoni D. (2001). Modelling Plastic Distortion in Fingerprint Images, in *Advances in Pattern Recognition*. ICAPR 2001, S Singh, N Murshed and W Kropatsch, Eds., Vol. 2013, pp. 371-378.
18. Champod C. (2009). Friction Ridge Examination (Fingerprints): Interpretation of, in *Wiley Encyclopedia of Forensic Sciences*. A Moenssens & A Jamieson, Eds, pp. 1508-1511.
19. Champod C., Lennard C., Margot P. and Stoilovic M. (2016). *Fingerprints and Other Skin Impressions*, Second Edition. CRC Press, Boca Raton, FL.
20. Charlton D. (2013). Standards to Avoid Bias in Fingerprint Examination? Are Such Standards Doomed to be Based on Fiscal Expediency? *Journal of Applied Research in Memory and Cognition*, Vol. 2, pp. 71-72. DOI: 10.1016/j.jarmac.2013.01.009.
21. Charter of the Subcommittee on Forensic Science. (2009). <https://www.whitehouse.gov/sites/default/files/microsites/ostp/forensic-science-subcommittee-charter.pdf>
22. Charter of the Subcommittee on Forensic Science. (2012). https://www.whitehouse.gov/sites/default/files/sofs_charter_2012_signed.pdf
23. Charter of the U.S. Department of Justice National Commission on Forensic Science. (2014). <http://www.justice.gov/sites/default/files/ncfs/legacy/2014/05/13/ncfs-charter.pdf>.
24. Chen Y. and Jain A.K. (2009). *Beyond Minutiae: A Fingerprint Individuality Model with Pattern, Ridge and Pore Features*, in *Advances in Biometrics: Third International Conference*, ICB 2009. LNCS 5558, M. Tistarelli and M.S. Nixon, Eds. Berlin- Heidelberg: Springer Verlag, pp. 523-533.

25. Cole S.A. (2004). Grandfathering Evidence: Fingerprint Admissibility Ruling from Jennings to Llera Plaza and Back Again. *American Criminal Law Review*, Vol. 41, No. 3, pp. 1189-1276.
26. Cole S.A. (2005). More Than Zero: Accounting for Error in Latent Fingerprint Identification. *Journal of Criminal Law & Criminology*, Vol. 95, No. 3, pp. 985-1078.
27. Cole S.A. (2006). The Prevalence and Potential Causes of Wrongful Conviction by Fingerprint. *Golden Gate University Law Review*, Vol. 37, pp. 39-105.
28. Cole S.A. (2009). Forensics without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification. *Law, Probability and Risk*, Vol. 8, No. 3, pp. 233-255.
29. Cole S.A. (2013a). Forensic Science Reform: Out of the Laboratory and into the Crime Scene. *Texas Law Review*. See Also, Vol. 91 (2013), pp. 123-136.
30. Cole S.A. (2013b). Implementing Counter-Measures Against Confirmation Bias in Forensic Science. *Journal of Applied Research in Memory and Cognition*, Vol. 2, No. 1, pp. 61-62.
31. Cole S.A. (2014). Individualization is Dead, Long Live Individualization! Reforms of Reporting Practices for Fingerprint Analysis in the United States. *Law, Probability and Risk*, Vol. 13, No. 2, pp. 1-34.
32. Cole S.A., Welling M., Dioso-Villa R. and Carpenter R. (2008). Beyond the Individuality of Fingerprints: A Measure of Simulated Computer Latent Print Source Attribution Accuracy. *Law, Probability and Risk*, Vol. 7, No. 3, pp. 165-189.
33. Colletto G.M.D.D. and Zimberknopf S. (1987). Diagnosis of Twin Zygosity by Dermatoglyphic Index. *Brasil Journal of Genetics*, Vol. 10, No. 2, pp. 289-299.
34. Cutro Sr. B.T. (2011). Chapter 4: Recording Living and Postmortem Friction Ridge Exemplars, in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs. NCJ 225320.
35. Dass S.C., et al. (2009). Individuality of Fingerprints, in *Encyclopedia of Biometrics*. SZ Li and A Jain, Eds. New York: Springer Verlag, pp. 741-751.
36. Dass S.C. (2010). Assessing Fingerprint Individuality in Presence of Noisy Minutiae. *IEEE Transactions on Information Forensics and Security*, Vol. 5, No. 1, pp. 62-70.
37. Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993).
38. Department of the Army (2015). Information Paper Re: Use of the Term "Identification" in *Latent Print Technical Reports*, November 3rd.

39. Dresch-Langley B. and Reeves A. (2011). Simultaneous Brightness and Apparent Depth from True Colors on Grey: Chevreul Revisited. *Seeing and Perceiving*, Vol. 25, No. 6, pp. 597-618.
40. Dror I.E. (2011). The Paradox of Human Expertise: Why Experts Get it Wrong, in *The Paradoxical Brain*. N Kapur, Editor. Cambridge, UK: Cambridge University Press, pp. 177-188.
41. Dror I.E. (2012). Expectations, Contextual Information, and Other Cognitive Influences in Forensic Laboratories. *Science & Justice*, Vol. 52, No. 2, pp. 132.
42. Dror I.E. and Charlton D. (2006). Why Experts Make Errors. *Journal of Forensic Identification*, Vol. 56, pp. 600-616.
43. Dror I. E., Charlton D. and Peron A. (2006). Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications. *Forensic Science International*, Vol. 156, pp. 174-178.
44. Dror I.E. and Cole S.A (2010). The Vision in 'Blind' Justice: Expert Perception, Judgment, and Visual Cognition in Forensic Pattern Recognition. *Psychonomic Bulletin & Review*, Vol. 17, No. 2, pp. 161-167.
45. Dror I.E. and Hampikian G. (2011). Subjectivity and Bias in Forensic DNA Mixture Interpretation. *Science & Justice*, Vol. 51, No. 4, pp. 204-208.
46. Dror I.E. and Rosenthal R. (2008). Meta-Analytically Quantifying the Reliability and Biasability of Forensic Experts. *Journal of Forensic Sciences*, Vol. 53, No. 4, pp. 900-903.
47. Dror I.E., Thompson W.C., Meissner C.A., et al. (2015). Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making. *Journal of Forensic Sciences*, Vol. 60, No. 4, pp. 1111-1112.
48. Dvornychenko V.N. and Garris M.D. (2006). *Summary of NIST Latent Fingerprint Testing Workshop*. US Department of Commerce, National Institute of Standards and Technology, NISTIR 7377.
49. Egli N.M., Champod C. and Margot P. (2007). Evidence Evaluation in Fingerprint Comparison and Automated Fingerprint Identification Systems – Modeling Within Finger Variability. *Forensic Science International*, Vol. 167, No. 2-3, pp. 189-195.
50. Eldridge H. (2017). The Shifting Landscape of Latent Print Testimony: An American Perspective. *Journal of Forensic Science and Medicine*, Vol 3, pp. 72-81.
51. European Network of Forensic Science Institutes (ENFSI) (2015). Best Practice Manual for Fingerprint Examination. ENFSI-BPM-FIN-01, Version 1, November 2015. Available at: <http://enfsi.eu/wp-content/uploads/2016/09/6. fingerprint examination 0.pdf>

52. Evett I., Jackson G., Lambert J.A. and McCrossan S. (2000). The Impact of the Principles of Evidence Interpretation on the Structure and Content of Statements. *Science & Justice*, Vol. 40, pp. 233-251.
53. Evett I.W. and Williams R.L. (2015). A Review of the Sixteen Points Fingerprint Standard in England and Wales. *Journal of Forensic Identification*, Vol. 65, No. 4, pp. 557-580. [Reprint of 1995 version in *Fingerprint Whorld*].
54. Expert Working Group on Human Factors in Latent Print Analysis. (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.
55. Fagert M. and Morris K. (2015). Quantifying the Limits of Fingerprint Variability. *Forensic Science International*, Vol. 254, pp. 87-99.
56. Feng J., Shi Y. and Zhou J. (2012). Robust and Efficient Algorithms for Separating Latent Overlapped Fingerprints. *IEEE Transaction on Information Forensics and Security*, Vol. 7, No. 5, pp. 1498-1510.
57. Found B. and Ganas J. (2013). The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework. *Science & Justice*, Vol. 53, No. 2, pp. 154-158.
58. Giannelli P.C. (2007). Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs. *North Carolina Law Review*, Vol. 86, p. 163.
59. Grieve D.L. (1990). The Identification Process: Traditions in Training. *Journal of Forensic Identification*, Vol. 40, No. 4, pp. 195-213.
60. Guan H., Dienstfrey A.M., Theofanos M.F. and Stanton B.C. (2014). A Measurement Metric for Forensic Latent Fingerprint PreProcessing. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 8017.
61. Gutiérrez E., Galera V., Martínez J.M. and Alonso C. (2007). Biological Variability of the Minutiae in the Fingerprints of a Sample of the Spanish Population. *Forensic Science International*, Vol. 172, No. 2-3, pp. 98-105.
62. Gutiérrez-Redomero E., Alonso-Rodríguez C., Hernández-Hurtado L.E. and Rodríguez-Villalba J.L. (2011). Distribution of the Minutiae in the Fingerprints of a Sample of the Spanish Population. *Forensic Science International*, Vol. 208, No. 1-3, pp. 79-90.
63. Gutowski S. (2006). Error Rates in Fingerprint Examination: The View in 2006. *Forensic Bulletin*, pp. 18-19.

64. Haber R.N. and Haber L. (2014). Experimental Results of Fingerprint Comparison Validity and Reliability: A Review and Critical Analysis. *Science & Justice*, Vol. 54, No. 5, pp. 375-389.
65. Haylock S.E. (1987). Abnormalities in Dermatoglyphics. *Fingerprint Whorld*, Vol. 12, No. 47, pp. 71-75.
66. Helsloot I. and Groenendaal J. (2011). Naturalistic Decision Making in Forensic Science: Toward a Better Understanding of Decision Making by Forensic Team Leaders. *Journal of Forensic Sciences*, Vol. 56, No. 4, pp. 890-897.
67. Hicklin R.A., Buscaglia J., Roberts M.A., et al. (2011). Latent Fingerprint Quality: A Survey of Examiners. *Journal of Forensic Identification*, Vol. 61, No. 4, pp. 385-418.
68. Hicklin R.A., Buscaglia J., and Roberts M.A. (2013). Assessing the Clarity of Friction Ridge Impressions. *Forensic Science International*, Vol. 226, No. 1, pp. 106-117.
69. H.R. Report No. 109-272. (2005). Making Appropriations for Science, The Departments of State, Justice, and Commerce, and Related Agencies for the Fiscal Year Ending September 30, 2006, and for other Purposes (Conference Report), Page 121.
<https://www.congress.gov/109/crpt/hrpt272/CRPT-109hrpt272.pdf>.
70. Hong S., Hong I., Han A., et al. (2015). A New Method of Artificial Latent Fingerprint Creation Using Artificial Sweat and Inkjet Printer. *Forensic Science International*, Vol. 257, pp. 403-408.
71. Indovina M., Hicklin R.A. and Kiebusinski G.I. (2011). ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #1]. U.S. Department of Commerce, National Institute of Standards and Technology, Washington DC, NISTIR 7775.
72. Indovina M., Dvornychenko V.N., Hicklin R.A. and Kiebusinski G.I. (2012). ELFT- EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #2]. U.S. Department of Commerce, National Institute of Standards and Technology, Washington DC, NISTIR 7859.
73. Jain A.K. (2005). "On the Uniqueness of Fingerprints." Presentation, NAS Sackler Forensic Science Colloquium, Washington DC:
http://biometrics.cse.msu.edu/Presentations/AnilJain_UniquenessOfFingerprints_NAS05.pdf
74. Jain A.K. (2011). *Automatic Fingerprint Matching Using Extended Feature Set*. National Institute of Justice, 235577.
75. Jain A.K. and Feng J. (2011). Latent Fingerprint Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, pp. 88-100.

76. Jain A.K., Prabhakar S. and Pankanti S. (2000). Can Identical Twins be Discriminated Based on Fingerprints? Technical Report MSU-CSE-00-23, Department of Computer Science, Michigan State University, East Lansing, Michigan.
77. Jain A.K., Prabhakar S. and Pankanti S. (2002). On the Similarity of Identical Twin Fingerprints. *Pattern Recognition*, Vol. 35, No. 11, pp. 2653-2663.
78. Kadane J. (2012). Comment on Neumann et al. *Journal of the Royal Statistical Society, Series A*, 175 (part 2) 404.
79. Kalka N.D. and Hicklin R.A. (2014). On Relative Distortion in Fingerprint Comparison. *Forensic Science International*, Vol. 244, pp. 78-84.
80. Kassin S.M., Dror I.E. and Kukucka J. (2013). The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions. *Journal of Applied Research in Memory and Cognition*, Vol. 2, No. 1, pp. 42-52.
81. Kaye, D.A. (2010). *The Double-Helix and the Law of Evidence*. Harvard University Press.
82. Kellman P.J., et al. (2014). Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates through Understanding and Predicting Difficulty. *PLoS ONE*, Vol. 9, No. 5, pp. 1-14.85.
83. Koehler J.J. (2017). Intuitive Error Rate Estimates for the Forensic Sciences (2016). *Jurimetrics*, Vol. 57, No. 2, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=2817443> or <http://dx.doi.org/10.2139/ssrn.2817443>
84. Koehler J.J. (forthcoming). Forensics or Fauxrensic? Ascertaining Accuracy in the Forensic Sciences, ARIZ. ST. L. J. 49. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255.
85. Koehler J.J., Chia A. and Lindsey S. (1993). The random match probability in DNA evidence: Irrelevant and prejudicial? *Jurimetrics*, Vol. 35, No. 2, pp. 201-219.
86. Krane D.E., Ford S., Gilder J.R., et al. (2008). Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation. *Journal of Forensic Sciences*, Vol. 53, No. 4, pp. 1006-1007.
87. Kücken M.U. (2007). Models for Fingerprint Pattern Formation. *Forensic Science International*, Vol. 171, No. 2-3, pp. 85-96.
88. Kumbhani H.K. (2007). Dermatoglyphics: A Review. *Anthropology Today: Trends, Scope and Applications*, No. 3, pp. 285-295.

89. Laber T., Kish P., Taylor M., et al. (2014). *Reliability Assessment of Current Methods in Bloodstain Pattern Analysis*. National Institute of Justice, US Department of Justice.
90. Langenburg G. (2009). A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability and Biasability of Conclusions Resulting from the ACE-V Process. *Journal of Forensic Identification*, Vol. 59, No. 2, pp. 219-257.
91. Langenburg G. (2011). Chapter 14: Scientific Research Supporting the Foundations of Friction Ridge Examinations, in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs, NCJ 225320.
92. Langenburg G., Champod C. and Genessay T. (2012). Informing the Judgements of Fingerprint Analysts Using Quality Metric and Statistical Assessment Tools. *Forensic Science International*, Vol. 219, pp. 183-198.
93. Langenburg G. Champod C. and Wertheim P. (2009). Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons. *Journal of Forensic Sciences*, Vol. 54, No. 3, pp. 571-582.
94. Leadbetter M. (2007). Letter to the Editor. *Fingerprint Whorld*, September Vol. 231.
95. Lempert R.O. (1976). Modeling Relevance. *Michigan Law Review*, Vol. 75, pp. 1021.
96. Lennard C. (2014). Fingerprint detection and identification: Current research efforts. *Australian Journal of Forensic Sciences*, Vol. 46, No. 3, pp. 293-303.
97. Levine R. (1986). *Ethics and Regulation of Clinical Research*, 2nd Edition. Yale University Press.
98. Li S.Z. and Jain A.K., Editors. (2009). *Encyclopedia of Biometrics*. Springer Science & Business Media.
99. Library of Congress. (2012). <https://www.congress.gov/bill/112th-congress/Senate-bill/3378>.
100. Liu S., Champod C., Wu J. and Luo Y. (2015). Study on Accuracy of Judgments by Chinese Fingerprint Examiners. *Journal of Forensic Science and Medicine*, Vol. 1, No. 1, pp. 33-37.
101. Liu Y. and Srihari S. (2009). A Computational Discriminability Analysis on Twin Fingerprints, in *Computational Forensics*, pp. 43-54.
102. Loesch D.Z. (1983). *Quantitative Dermatoglyphics: Classification, Genetics, and Pathology*. Oxford University Press, Oxford.

103. MacArthur J.W. (1938). Reliability of Dermatoglyphics in Twin Diagnosis. *Human Biology*, Vol. 10, No. 1, pp. 12- 35.
104. Maceo A.V. (2005). The Basis for the Uniqueness and Persistence of Scars in the Friction Ridge Skin. *Fingerprint Whorld*, Vol. 31, No. 121, pp. 147-161.
105. Maceo A.V. (2009). Qualitative Assessment of Skin Deformation: A Pilot Study. *Journal of Forensic Identification*, Vol. 59, No. 4, pp. 390-440.
106. Maceo A.V. (2011). Chapter 2: Anatomy and Physiology of Adult Friction Ridge Skin, in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs, NCJ 225320.
107. Martire K.A., Kemp R.I., Watkins I., et al. (2013). The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect. *Law and Human Behavior*, Vol. 37, No. 3, pp. 197. doi:10.1037/lhb0000027.
108. Martire K.A., Kemp R.I., Sayle M. and Newell B.R. (2014). On the Interpretation of Likelihood Ratios in Forensic Science Evidence: Presentation Formats and the Weak Evidence Effect. *Forensic Science International*, Vol. 240, pp. 61-68. doi:10.1016/j.forsciint.2014.04.005.
109. Mavalwala J. (1977). *Dermatoglyphics: An International Bibliography*. The Hague: Mouton; Chicago: distributed in the USA and Canada by Aldine. ISBN 0202900517.
110. Miller L.S. (1984). Bias Among Forensic Document Examiners: A Need for Procedural Change. *Journal of Police Science and Administration*, Vol. 12, No. 4, pp. 407-411.
111. Modi S.K., Elliott S.J., Whetsone J. and Kim H. (2007). Impact of Age Groups on Fingerprint Recognition Performance. *Automatic Identification Advanced Technologies, 2007 IEEE Workshop*, pp. 19-23.
112. Morrison G.S. (2011). Measuring the Validity and Reliability of Forensic Likelihood-Ratio Systems. *Science & Justice*, Vol. 51, pp. 91–98. <http://dx.doi.org/10.1016/j.scijus.2011.03.002>
113. Morrison G.S., Kaye D.H., Balding D.J., et al. (2016). A comment on the PCAST report: Skip the “match”/“non-match” stage. *Forensic Science International*. <http://dx.doi.org/10.1016/j.forsciint.2016.10.018>
114. Morrison G.S. and Stoel R.D. (2014). Forensic Strength of Evidence Statements Should Preferably be Likelihood Ratios Calculated Using Relevant Data, Quantitative Measurements, and Statistical Models – A Response to Lennard (2013) Fingerprint identification: How Far Have We Come? *Australian Journal of Forensic Sciences*, Vol. 46, pp. 282–292. <http://dx.doi.org/10.1080/00450618.2013.833648>

115. Moses K. (2011). Chapter 6: Automatic Fingerprint Identification Systems (AFIS), in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs, NCJ 225320.
116. Nagar A., Choi H-S. and Jain A.K. (2012). Evidential Value of Automated Latent Fingerprint Comparison: An Empirical Approach. *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 6, pp. 1752-1765.
117. Nakhaeizadeh S., Dror I.E. and Morgan R.M. (2014). Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias. *Science & Justice*, Vol. 54, No. 3, pp. 208-214.
118. National Commission on Forensic Science (NCFS). (2015). Subcommittee on Human Factors- *Ensuring that Forensic Analysis is Based Upon Task-Relevant Information*, available at: <https://www.justice.gov/ncfs/file/818196/download>
119. National Commission on Forensic Science (NCFS). (2016). Views of the Commission: *Facilitating Research on Laboratory Performance* (adopted unanimously September 13, 2016). Available at: <https://www.justice.gov/ncfs/page/file/909311/download>
120. National Institute of Standards and Technology. (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, pp. 72.
121. National Research Council. (2009). *Strengthening Forensic Science in the United States: A Path Forward*. US Department of Justice, Committee on Identifying the Needs of the Forensic Sciences Community. Available at: <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>
122. National Science and Technology Council Committee on Science, Subcommittee on Forensic Science. (2014). *Strengthening Forensic Science: A Progress Report*, page 2, available at: https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/strengthening_the_forensic_sciences_may_-_2014.pdf
123. Neufeld P.J. (2005). The (Near) Irrelevance of Daubert to Criminal Justice: And Some Suggestions for Reform. *American Journal of Public Health*, Vol. 95 (Supp. 1): S107, S110.
124. Neumann C., Evett I.W. and Skerrett J. (2012). Quantifying the Weight of Evidence from a Forensic Fingerprint Comparison: A New Paradigm. *Journal of the Royal Statistical Society*, Vol. 175, Part 2, pp. 1-26.
125. Neumann C., Champod C., Yoo M., et al. (2015). Quantifying the Weight of Fingerprint Evidence through the Spatial Relationship, Directions and Types of Minutiae Observed on Fingermarks. *Forensic Science International*, Vol. 248, pp. 154-171.

126. Nickerson R.S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, Vol. 2, No. 2, pp. 175-220.
127. Nisbett R.E. and Wilson T.D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, Vol. 84, No. 3, pp. 231.
128. Nordgaard A. and Rasmusson B. (2012). The Likelihood Ratio as Value of Evidence- More Than A Question of Numbers. *Law, Probability and Risk*, Vol. 11, No. 4, pp. 303-315.
129. Office of the Inspector General U.S. Department of Justice. (2006). *A Review of the FBI's Handling of the Brandon Mayfield Case*. Office of the Inspector General U.S. Department of Justice, Washington DC, pp. 1-330.
130. Okajima M. and Usukara K. (1984). Quantitative and Genetic Features of Epidermal Ridge Minutiae on the Palm of Twins. *Human Heredity*, Vol. 34, pp. 285-290.
131. Olsen R.D. (1978). Occupational Damage and Bad Skin Condition, in *Scott's Fingerprint Mechanics*. Springfield: Charles C. Thomas, pp. 80-82.
132. Orne M.T. (1962). On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications. *American Psychologist*, Vol. 17, No. 11, pp. 776-783.
133. Osborne N.K., Woods S., Kieser J. and Zajac R. (2014). Does Contextual Information Bias Bitemark Comparisons? *Science & Justice*, Vol. 54, No. 4, pp. 267-273.
134. Ostrum B. (2009). Commentary on: Authors' Response [J Forensic Sci, 2009, Vol. 54, No. 2, pp. 501] to Wells' Comments [J Forensic Sci, 2009, Vol. 54, No. 2, pp. 500] regarding Krane D.E., et al. (2008). Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation, *J Forensic Sci. Journal of Forensic Sciences*, Vol. 54, No. 6, pp. 1498-1499.
135. Pacheco I., Cerchiai B. and Stoiloff S. (2014). *Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations*. Miami Dade Police Department/National Institute of Justice, Final Technical Report, Document # 248534.
136. Pankanti S., Prabhakar S. and Jain A.K. (2002). On the Individuality of Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1010-1025.
137. Parsons N.R., Smith J.Q., Thonnes E., et al. (2008). Rotationally Invariant Statistics for Examining the Evidence from the Pores in Fingerprints. *Law, Probability and Risk*, Vol. 7, No. 1, pp. 1-14.

138. Paulhus D.L. (1991). Measurement and Control of Response Biases, in *Measures of Personality and Social Psychological Attitudes*. JP Robinson, PR Shaver and LS Wrightsman, Editors. San Diego, CA: Academic Press, 746 pages.
139. Paulino A.A., Jain A.K. and Jianjiang F. (2010). Latent Fingerprint Matching: Fusion of Manually Marked and Derived Minutiae, in *23rd SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 63-70.
140. Paulino A.A., Feng J. and Jain A.K. (2013). Latent Fingerprint Matching Using Descriptor-Based Hough Transform. *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 1, pp. 31-45.
141. People v. Jennings (State of Illinois v. Jennings). 252 Ill. 534, 96 N.E. 1077 (1911). Personal communication, Henry Swofford, Chief, Latent Print Branch- U.S. Army Criminal Investigation Laboratory. Forest Park, GA: November 4, 2015.
142. Peterson J.L., Lin G., Ho M., et al. (2003). The Feasibility of External Blind DNA Proficiency Testing. I. Background and Findings. *Journal of Forensic Sciences*, Vol. 48, No. 1, pp. 21-31.
143. President's Council of Advisors on Science and Technology (PCAST) (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Executive Office of the President, September 2016. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf
144. Pretty I. and Sweet D. (2010). A Paradigm Shift in the Analysis of Bitemarks. *Forensic Science International*, Vol. 201, pp. 38-44.
145. Pronin E. and Kugler M.B. (2007). Valuing Thoughts, Ignoring Behavior: The Introspection Illusion as a Source of the Bias Blind Spot. *Journal of Experimental Social Psychology*, Vol. 43, No. 4, pp. 565-578.
146. Pronin E., Gilovich T. and Ross L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review*, Vol. 111, No. 3, pp. 781-799.
147. Pronin E., Yin D.Y. and Ross L. (2002). The Bias Blind Spot: Perception of Bias in Self and Others. *Personality and Social Psychology Bulletin*, Vol. 28, pp. 369-381.
148. Ray E. and Dechant P.J. (2013). Sufficiency and Standards for Exclusion Decisions. *Journal of Forensic Identification*, Vol. 63, No. 6, pp. 675-697.
149. Reed T., Viken R.J. and Rinehart S.A. (2006). High Heritability of Fingertip Arch Patterns in Twin-Pairs. *American Journal of Medical Genetics*, Vol. 140A, pp. 263-271.

150. Risinger D.M. (2000). Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left at the Dock? *Albany Law Review*, Vol. 64, No. 1, pp. 99-149.
151. Risinger D.M., Saks M.J., Thompson W.C and Rosenthal R. (2002). The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion. *California Law Review*, Vol. 90, No. 1, pp. 1-56.
152. Robertson B. and Vignaux G.A. (1995). *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. New York, NY: John Wiley and Sons, 240 pages.
153. Roddy A.R. and Stosz J.D. (1997). Fingerprint Features- Statistical Analysis and System Performance Estimates. *Proceedings of the IEEE*, Vol. 85, No. 9, pp. 1390-1420.
154. Saks M.J., et al. (2003). Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States. *Science & Justice*, Vol. 43, No. 2, pp. 77-90.
155. Saks M.J. and Koehler J.J. (2005). The Coming Paradigm Shift in Forensic Identification Science. *Science*, Vol. 309, No. 5736, pp. 892-895.
156. Sarkar N.C. (2004). *Finger Ridge Minutiae: Classification, Distribution and Genetics*. Kolkata, India: Anthropological Survey of India.
157. Schreiber M.H. (1963). The Clinical History as a Factor in Roentgenogram Interpretation. *JAMA*, Vol. 185, No. 5, pp. 399-401.
158. Schum D.A. (1994). *Evidential Foundations of Probabilistic Reasoning*. New York, NY: John Wiley & Sons, 545 pages.
159. Scientific Working Group on Friction Ridge Analysis, Study and Technology. (2013). *Document #10: Standard for Examining Friction Ridge Impressions and Resulting Conclusions*, Version 2.0.
160. Seweryn P. (2005). Frequency of Minutiae on Left and Right Hand Index Fingers. *Problemy Kryminalistyki*, No. 247, pp. 40-46.
161. Sheets H.D., Torres A., Langenburg G., et al. (2014). Distortion in Fingerprints: A Statistical Investigation Using Shape Measurement Tools. *Journal of Forensic Sciences*, Vol. 59, No. 4, pp. 1113-1120.
162. Sheldrake R. (1999). How Widely is Blind Assessment Used in Scientific Research? *Alternative Therapies in Health and Medicine*, Vol. 5, No. 3, pp. 88.

163. Shultz K. and Grimes D. (2002). Blinding in Randomized Trials: Hiding Who Got What. *The Lancet*, Vol. 359, No. 9307, pp. 696-700.
164. Srihari S.N., Srinivasan H. and Fang G. (2008). Discriminability of Fingerprints in Twins. *Journal of Forensic Identification*, Vol. 58, No. 1, pp. 109-127.
165. Stacey R.B. (2004). Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case. *Journal of Forensic Identification*, Vol. 54, No. 6, pp. 706-718.
166. Steele L. (2004). The Defense Challenge to Fingerprints. *Criminal Law Bulletin*, Vol. 40 No. 3, pp. 213-240.
167. Stern H. (2012). Comment on Neumann et al. *Journal of the Royal Statistical Society*, Series A, 175 (part 2), pp. 408-409.
168. Stoel R.D., Berger C.E.H., Kerkhoff W., et al. (2014). Minimizing Contextual Bias in Forensic Casework, in *Forensic Science and the Administration of Justice: Critical Issues and Directions*. KJ Strom and MJ Hickman, Eds. Sage Publications, pp. 67-86.
169. Stoel R.D., Dror I.E. and Miller L.S. (2014). Bias Among Forensic Document Examiners: Still a Need for Procedural Changes. *Australian Journal of Forensic Sciences*, Vol. 46, No. 1, pp. 91-97.
170. Stoney D.A. (2001). Measurement of Fingerprint Individuality, in *Advances in Fingerprint Technology*, 2nd edition. HC Lee and RE Gaensslen, Eds. Boca Raton: CRC Press, pp. 327-387.
171. Stucker M., Geil M., Kyeck S., et al. (2001). Interpapillary Lines—The Variable Part of the Human Fingerprint. *Journal of Forensic Sciences*, Vol. 46, No. 4, pp. 857-861.
172. Su C. and Srihari S. (2009). Probability of Random Correspondence for Fingerprints, in *Computational Forensics: Third International Workshop, IWCF 2009*. LNCS 5718, ZJM Geradts, KY Franke and CJ Veenman, Eds. Berlin Heidelberg: Springer Verlag, pp. 55-66.
173. Su C. and Srihari S. (2010). Evaluation of Rarity of Fingerprints in Forensics. *Proceedings of Neural Information Processing Systems*. Vancouver, Canada, December 2010.
174. Swofford H. (2015). The Emerging Paradigm Shift in the Epistemology of Fingerprint Conclusions. *Journal of Forensic Identification*, Vol. 65, No. 3, pp. 201-213.
175. Tangen J.M., Thompson M.B. and McCarthy D.J. (2011). Identifying Fingerprint Expertise. *Psychological Science*, Vol. 22, No. 8, pp. 995-997.
176. Tao X., Chen X., Yang X. and Tian J. (2012). Fingerprint Recognition with Identical Twin Fingerprints. *PLoS ONE*, Vol. 7, No. 4, pp. 1-7.

177. Tape T.G. and Panzer R.J. (1986). Echocardiography, Endocarditis, and Clinical Information Bias. *Journal of General Internal Medicine*, Vol. 1, Vol. 5, pp. 300-304.
178. Thompson W.C. (1993). Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the "DNA War". *Journal of Criminal Law & Criminology*, Vol. 84, pp. 22-104.
179. Thompson W.C. (2008). Beyond Bad Apples: Analyzing the Role of Forensic Science in Wrongful Convictions. *Southwestern Law Review*, Vol. 37, No. 4, pp. 1027-1050.
180. Thompson W.C. (2009a). Interpretation: Observer Effects, in *Wiley Encyclopedia of Forensic Science*. A Jamieson, A Moenssens, Editors. Chichester, UK: John Wiley & Sons Ltd, pp. 1575-1579.
181. Thompson W.C. (2009b). Painting the Target Around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation. *Law, Probability and Risk*, Vol. 8, pp. 257-276.
182. Thompson W.C. (2011). What Role Should Investigative Facts Play in the Evaluation of Scientific Evidence? *Australian Journal of Forensic Sciences*, Vol. 43, No. 2-3, pp. 123-134.
183. Thompson W.C. (2012). Discussion paper: Hard Cases Make Bad Law: Reactions to R v. T. *Law, Probability and Risk*, Vol. 11, pp. 347-359.
184. Thompson W.C., Vuille J., Biedermann A. and Taroni F. (2013). The Role of Prior Probability in Forensic Assessments. *Frontiers in Genetics*, Vol. 4, pp. 220-223.
185. Thompson M.B., Tangen J.M. and McCarthy D.J. (2013). Expertise in Fingerprint Identification. *Journal of Forensic Sciences*, Vol. 58, No. 6, pp. 1519- 1530.
186. Thompson M.B., Tangen J.M. and McCarthy D.J. (2014). Human Matching Performance of Genuine Crime Scene Latent Fingerprints. *Law and Human Behavior*, Vol. 38, No. 1, pp. 84-93.
187. Thompson W.C. and Newman E.J. (2015). Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents. *Law & Human Behavior*, Vol. 39, No. 4, pp. 332-349.
188. Thompson W.C. (2016). Determining the Proper Evidentiary Basis for an Expert Opinion: What Do Experts Need to Know and When Do They Know Too Much?, in *Blinding as a Solution to Bias in Biomedical Science and the Courts: A Multidisciplinary Approach*. C Robertson and A. Kesselheim, Eds. Elsevier.
189. Thornton J.I. (1977). One-Dissimilarity Doctrine in Fingerprint Identification. *International Criminal Police Review*, No. 306, pp. 89-95.

190. Thornton J.I. (2010). Letter to the Editor—A Rejection of “Working Blind” as a Cure for Contextual Bias. *Journal of Forensic Sciences*, Vol. 55, No. 6, pp. 1663.
191. Ulery B.T., Hicklin R.A., Buscaglia J. and Roberts M.A. (2011). Accuracy and Reliability of Forensic Latent Fingerprint Decisions. *Proceedings of the National Academy of Sciences, USA*, Vol. 108, No. 19, pp. 7733-7738.
192. Ulery B.T., Hicklin R.A., Buscaglia J. and Roberts M.A. (2012). Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. *PLoS ONE*, Vol. 7, No. 3, pp. 1-12.
193. Ulery B.T., Hicklin R.A., Kiebusinski G.I., Roberts M.A., and Buscaglia J. (2013). Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations. *Forensic Science International*, Vol. 230, No. 1, pp. 99-106.
194. Ulery B.T., Hicklin R.A., Roberts M.A., Buscaglia J. (2014). Measuring What Latent Fingerprint Examiners Consider Sufficient Information for Individualization Determinations. *PLoS ONE*, Vol. 9, No. 11, e110179.
195. Ulery B.T., Hicklin R.A., Roberts M.A., Buscaglia J. (2015). Changes in Latent Fingerprint Examiners’ Markup Between Analysis and Comparison. *Forensic Science International*, Vol. 247, pp. 54-61.
196. Ulery B.T., Hicklin R.A., Roberts M.A. and Buscaglia J. (2016). Interexaminer Variation of Minutia Markup on Latent Fingerprints. *Forensic Science International*, Vol. 264, pp. 89-99.
197. Ulery B.T., Hicklin R.A., Roberts M.A. and Buscaglia J. (2017). Factors Associated with Latent Fingerprint Exclusion Determinations. *Forensic Science International*, Vol. 275, pp. 65-75.
198. U.S. Senate Committee on Commerce, Science, and Transportation. (2011). http://www.commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=63e87410-acf3-45eb-a849-1b4edf6a8959&Statement_id=f4c8fa35-3359-4a0f-ae8c-ee2b4b0641c6&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a&MonthDisplay=12&YearDisplay=2011
199. U.S. Senate Committee on Commerce, Science, and Transportation. (2012). http://www.commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=7665a46c-571e-4e73-8068-6c6402df2ae1&Statement_id=30eee171-a085-4d66-a847-7378e94ecd7c&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a&MonthDisplay=3&YearDisplay=2012
200. U.S. Senate Committee on Commerce, Science, and Transportation. (2013). http://www.commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=f0921154-2c02-456e-a40a-7bfa6cb417d0&Statement_id=af01eb83-8aa4-45d0-8e61-ef092eaae236&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a&MonthDisplay=6&YearDisplay=2013

201. Watson C., Fiumara G.P., Tabassi E., et al. (2014). *Fingerprint Vendor Technology Evaluation*. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 8034.
202. Wells J.D. (2009). Commentary on: Krane D.E., et al. (2008). Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation, *J Forensic Sci. Journal of Forensic Sciences*, Vol. 54, No. 2, p. 500.
203. Wertheim K., Langenburg G. and Moenssens A. (2006). A Report of Latent Print Examiner Accuracy During Comparison Training Exercises. *Journal of Forensic Identification*, Vol. 56, No. 1, pp. 55-93.
204. Wertheim K. and Maceo A. (2002). The Critical Stage of Friction Ridge and Pattern Formation. *Journal of Forensic Identification*, Vol. 52, No. 1, pp. 35-85.
205. Wertheim K. (2011). Chapter 3: Embryology and Morphology of Friction Ridge Skin, in *The Fingerprint Sourcebook*. A McRoberts, Editor. US Department of Justice, Office of Justice Programs, NCJ 225320.
206. Wilson T.D. and Brekke N. (1994). Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations. *Psychological Bulletin*, Vol. 116, No. 1, pp. 117.
207. Wilson C., et al. (2004). *Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report*. U.S Department of Commerce, National Institute of Standards and Technology, NISTIR 7123.
208. Yoon S., Feng J. and Jain A.K. (2012). Altered Fingerprints: Analysis and Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 451-464.
209. Yoon S., et al. (2013). LFIQ: Latent Fingerprint Image Quality. *Biometrics: Theory, Applications and Systems (BTAS)*, IEEE Sixth International Conference on Biometrics Compendium, pp. 1-8.
210. Yoon S. and Jain A.K. (2015). Longitudinal Study of Fingerprint Recognition. *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 112, No. 28, pp. 8555-8560.
211. Zabell S.L. (2005). Fingerprint Evidence. *Journal of Law and Policy*, Vol. 13, No. 1, pp. 143-179.
212. Zhu Y., Dass S.C. and Jain A.K. (2007). Statistical Models for Assessing the Individuality of Fingerprints. *IEEE Transactions on Information Forensics and Security*, Vol. 2, No. 3, pp. 391-401.

- (1) <http://www.nist.gov/itl/iad/ig/fingerprint.cfm>
- (2) http://biometrics.nist.gov/cs_links/latent/elft-efs/IAI_2012/ELFT-EFS2_IAI_2012_Final.pdf
- (3) http://www.nist.gov/itl/iad/ig/bio_quality.cfm
- (4) <https://www.nist.gov/news-events/news/2014/01/us-departments-justice-and-commerce-name-experts-first-ever-national>
- (5) www.onin.com

APPENDICES

A. Methods

B. Latent Fingerprint Examination – A Primer

C. Bibliography and Questions

D. Working Group Roster

E. Working Group Bios

F. Project Advisory Committee and Staff

A. METHODS

For this review of the foundational literature for latent fingerprint examination, the first task was to update the bibliography compiled by the White House Subcommittee on Forensic Science (SoFS). That document, compiled in 2012, consisted of citations to over 100 articles, books, book chapters and reports that forensic practitioners believed to be foundational to their field. To update this bibliography, staff used search engines at two online university libraries to identify publications related to latent print examination. The first was Boston University Medical Center's Alumni Medical Library, using the Google Scholar search engine. The second was The George Washington University's *ArticlesPlus* search engine. A time range was implemented (January 2012- June 2015) for both search engines to limit the results to only those published since the release of the original bibliographies. Relevant articles were reviewed by project staff and combined with the original bibliography, to provide the WG with an updated bibliography that listed over 150 publications, with a complete citation and abstract for each publication. Staff also located and made available to the WG electronic copies of nearly all the articles listed in the updated bibliography. (A few articles could not be obtained from major university libraries, but the WG determined that those rather obscure publications were not important to its analysis.) Working Group members were sent the annotated bibliographies developed in the SoFS process, updated by AAAS staff (See Appendix C). These bibliographies were not intended to limit the Working Groups in the sources they relied on, although the bibliographies served as a good starting place for the evaluation.

In conjunction with the forensic scientist on the WG, staff also prepared a detailed primer on the specific operational steps forensic laboratories typically follow in performing latent print examination, using Automated Fingerprint Identification Systems (AFIS), and forming conclusions. This primer was provided to the non-forensic scientists in the WG to help them achieve a more sophisticated understanding of actual practices in forensic laboratories. The WG began its analysis by reading abstracts of all of the publications in the updated bibliography in order to identify those most important for assessing the scientific foundations of the field. They then carefully studied those publications that they had identified as important, with each member focusing most heavily on those publications most closely related to his interests and expertise.

Early in this process, the WG determined that it was helpful, and indeed necessary, to consider publications beyond those listed in the bibliography. The bibliography consisted primarily of publications identified by the IWG as supporting the foundational validity of latent print examination. But this list omitted a number of important review articles, reports, and scholarly commentaries on latent print examination, and more generally on pattern analysis in forensic science, that the WG found helpful in understanding the field. A number of these publications are cited in this report. While these articles are not, strictly speaking, a part of the foundation literature for latent print examination, they offer important perspectives on that literature and on the scientific underpinnings of the field.

At the beginning of the review process, project staff proposed that members of the WG prepare a written evaluation of each and every publication in the bibliography, using an assessment form that listed five overarching questions to be answered for each publication, including, for example, "What are the strengths/weaknesses of the article/report?" "Did the article/report answer the specific question under which it was listed?" Staff also provided assessment guidelines for the WG to use when reviewing

each publication. Both the assessment forms and the assessment guidelines were vetted by the Advisory Committee.

After completing only a handful of the assessment forms, however, the WG determined that the evaluation of individual publications, one at a time, was not an effective approach to reviewing this literature. This atomistic approach ignores the concept of convergent validity- i.e., the possibility that various publications, each with distinct limitations when considered by itself, can reinforce each other and collectively support conclusions that would not be warranted on the basis of a single article. The WG found, for example, that studies testing the accuracy of latent print examiners were best considered as a group. While each study had unique limitations that affected the conclusions that could be drawn from it, the studies as a group told a consistent story that allowed more definitive conclusions to be drawn from the literature as a whole. For each of the issues it considered, the WG found it necessary to consider the collective strength of multiple studies. The WG, therefore, dispensed with the individual evaluation forms and proceeded to evaluate the literature in a holistic manner. The resulting report thus constitutes a review of the whole of the scientific literature underpinning the forensic discipline of latent fingerprint examination, rather than an atomistic discussion of the strengths of limitations of individual publications in the expanded bibliography.¹

The IWG bibliography was organized around fifteen questions related to the scientific basis for latent print analysis. Each section of the bibliography listed articles thought to address one of the fifteen questions. At the outset of this review, staff suggested that the WG organize its analysis around the same fifteen questions identified by the IWG. As it delved into the literature, however, the WG came to the conclusion that the fifteen questions posed by the IWG were not the best way to frame the key underlying issues for purpose of this review. Prior to an in-person meeting of the WG, the Chair suggested that the review of the literature instead be organized around six underlying questions. The WG organized its intensive discussion at the in-person meeting around those questions and found that they allowed the important scientific issues to be considered and addressed in a concise and efficient manner. Consequently, the WG decided to organize this report around the same six questions.

After the in-person meeting, the Chair of the WG prepared a preliminary draft of this report. The other members of the WG helped expand this draft and contributed additional sections. The WG and staff then held a series of conference calls to develop and refine the report through multiple drafts. Extensive discussion was required to reach consensus among the WG with regard to a number of specific conclusions. Sections of the report were re-drafted and expanded with additional citations in order to improve the report. This process took a number of months but led ultimately to a draft that reflected the considered judgment and consensus view of the entire WG.

Finally, drafts of the report were sent to the Advisory Committee and several stakeholders for their review, and their remarks contributed to the preparation of this final report.

¹ The audience for the report includes members of both the defense and prosecutorial bars, judges, policy makers (such as members of Congress and members of the Executive Branch), the academic, scientific and technical community, and funders. Because of the broad constituency of the report, the holistic approach adopted by the WG has distinct advantages over a review of individual bibliographic sources. This constituency will not find it useful to learn, for example, that article #15 meets the requirements of a “good” scientific article, but article #43 has an inadequate sample size, article #53 used the wrong statistic, etc. The methodology used in this report provides a useful guide to answer the questions What do we know? What don’t we know? What are the gaps? What research or other activities should be pursued to improve latent fingerprint analysis?

B. LATENT FINGERPRINT EXAMINATION – A PRIMER

Latent fingerprint examination involves the evaluation and comparison of friction ridge impressions made by fingers, palms, and soles, as well as the examination of the skin itself in these three areas. Material on the surface of the ridges can transfer when these skin areas come into contact with other objects, leaving an impression of the friction ridge pattern. Latent print examiners have learned to use lighting, powder and chemicals to reveal and make a record of these impressions.

Friction ridge skin impressions (fingerprints, palm prints, and footprints) have been used for over 100 years as a means of personal identification. Two characteristics of these impressions make them useful for identification. First, barring injury or disease, friction ridge patterns persist throughout the lifetime of the individual. Second, friction ridge patterns are highly variable among individuals, making them useful for distinguishing one individual from another.

Examiners can manually compare an impression to known prints of specific individuals (e.g. victims, suspects or other persons of interest) or conduct an automated search in an AFIS (Automated Fingerprint Identification System) database. Most states have a statewide or local database of exemplars available for searching. The FBI administers a national AFIS database that includes over 70 million records. Candidates generated from AFIS searches are examined using manual comparisons.

The framework used to guide decision-making during the examination process is referred to as ACE-V: Analysis, Comparison, Evaluation and Verification.

The process begins with the analysis of the unknown impression to determine whether it is suitable, or sufficient, for comparison. The examiner observes this unknown (questioned) print and seeks to understand its presentation or appearance by gathering all available information in the print. This understanding is aided by considering various dynamics that can occur when three-dimensional skin contacts a two-dimensional surface. These may include the following:

- Skin condition – natural ridge structure, injury, signs of disease or aging
- Contact surface – porous, nonporous, flat, concave, convex, textured, smooth, clean, dirty, background noise, etc.
- Impression matrix (residue) – perspiration and/or oil material, blood, grease, ink, etc.
- Development technique – consideration of the technique’s signature (appearance) and its consistency throughout the impression
- Mechanics of touch – amount of pressure applied during initial contact of skin with surface (deposition pressure), as well as the degree of displacement of the skin after contact (lateral pressure)
- Quality/clarity – the reliability of the observed features
- Quantity – the overall amount of information available at all levels of detail
- Preservation or capture method – photograph, lift, etc.

If the examiner determines that the features of the latent are sufficiently informative to allow a meaningful comparison, then the examiner declares the impression to be *of value* and documents those features. The examiner next assesses the quality of the known print in a similar manner before proceeding to the comparison phase. If the examiner determines that the features of the latent or

known print are insufficiently informative to allow a meaningful comparison, the examiner declares the print to be of “no value” and the examination ends at that point.

In the comparison and evaluation phases, the information gathered from the unknown print during analysis is compared against a known print. The examiner considers the possibility that the observed consistencies and differences could arise between prints that have a common source (e.g., same finger) and between prints that have a different source. Examiners compare and evaluate the following features/information during a given examination, depending on the clarity of the prints:

- Overall friction ridge flow (Level 1 detail) to include pattern type, if discernible
- Friction ridge paths (Level 2 detail or minutiae), including ridge endings, bifurcations and dots, or a combination of these features
- Finer friction ridge information (Level 3 detail), including sweat pore location, ridge edge features, and ridge/furrow widths
- Occasional features such as creases and scars
- Spatial relationships of observed features
- Ridge counts between observed features
- Dynamics of the deposition process that may lead to distortion
- Tolerance (allowance) for variations in appearance between the unknown/known prints that may have resulted from any distortion

Although comparison and evaluation are typically listed as separate steps, examiners generally move seamlessly from comparison to evaluation. As the examiner compares more information between the unknown and known prints, he/she begins to formulate a tentative opinion on whether the prints have a common source. The comparison continues until the examiner either has confidence in the opinion reached, or has decided that the available evidence is insufficient to determine with confidence whether the prints have a common source or not. The traditional practice has been to report a categorical conclusion of “identification,” “exclusion,” “inconclusive,” or “no value.” This report challenges the scientific basis of the identification decision.

In the ACE-V framework there are no formal standards for evaluating the degree of similarity or dissimilarity between prints and no formal rules (e.g. numerical threshold) for determining when to report the comparison as an “identification,” “exclusion,” or “inconclusive.” It is entirely up to the examiner to decide, based on the features of the prints being compared, what conclusion to report. That decision depends both on examiners’ ability to discern the many features and their experience evaluating the significance of those features. A major factor in assessing significance of these features is the examiner’s judgment of the specificity, or rarity, of the observed features. These judgments are made subjectively based on experience rather than by consulting data on the specificity of features. The greater the perceived specificity of the features, the greater the weight or value assigned to those features during the examination process. The evaluation may also be affected by the visual acuity of the examiner, which affects the ability to observe both gross and minute details.

The decision to “exclude” is made when the examiner is confident that there are sufficient inconsistencies between the compared impressions that they are unlikely to have a common source, thus concluding they are from different sources. The decision to report the comparison as “inconclusive” is made if an examiner cannot find sufficient agreement or disagreement to support either an identification or exclusion. Typically, the inconclusive decision indicates issues with the known

prints, such as incomplete recording of the source skin. Some examiners may also render an inconclusive decision due to limited quality and quantity of information in the latent print. It is good practice for an examiner to state the reasoning behind a decision in the case report, to provide guidance for the end user.

Currently, examiners undergo a lengthy training process to help them see, recognize, and become familiar with the specificity of features and learn how to use those features to make accurate source determinations. Through training, novices learn how more experienced examiners compare and evaluate prints and how they make decisions. Novice examiners are also able to hone their skills by evaluating training sets containing pairs of prints that are known to be from the same finger or different fingers, in order to test their accuracy in making that determination. Through training, novices develop an intuitive sense of the specificity of features. They also learn how to evaluate and take account of within-finger variability that may arise from distortion. Black-box studies have consistently shown that trained examiners out-perform novices and have relatively low rates of false matches, albeit with somewhat higher rates of false exclusions. These findings suggest that the training process, while perhaps having some room for improvement, is largely successful.

The final phase of the ACE-V decision-making framework is verification. Verification is a quality control mechanism in which a second qualified examiner reviews the work of the first examiner. In many agencies, only identifications are verified. The verification stage is not supposed to rubber-stamp the initial examiner's conclusion. Rather, verification is an attempt to ensure that the initial conclusion was properly drawn. In the vast majority of fingerprint examinations, both the initial examiner and verifier reach the same conclusion.

Verification procedures differ among laboratories and agencies. In a traditional verification, a second competent examiner, who knows the initial examiner's conclusion reviews the impressions that were compared, along with the initial examiner's work-product (e.g., notes regarding features of value for comparison), and decides to either confirm or reject the first examiner's conclusion. It is a type of open peer review- where the verifier has access to the initial examiner's entire case file- thus permitting the verifier to scrutinize the accuracy of both the process and the conclusion.¹ The examiner and verifier may use different information because no two examiners are likely to interpret information in exactly the same way. However, they may still reach the same conclusion.

An alternative approach, used in some laboratories, is to practice "blind" verification, in which the initial examiner's work product is unavailable to the verifying examiner. Blind verification assures that the second examiner is not subconsciously biased by knowing the conclusion of the first examiner. However, because it does not include a review of the first examiner's work product (e.g., notes and mark-up of the latent) it may be less helpful in identifying weaknesses in the first examiner's work process and analysis. Other approaches may include using multiple verifiers, or having the same verifiers first conduct a blind verification, and then un-blind to allow open peer review.

When defending their conclusions in the courtroom, latent print examiners have sometimes made the claim that ACE-V is a validated method and that following ACE-V is sufficient to assure the accuracy and reliability of examiner's conclusions. Indeed, some practitioners claimed that the ACE-V method, when followed properly, has an error rate of zero. These claims were powerfully challenged by academic

¹ The peer review referred to here simply means that an examiner's work is checked by a peer. It is not the scientific process of peer-review and publication discussed in *Daubert v. Merrell Dow Pharmaceuticals* (1993).

commentators (e.g., Cole, 2005) as well as by the 2009 NRC report, which made the following comments about the ACE-V method:

ACE-V provides a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis. ACE-V does not guard against bias; is too broad to ensure repeatability and transparency; and does not guarantee that two analysts following it will obtain the same results. For these reasons, merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results (NRC 2009, p. 142).

The NRC report concluded that “[b]etter documentation is needed of each step in the ACE-V process” in order to create the “transparent record” needed to allow courts to “assess the reliability of the method for a specific case” (p. 143). The NRC report also called for additional research to validate latent print examination and to assess error rates, pointing out that “[t]he method, and the performance of those who use it, are inextricably linked, and both involve multiple sources of error...” (p. 143). Since 2009, additional research of the type envisioned in the NRC report has begun to emerge. One purpose of the current report is to offer a balanced assessment of the current state of validation of the field- to suggest what is now known, and what additional research is needed, to move the field forward.

C. BIBLIOGRAPHY

Recent Publications (indicated by *)/Late additions (indicated by **)

UNDERLYING FINGERPRINT CHARACTERISTICS

I. What scientific literature describes how distinct or similar fingerprints are across overall population, related individuals or identical twins?

***1. Anthonioz A. and Champod C. (2014). Integration of Pore Features into the Evaluation of Fingerprint Evidence. *Journal of Forensic Sciences*, Vol. 59, No. 1, pp. 82-93.**

Fingerprint practitioners rely on level 3 features to make decisions in relation to the source of an unknown friction ridge skin impression. This research proposes to assess the strength of evidence associated with pores when shown in (dis)agreement between a mark and a reference print. Based upon an algorithm designed to automatically detect pores, a metric is defined in order to compare different impressions. From this metric, the weight of the findings is quantified using a likelihood ratio. The results obtained on four configurations and 54 donors show the significant contribution of the pore features and translate into statistical terms what latent fingerprint examiners have developed holistically through experience. The system provides LR's that are indicative of the true state under both the prosecution and the defense propositions. Not only such a system brings transparency regarding the weight to assign to such features, but also forces a discussion in relation to the risks of such a model to mislead.

2. Chen Y. and Jain A.K. (2009). *Beyond Minutiae: A Fingerprint Individuality Model with Pattern, Ridge and Pore Features*, in *Advances in Biometrics: Third International Conference, ICB 2009*. LNCS 5558, M. Tistarelli and M.S. Nixon, Eds. Berlin- Heidelberg: Springer Verlag, pp. 523-533.

Fingerprints are considered to be unique because they contain various distinctive features, including minutiae, ridges, pores, etc. Some attempts have been made to model the minutiae in order to get a quantitative measure for uniqueness or individuality of fingerprints. However, these models do not fully exploit information contained in non-minutiae features that is used for matching fingerprints in practice. We propose an individuality model that incorporates all three levels of fingerprint features: pattern or class type (Level 1), minutiae and ridges (Level 2), and pores (Level 3). Correlations among these features and their distributions are also taken into account in our model. Experimental results show that the theoretical estimates of fingerprint individuality using our model consistently follow the empirical values based on the public domain NIST-4 database.

3. Colletto G.M.D.D. and Zimberknopf S. (1987). Diagnosis of Twin Zygoty by Dermatoglyphic Index. *Brasil Journal of Genetics*, Vol. 10, No. 2, pp. 289-299.

(Summary from bibliography***)This paper proposes the use of dermatoglyphic analysis to determine twin zygoty. The study sample consists of 120 same sex twin pairs, of which 66 are monozygous and 54 are dizygous. Numerical values were given to 35 dermatoglyphic variables in order to calculate a dermatoglyphic index. A discriminant analysis was then applied to within-pair differences of these dermatoglyphic variables. 15 of these variables were selected to be used in a function allowing the proper classification of 100% of the twin pairs in this study. Other methodologies are recommended for cases in which the probability of mono- or dizygoty is below 90%.

4. International Association for Identification, et al. (2011). *The Fingerprint Sourcebook*. NCJ 225320. *The Fingerprint Sourcebook* aims to be the definitive resource on the science of fingerprint identification. The *Sourcebook* was prepared by the International Association for Identification and topics covered include the anatomy and physiology of friction ridge skin (the uniquely ridged skin found on the palms and soles); techniques

for recording exemplars from both living and deceased subjects; the FBI's Automated Fingerprint Identifications Systems (AFIS); latent print development, preservation and documentation; equipment and laboratory quality assurance; perceptual, cognitive and psychological factors in expert identifications; and legal issues.

5. Kumbhani H.K. (2007). *Dermatoglyphics: A Review. Anthropology Today: Trends, Scope and Applications*, No. 3, pp. 285-295.

(No abstract in original article***) This article offers an etymological review of dermatoglyphics that shows the valuable implications, applications and utility of the work carried out in this area. It is based on more than 220 references covering the years 1891 to 2005. More specifically, the author covers published work on the frequency of various patterns for fingerprints in different communities and populations. Studies on palm, sole, and toe prints as well as on middle and proximal phalanges are covered, showing their strong and weak points. A review of work relating to inheritance and to the correlation between dermatoglyphics and diseases is also presented. The role of dermatoglyphics for personal identification, disputed paternity, and diagnosis of monozygotic (identical) versus dizygotic (fraternal) twins is also reviewed.

6. Loesch D.Z. (1983). *Quantitative Dermatoglyphics: Classification, Genetics, and Pathology*. Oxford University Press, Oxford.

(Summary from bibliography***) A topological system to describe papillary surfaces is presented. A series of dermatoglyphic measurements are associated with this system, such as ridge counts or ridge breadth. The system then allows interpreting the relationships between pattern elements and combinations thereof. The book is divided into three parts and covers the subjects of the classification of dermatoglyphic patterns and their topology, the variability within and between human populations of topologically significant pattern elements, their heritability and diagnostic application.

7. MacArthur J.W. (1938). *Reliability of Dermatoglyphics in Twin Diagnosis. Human Biology*, Vol. 10, No. 1, pp. 12- 35.

(No abstract in original article***) In order to state the degree of similarity in dermatoglyphic features between individuals and between hands and feet of one individual, a quantitative and objective method was developed and is presented in this paper. Comparisons were made bilaterally (hands of one person), homolaterally (same side hands of pair), and heterolaterally (opposite side hands of pair) between parents and children, siblings, mono- and dizygotic twins, and left and right hands of individuals. Differences in terms of ridge count, finger pattern, palmar digital and axial triradii, and palm patterns were accounted for. Averages and ranges of differences for all pairwise comparisons are given and suggest that the method developed has potential in distinguishing identical pairs or sets.

8. Mavalwala J. (1977). *Dermatoglyphics: An International Bibliography*. The Hague: Mouton; Chicago: distributed in the USA and Canada by Aldine. ISBN 0202900517.

(Summary from bibliography***) The book provides an extensive bibliography on the subject of dermatoglyphics, including over 3000 references. It covers all areas where dermatoglyphics are used except identification work as used by law enforcement agencies. The time frame spans the end of the 19th century and concludes in 1973. It is an extremely valuable resource for research dealing with the distribution of level 1 features (essentially classes of general pattern and metrics of ridge counts) on both fingers and palms in a range of populations (small tribes to large ethnic groups).

9. Parsons N.R., Smith J.Q., Thönnnes E., et al, (2008). Rotationally Invariant Statistics for Examining the Evidence from the Pores in Fingerprints. *Law, Probability and Risk*, Vol. 7, No. 1, pp. 1-14.

Recent methodological advances in the processing of DNA evidence have begun to force a closer examination of assertions about the strength of other sorts of evidence. One traditional source of evidence is the fingerprint. Currently a print taken from a suspect is compared against a mark from a crime scene and a match declared using the judgment of an expert based on matching minutiae and the ridge patterns around these. However, such methods have proved difficult to quantify effectively. This has provoked the investigation of even finer features in the print and the mark. One set of such features are the many pores located along the ridges of the fingerprint. Is it possible to supplement expert judgments associated with a match with a more automatic and quantitative measure of the strength of evidence, based on pore information? The results of this preliminary analysis suggest we can. Our methodology is relatively transparent, using common statistics for two sample comparisons of point patterns. The results discussed here concern the matching of inked prints using grey-level imaging and complement previous studies, which tend to focus on the comparison of binarised images.

10. Reed T., Viken R.J. and Rinehart S.A. (2006). High Heritability of Fingertip Arch Patterns in Twin-Pairs. *American Journal of Medical Genetics*, Vol. 140A, pp. 263-271.

The presence of an arch pattern on at least one fingertip has previously been suggested as an autosomal dominant trait with reduced penetrance, although the examination of pedigrees with this trait segregating is also consistent with major gene or multifactorial inheritance. We used fingerprints in 2,484 twin-pairs to estimate heritability for the presence of at least one fingertip arch pattern. The frequency of arches in the entire sample was 4.3% (2,175/50,850), 5.5% in females and 3.2% in males. There were 267 twin-pairs concordant for the presence of an arch on any finger. Structural equation modeling was performed to contingency table data for five groups (MZ male, MZ female, DZ male, DZ female, and unlike-sexed DZ). The best fitting model, which allowed for the prevalence of arches to differ between males and females, had a heritability of 91%. There was some evidence for small dominant genetic effects in females and shared environmental effects in males, although both were not significant. With such high heritability, the search for specific genes influencing the occurrence of fingertip arch patterns is justified.

11. Roddy A.R. and Stosz J.D. (1997). Fingerprint Features- Statistical Analysis and System Performance Estimates. *Proceedings of the IEEE*, Vol. 85, No. 9, pp. 1390-1420.

As the need for personal authentication increases, many people are turning to biometric authentication as an alternative to traditional security devices. Concurrently, users and vendors of biometric authentication systems are searching for methods to establish system performance. This paper presents a model that defines the parameters necessary to estimate the performance of fingerprint-authentication systems without going through the rigors of intensive system testing inherent in establishing error rates. The model presented here was developed to predict the performance of the pore-based automated fingerprint-matching routine developed internally in the research and development division at the National Security Agency. This paper also discusses the statistics of fingerprint pores and the efficacy of using pores in addition to the traditionally used minutiae to improve system performance. In addition, this paper links together the realms of automated matching and statistical evaluations of fingerprint features. The result of this link provides knowledge of practical performance limits of any automated matching routine that uses pores or minutia features.

12. Steinman G. (2001). Mechanisms of Twinning: I. Effect of Environmental Diversity on Genetic Expression on Monozygotic Multifetal Pregnancies. *The Journal of Reproductive Medicine*, Vol. 46, No. 5, pp. 467-472.

(Summary from bibliography***) The author observed 250 fingerprints to examine the role of genetics and environmental factors affecting the general pattern of fingerprints in uniovular triplets, quadruplets, and

quintuplets. Results suggest that as the monozygotic set increases, the concordances in finger print pattern decreases within the set; twins were shown to have 88% concordance, 84% in triplets, 74% in quadruplets and only 71% concordance in quintuplets. These differences may be explained by a higher potential for mutation, as well as increased local dissimilarities within the uterine environment in larger sets.

II. What scientific literature describes the theoretical or biological basis for the distinctiveness of fingerprint characteristics?

13. Babler W.J. (1991). Embryologic Development of Epidermal Ridges and Their Configurations. *Dermatoglyphics- Science in Transition*, Vol. 27, No. 2, pp. 95-112.

(No abstract in original article***) This chapter provides a full overview of the studies in the relation to the embryologic development of friction ridge skin starting around 10 weeks post-fertilization. The understanding of the prenatal morphogenesis of dermatoglyphic traits is fundamental to the understanding of the relationship between dermatoglyphic features and birth defects. A full account is given of the development of the hand, the volar pads, the epidermal ridges (their proliferation on the dermis leads to the formation of minutiae) and dermal papillae. The process of ridge formation is not an event occurring simultaneously across the whole surface, but initiates at several points on the surfaces and spreads out until these fields of ridges finally meet. The factors affecting ridge configuration are discussed, including growth stress, topography of the volar pads, the surface distribution of nerves and bone development.

14. International Association for Identification, et al. (2011). Chapter 3: Embryology and Morphology of Friction Ridge Skin, in *The Fingerprint Sourcebook*. NCJ 225320.

(No abstract in original document***) This chapter discusses in detail how the basis of uniqueness lies in embryology and how the late embryological and early fetal development periods are crucial in the pattern formation of friction ridge skin. The author shows, using an extensive bibliography, that even though genetic information directs cellular function and determines someone's appearance, so many steps and factors are involved between the genesis of the DNA-encoded protein and the final development of the body that even identical twins would develop distinguishable friction ridge skin patterns.

15. Kücken M.U. (2007). Models for Fingerprint Pattern Formation. *Forensic Science International*, Vol. 171, No. 2-3, pp. 85-96.

There is currently no general agreement on the process by which fingerprint (epidermal ridge) patterns form. Nevertheless, many possible mechanisms have been proposed. Based on an extensive literature review and mathematical modeling, we argue that the pattern arises as the result of a buckling (folding) process in a cell layer of the epidermis. Using this model, we were able to explain the long-known observation that the pattern type is related to the geometry of the embryonal fingertip.

****Maceo A.V. (2005). The Basis for the Uniqueness and Persistence of Scars in the Friction Ridge Skin. *Fingerprint Whorld*, Vol. 31, No. 121, pp. 147-161.**

At the site of an injury, the skin must accomplish two things: 1) repair itself and 2) return to a state of normal maintenance. In the process of repair, the skin undergoes guided, but random, cell growth. The random nature of this process means that the exact appearance of the scar is not a foretold. Once repaired, the skin is brandished with a unique reminder of the injury and must return to the normal maintenance program. The skin of a scar follows the same maintenance program as uninjured skin and consistently reproduces the features on the surface of the scar.

****Maceo A.V. (2009). Friction Ridge Skin: Morphogenesis and Overview, in *Wiley Encyclopedia of Forensic Science*. John Wiley and Sons, Inc, pp. 1-9.**

The author provides an overview of the anatomy of friction ridge skin and its development *in utero*.

16. Wertheim K. and Maceo A. (2002). The Critical Stage of Friction Ridge and Pattern Formation. *Journal of Forensic Identification*, Vol. 52, No. 1, pp. 35-85.

This study provides an enhanced understanding of the biological structure and development of friction ridge skin for the latent print examiner, who is called upon to explain the scientific principles of latent print identification as based on permanence and uniqueness. Cellular attachments ensure permanence, while variable stresses and cellular distributions account for individuality on all “three levels” of detail. Volar patterning is dependent upon the tension across the surface of the developing skin during a critical stage of approximately 10.5 to 16 weeks estimated gestational age. Fingerprint ridge counts are predominantly affected by two combined timing events: the onset of epidermal cellular proliferation and the timing of the regression of the volar pads. Fingerprint pattern types are predominantly affected by the symmetry of the volar pad.

III. What scientific literature describes the permanence of fingerprint characteristics?

17. Galton F. (1892). *Finger Prints*. London: Macmillian and Co.

In addition to setting up a classification for fingerprint files, in 1892 Galton confirmed the three basic tenets of the science of friction ridge skin examination: immutability (based largely on data provided by Herschel), uniqueness, and inalterability. His studies established that fingerprints were stable from their formation during fetal life until death and even after. They therefore grow as the body grows and are sharply defined until an advanced age when incipient ridges appear as the skin becomes thinner. He also attested that fingerprints were, in severe cases, obliterated by some injuries that could leave permanent scars (hard labor, use of peculiar tools, severe cuts, etc.) and, therefore, make the ridge at that part undecipherable. He concluded that though the general pattern of volar skin may be altered through time, the number of ridges, their embranchments, and other minutiae remain unchanged.

18. Haylock S.E. (1987). Abnormalities in Dermatoglyphics. *Fingerprint Whorld*, Vol. 12, No. 47, pp. 71-75.

(Summary from bibliography***) Abnormalities in finger and palm prints can be congenital or acquired. The author divided these modifications into 3 groups: malformation of digits, disturbances in the epidermal ridges, and traumas and diseases. These alterations may make the print’s pattern unreadable and the author explains how some traumas can momentarily alter the quality of a print (minor cuts, contact with rough surface, etc.), and how some other alterations can be permanent (deep cuts, severe burns, etc.). This article contains a list of possible alterations and their effects on fingerprints.

19. International Association for Identification, et al. (2011). Chapter 2: Anatomy and Physiology of Adult Friction Ridge Skin, in *The Fingerprint Sourcebook*. NCJ 225320.

(No abstract in original document ***) This chapter provides an extensive account of the anatomy and physiology of friction ridge skin, building on an extensive biological literature. It describes the biological processes leading to the permanence of friction ridge skin. The effects of aging (flattened ridges and wrinkles) are explained. A full section also deals with the wound healing process and its effect on scar morphology.

****Maceo A.V. (2005). The Basis for the Uniqueness and Persistence of Scars in the Friction Ridge Skin. *Fingerprint Whorld*, Vol. 31, No. 121, pp. 147-161.**

At the site of an injury, the skin must accomplish two things: 1) repair itself and 2) return to a state of normal maintenance. In the process of repair, the skin undergoes guided, but random, cell growth. The random nature of this process means that the exact appearance of the scar is not a foretold. Once repaired, the skin is brandished with a unique reminder of the injury and must return to the normal maintenance program. The skin of a scar follows the same maintenance program as uninjured skin and consistently reproduces the features on the surface of the scar.

20. Modi S.K., Elliott S.J., Whetsone J. et al. (2007). Impact of Age Groups on Fingerprint Recognition Performance. *Automatic Identification Advanced Technologies, 2007 IEEE Workshop*, pp. 19-23.

Ever since the introduction of automated fingerprint recognition in law enforcement in the 1970's, it has been used in applications ranging from personal authentication to civilian border control. The increasing use of automated fingerprint recognition raises a challenge of processing a diverse range of fingerprints. The quality control module is important to this process because it supports consistent fingerprint detail extraction, which helps in identification/verification. Inherent feature issues, such as poor ridge flow, and interaction issues, such as inconsistent finger placement, have an impact on captured fingerprint quality, which eventually affects overall system performance. Aging results in loss of collagen; compared to younger skin, aging skin is loose and dry. Decreased skin firmness directly affects the quality of fingerprints acquired by sensors. Medical conditions such as arthritis may affect the user's ability to interact with the sensor, further reducing fingerprint quality. Because quality of fingerprints varies according to the user population's ages and because fingerprint quality has an impact on overall system performance, it is important to understand the significance of fingerprint samples from different age groups. This research examines the effects of fingerprints from different age groups on quality levels, minutiae count and performance of a minutiae-based matcher. The results show a difference in fingerprint image quality across age groups, most pronounced in the 62-and-older age group, confirming the work of Elliott and Sickler.

21. Olsen R.D. (1978). Occupational Damage and Bad Skin Condition, in *Scott's Fingerprint Mechanics*. Springfield: Charles C. Thomas, pp. 80-82.

(Summary from bibliography***) This section (pp. 80-82) explains how professional occupations (such as glazier, plasterer, labourer and battery worker) or other physical activities can result in temporary or permanent damages to the ridge structure of skin. These injuries can lead to minor or total obscuration of ridges and can sometimes interfere with classification, which must be taken into account by experts. The author provides a list of several physical injuries (like creasing, scars) and examples of altered inked prints.

22. Stücker M., Geil M., Hyeck S., et al. (2001). Interpapillary Lines- the Variable Part of the Human Fingerprint. *Journal of Forensic Sciences*, Vol. 46. No. 4, pp. 857-861.

The dermatoglyphic pattern of human palms and soles is individually unique and unchanging. Their prints show the course of the papillary ridges as papillary lines. Case reports and a few older studies of repeatedly taken fingerprints could, however, show that so-called interpapillary lines can develop between the papillary lines. The questions of this study were: How often do interpapillary lines occur? Can the differences between papillary and interpapillary ridges be quantified? Five-hundred and two ink prints of the palms and fingers from the archive of the Bochum Police Department were examined retrospectively. In 121 volunteers, the appearance of interpapillary lines was examined prospectively. From the latter collective, the fingerprints of 13 people with interpapillary lines and nine people without were examined further by taking two silicon prints and measuring them with laser profilometry. In 215 of the 502 ink prints (42.8%) interpapillary lines could be demonstrated. In those subjects younger than 20 years, they were less frequently observed (34.1%) than in those above the age of 20 (51.8%). In all cases using laser profilometry the interpapillary lines could be related to a corresponding interpapillary ridge. The

interpapillary ridge heights were $24.9 \pm 10.0 \mu\text{m}$, significantly lower than the papillary ridges, which measured $59.0 \pm 19.2 \mu\text{m}$. Interpapillary ridge widths were with $194.8 \pm 65.1 \mu\text{m}$ significantly narrower as compared to $435.5 \pm 57.4 \mu\text{m}$ in the papillary ridge. Those papillary ridges, between which interpapillary ridges were found, were significantly further apart from each other ($610.5 \pm 78.9 \mu\text{m}$) than those without interpapillary ridges ($484.9 \pm 70.6 \mu\text{m}$). During the course of a lifetime, new ridges between the regular papillary ridges can develop or manifest. The fact that interpapillary lines are more frequently found on the right hands in men and those with increasing age is consistent with the theory that they correspond to degenerative changes and with sensitivity of touch.

MINUTIAE SAMPLE SUFFICIENCY

IV. What scientific literature describes the likelihood that observed minutiae originate from the same or different fingerprints?

23. Champod C. (2000). Fingerprints (Dactyloscopy): Standard Proof, in *Encyclopedia of Forensic Sciences*, J. Siegel, P. Saukko and G. Knupfer, Eds. London: Academic Press, pp. 884-890.

(Summary from bibliography***) On a large sample of inked fingerprints from male subjects (321 right loops on right index fingers, 365 right loops on right middle fingers, 118 left loops on left middle fingers, and 173 whorls on right index fingers), the author investigates different links between minutiae and other variables. General pattern, finger number, ridge count, minutiae type, minutiae position, minutiae direction, minutiae density and the distance between several points making up a combined minutiae are all investigated. The results obtained were as follows: the number of minutiae depends on the location of the fingerprint area being studied (the delta and the center being the most dense); the presence/absence of a delta and center influences greatly the number of neighboring minutiae; the top parts of fingerprints show the least variation in terms of minutiae number and minutiae distribution in the studied areas mainly follows a Poisson distribution. When comparing two fingerprints using the model presented in this thesis, results tend to support the proposition that individualizing a fingerprint requires more than a fixed number of minutiae that should be in agreement (without discrepancy) between two impressions. On the other hand this probabilistic estimate for a given configuration of minutiae depends on the anatomical position of the area under comparison.

24. Dass S.C., Pankanti S., Prabhakar S., et al. (2009). Individuality of Fingerprints, in *Encyclopedia of Biometrics*. SZ Li and A Jain, Eds. New York: Springer Verlag, pp. 741-751.

Fingerprint individuality is the study of the extent of uniqueness of fingerprints. It is the most important measure to be ascertained when fingerprint evidence is presented in court by experts. A measure of fingerprint individuality reflects the amount of uncertainty associated with the experts' decisions, which is primarily due to the variability of feature characteristics in a pair of fingerprints. This inherent variability can cause random matching between the pair of fingerprints even if they are not from the same person. Fingerprint individuality aims to characterize this randomness in matching them quantitatively in terms of statistical models.

25. Egli N.M. (2009). *Interpretation of Partial Fingermarks Using an Automated Fingerprint Identification System*. Ph.D. in Forensic Science, Faculty of Law and Criminal Sciences, University of Lausanne, Lausanne.

(Summary from bibliography***) The model is established based on AFIS scores, thus taking into account minutia position and orientation for groups of minutiae. A very large background database (over 600,000 fingerprints, from over 60,000 ten-print cards) was used for establishment and testing of the model. Within-finger variability is based on repeated impressions of a given finger, then the scores obtained are modeled (rather than modeling the characteristics directly); impressions from the finger of one donor are used for the establishment of the model and impressions from a different finger are used for testing. Between-finger variability is extracted by confronting the

crime-scene mark to the entire database. Testing using same source and different source comparisons (where the 'different source' reference prints were chosen randomly) showed low rates of misleading evidence.

26. Gutiérrez E., Galera V., Martínez J.M., et al. (2007). Biological Variability of the Minutiae in the Fingerprints of a Sample of the Spanish Population. *Forensic Science International*, Vol. 172, No. 2-3, pp. 98-105.

The minutiae, a term coined by Galton to refer to the small peculiarities present along the length of every isolated ridge, or characteristic points, a term used primarily by the Spanish Police Scientists, have an inter- and intra-population variability that has not been extensively studied. However, these peculiarities constitute the bases for the fingerprint identification of individuals in the field of criminology. Using the adhesive paper and graphite method, the fingerprints of 200 students, 100 males and 100 females, with ages ranging between 20 and 35, have been taken at the University of Alcalá (Madrid). From this sample, the distal phalanx of the index finger of the right hand has been studied. The total count of the minutiae, as well as that of each different type, was made of the entire print area and inside and outside of a circle with a radius of 18 ridges. The highest frequencies were of ridge endings, followed by bifurcations and convergences, all others appearing with frequencies of less than 5%. The distribution of the minutiae was not homogeneous for the area of the fingerprint (inside and outside the circle). In the study of minutiae, statistically significant differences were found between the sexes and between the different types of general pattern (arches, loops, and whorls).

27. Gutiérrez-Redomero E., Alonso-Rodríguez C., Hernandez-Hurtado L.E., et al. (2011). Distribution of the Minutiae in the Fingerprints of a Sample of the Spanish Population. *Forensic Science International*, Vol. 208, No. 1-3, pp. 79-90.

One of the fundamental aspects of the process of identification through fingerprints is the comparison of the minutiae between the fingermark obtained at the scene of the crime and the suspect's corresponding finger. There is no scientific basis in this process that allows the use of numerical standards, such as those kept in different countries, to obtain the identification. The recent mistakes made in the field of dactyloscopy, together with the growing rigor and scrutiny that forensic evidence undergoes in the legislative and scientific areas, have resulted in the need to reconsider some of the basic principles that support this discipline. A probabilistic estimation of the evidential value is especially necessary; therefore, it is indispensable to know and quantify the variability of the features used in the identification process. The sample studied for this research was obtained from 100 Caucasian men and 100 Caucasian women from the Spanish population, which amounts to a total of 2000 fingerprints. The different types of minutiae were located, identified, and quantified visually on the fingerprint, in four sectors, and inside and outside of a circle, whose radius cut, perpendicularly, fifteen ridges starting from the center cut of the axes that defined the sectors. According to the results obtained in this study, through dactyloscopic identification the weight of the evidence of a minutia, such as the ridge endings, with frequencies between 55% and 65%, according to the area and gender evaluated, cannot be the same as that of a bifurcation or convergence, with frequencies of 13–18% or those of other minutiae that show frequencies less than 3%. The significant differences found in the topological distribution of the endings, bifurcations, and convergences show the need to take into account, for its demonstrational value, the finger area in which they are evaluated. The significant association observed between the types of minutiae and the different fingers revealed a greater frequency of endings on the thumb and index fingers, and bifurcations and convergences on the middle, ring, and little fingers.

28. Jain A.K., Prabhakar S. and Pankanti S. (2002). On the Similarity of Identical Twin Fingerprints. *Pattern Recognition*, Vol. 35, No. 11, pp. 2653-2663.

Reliable and accurate verification of people is extremely important in a number of business transactions as well as access to privileged information. Automatic verification methods based on physical biometric characteristics, such as fingerprint or iris, can provide positive verification with a very high accuracy. However, the biometrics-based methods assume that the physical characteristics of an individual (as captured by a sensor) used for verification are sufficiently unique to distinguish one person from another. Identical twins have the closest genetics-based

relationship and, therefore, the maximum similarity between fingerprints is expected to be found among identical twins. We show that a state-of-the-art automatic fingerprint verification system can successfully distinguish identical twins, though with a slightly lower accuracy than non-twins.

29. Liu Y. and Srihari S. (2009). A Computational Discriminability Analysis on Twin Fingerprints, in *Computational Forensics*, pp. 43-54.

Sharing similar genetic traits makes the investigation of twins an important study in forensics and biometrics. Fingerprints are one of the most commonly found types of forensic evidence. The similarity between twins' prints is critical to establish the reliability of fingerprint identification. We present a quantitative analysis of the discriminability of twin fingerprints on a new data set (227 pairs of identical twins and fraternal twins) recently collected from a twin population using both level 1 and level 2 features. Although the patterns of minutiae among twins are more similar than in the general population, the similarity of fingerprints of twins is significantly different from that between genuine prints of the same finger. Twins fingerprints are discriminable with a 1.5% ~ 1.7% higher EER than non-twins. And identical twins can be distinguished by examining fingerprint with a slightly higher error rate than fraternal twins.

30. Loesch D.Z. and Martin N.G. (1984). Relationships Between Minute Characteristics of Finger Ridges and Pattern Size and Shape. *Annals of Human Biology*, Vol. 11, No. 2, pp. 125-132.

We present results of a multivariate analysis of several characteristics of the ridge pattern on the tip of fingers III and IV of the right hand, with special consideration of the minute properties of ridges. The analysis is based on a sample of 38 females and 63 males of European origin. Results of factor analysis for the male sample reveal that pattern type and size load on one factor. Furthermore, there are separate factors for the two types of minutiae. If characters on both fingers are considered jointly, there are seven factors, identical for both fingers on some variables, but unique for minutiae, especially for junctions, on each finger. These results are consistent with evidence obtained in our previous study: junctions and pattern type are largely independent correlates with tactile sensitivity.

31. Neumann C., Champod C., Puch-Solis R., et al. (2007). Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae. *Journal of Forensic Sciences*, Vol. 52, No. 1, pp. 54-64.

Recent challenges to fingerprint evidence have brought forward the need for peer-reviewed scientific publications to support the evidential value assessment of fingerprint. This paper proposes some research directions to gather statistical knowledge of the within-source and between-sources variability of configurations of three minutiae on fingermarks and fingerprints. This paper proposes the use of the likelihood ratio (LR) approach to assess the value of fingerprint evidence. The model explores the statistical contribution of configurations of three minutiae using Tippett plots and related measures to assess the quality of the system. Features vectors used for statistical analysis have been obtained following a preprocessing step based on Gabor filtering and image processing to extract minutia position, type, and direction. Spatial relationships have been coded using Delaunay triangulation. The metric used to assess similarity between two feature vectors is based on an Euclidean distance measure. The within-source variability has been estimated using a sample of 216 fingerprints from four fingers (two donors). Between-sources variability takes advantage of a database of 818 ulnar loops from randomly selected males. The results show that the data-driven approach adopted here is robust. The magnitude of LRs obtained under the prosecution and defense propositions rests upon the major evidential contribution that small portions of fingerprint, containing three minutiae, can provide regardless of its position on the general pattern.

***32. Neumann C., Evett I.W. and Skerrett J. (2012). Quantifying the Weight of Evidence from a Forensic Fingerprint Comparison: A New Paradigm. *Journal of the Royal Statistical Society*, Vol. 175, Part 2, pp. 1-26.**

The fingerprint has, with considerable justification, come to be regarded as the acme of forensic identification. Over the past century, millions of cases have been resolved worldwide because of marks left at crime scenes. The comparison methodology has not evolved greatly during its history and it is universal practice to present fingerprint evidence to a court as a categorical opinion of identification or exclusion, or to classify the evidence as inconclusive and not to report it. There has been a growing movement to supplement the fingerprint examination process by one that has a statistical model, supported by appropriate databases for calculating numerical measures of weight of evidence. The movement calls for the establishment of a logical framework for informing conclusions, based on explicit assumptions and data and open to revision and improvement. The aim is to enable the numerical evaluation of evidence that would currently be reported as a categorical identification and also of evidence that would currently be classified as inconclusive. The paper presents the results of a project carried out by the Forensic Science Service that aims to attain this goal. After a historical review, we describe a formal model for assigning numerical values to configurations of minutiae in fingerprints. We describe how the parameters of the model have been optimized to take account of inter-operator variability and distortion of the finger pad, and we present the results of a substantial validation experiment that was based on searches that have been carried out on the US national fingerprint database of approximately 600 million fingerprints.

33. Okajima M. (1967). Frequency of Epidermal-Ridge Minutiae in the Calcar Area of Japanese Twins. *American Journal of Human Genetics*, Vol. 19, No. 5, pp. 660-673.

In order to investigate the individual differences in minutiae, the author has proposed a method of classifying minutiae based only on their frequency in the calcar area of the sole. This area is suitable because the dermal ridges run transversely in parallel and patterns are rare. Therefore, the frequency of minutiae is not influenced by the presence of turns of ridges. Furthermore, it was found that dermal ridges in this area show distinct individual differences in the arrangement and frequency of minutiae, as will be illustrated. The method and the results obtained from German and Japanese twins were reported earlier (Okajima, 1966). In the present report, additional results are presented on Japanese twins, and some of the previous conclusions are revised.

34. Okajima M. and Usukara K. (1984). Quantitative and Genetic Features of Epidermal Ridge Minutiae on the Palm of Twins. *Human Heredity*, Vol. 34, pp. 285-290.

The total minutia count (TMC) and fork index (FI), i.e., the proportion of forks in total minutiae, were examined in palm prints of 20 pairs each of monozygotic and dizygotic Japanese twins. No bilateral difference was found in either trait. The mean of the TMC was higher in males than in females, but the FI showed no sex difference. Both traits presented relatively high intra-class correlation coefficients between monozygotic twins, and genetic control of these traits is suggested. However, no correlation was observed between these two traits. A slight correlation was observed between the TMC and the total palmar interdigital ridge count. Some methodological problems concerned with classification criteria were discussed.

35. Pankanti S., Prabhakar S. and Jain A.K. (2002). On the Individuality of Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1010-1025.

Fingerprint identification is based on two basic premises: 1) persistence (the basic characteristics of fingerprints do not change with time) and 2) individuality (the fingerprint is unique to an individual). The validity of the first premise has been established by the anatomy and morphogenesis of friction ridge skin. While the second premise has been generally accepted to be true based on empirical results, the underlying scientific basis of fingerprint individuality has not been formally established. As a result, the validity of fingerprint evidence is now being

challenged in several court cases. A scientific basis for establishing fingerprint individuality will not only result in the admissibility of fingerprint identification in the courts of law, but will also establish an upper bound on the performance of an automatic fingerprint verification system. We address the problem of fingerprint individuality by quantifying the amount of information available in minutiae features to establish a correspondence between two fingerprint images. We derive an expression that estimates the probability of a false correspondence between minutiae-based representations from two arbitrary fingerprints belonging to different fingers. For example, the probability that a fingerprint with 36 minutiae points will share 12 minutiae points with another arbitrarily chosen fingerprint with 36 minutiae points is 6.10×10^{-8} . These probability estimates are compared with typical fingerprint matcher accuracy results. Our results show that 1) contrary to the popular belief, fingerprint matching is not infallible and leads to some false associations, 2) while there is an overwhelming amount of discriminatory information present in the fingerprints, the strength of the evidence degrades drastically with noise in the sensed fingerprint images, 3) the performance of the state-of-the-art automatic fingerprint matchers is not even close to the theoretical limit, and 4) because automatic fingerprint verification systems based on minutia use only a part of the discriminatory information present in the fingerprints, it may be desirable to explore additional complementary representations of fingerprints for automatic matching.

36. Sarkar N.C. (2004). *Finger Ridge Minutiae: Classification, Distribution and Genetics*. Kolkata, India: Anthropological Survey of India.

(Summary from bibliography***) Based on different tests carried out on a sample of 2010 fingerprints from 201 males and 2020 fingerprints from 202 females, concerning both level I and level II features, and integrating familial analysis, the author remarks on the following trends on level 2 features (the work offers in depth data on other dermatoglyphic traits):

1. The distribution of total finger ridge counts and total finger minutiae count are non-Gaussian and characterized by flat peaks. Deviation from normality is not statistically significant for both total ridge counts and total finger minutiae counts.
2. Ridge endings, out of eight minutiae types that were considered, show a preponderance over other types, and the least frequently occurring minutiae is the connecting ridge. Two minutiae types, ridge endings and forks, together represent more than 80% of the total minutiae count, while the remaining six types represent only about 20%.
3. Maximum occurrence of minutiae has been noted in the thumb following by the ring finger, while minimum occurrence has been noted either in index or little finger.
4. For total ridge counts and total minutiae count, fingers could be classified into two groups (a) thumb and ring finger, (b) index, middle and little finger. Both total ridge count and total minutiae count also bear high hereditary significance.
5. Mean minutiae count in males is significantly higher than that in females.
6. Values of correlation for familial combinations except parent-parent are positive and most of them are statistically significant. The values of correlation for bifurcation and ridge ending are similar to each other and are at par with correlation values for total minutiae count, but higher than those of other minutiae types.
7. Two minutiae types, bifurcation and ridge ending along with a total minutiae count bear strong hereditary significance. These two minutiae types, out of eight different minutia types, show very high heritability. Because of this high hereditary significance, fork and ridge ending may be useful as marker characters in genetic analysis as well as in personal identification.

8. Mean values of total ridge counts and total minutiae count for whorls are significantly higher than those for loops.
9. The values of the ratio between mean minutiae count and mean ridge count is maximal in thumbs and decreases gradually from radial to ulnar fingers and finally reduced to its minimum in little fingers, irrespective of pattern type, hand, and sex.

37. Seweryn P. (2005). Frequency of Minutiae on Left and Right Hand Index Fingers. *Problemy Kryminalistyki*, No. 247, pp. 40-46.

This is a study of the statistical distribution of minutiae observed on 400 fingerprints (right and left index) from 200 individuals (178 males and 22 females) held on police files from Southern Poland. Minutiae were manually counted and classified according to their types (ridge ending, bifurcations and a set of combined minutiae). Relative frequencies are given, showing the large relative abundance of ridge endings (opening and ending) and bifurcations compared to the other types of minutiae considered. One main outcome is the observation that the density of minutiae in the index varies according to the general pattern of the fingerprint (arch, loop or whorl). A rounded mean number of minutiae of 86 minutiae was counted on whorls, 80 on loops and 72 on arches.

38. Srihari S.N., Srinivasan H. and Fang G. (2008). Discriminability of Fingerprints in Twins. *Journal of Forensic Identification*, Vol. 58, No. 1, pp. 109-127.

A study of the discriminability of fingerprints of twins is presented. The fingerprint data used are of high quality and quantity because of a predominantly young subject population of 298 pairs of twins whose ten-prints were captured using a livescan device. Discriminability using level 1 and level 2 features is independently reported. The level 1 study visually classified by humans each fingerprint into one of six categories (right loop, left loop, whorl, arch, twin loop, and tented arch). It was found that twins are much more likely (55%) to have the same level 1 classification when compared to the general population (32%). The level 2 study compared minutiae (ridge endings and bifurcations). This was done by a minutiae-based automatic fingerprint identification algorithm that provided a score (0-350) given a pair of fingerprints. Scores were computed for corresponding fingers from both twins and non-twins. Five distributions of scores were determined: twins, non-twins, identical twins, fraternal twins, and genuine scores from the same finger. Using the Kolmogorov-Smirnov test to compare distributions, the following inferences are made: twins are different from genuines, twins are different from non-twins, and identical twins are the same as fraternal twins. The main conclusion is that, although the patterns of minutiae among twins are more similar than in the general population, they are still discriminable.

39. Stoney D.A. (2001). Measurement of Fingerprint Individuality, in *Advances in Fingerprint Technology*, second ed. HC Lee and RE Gaensslen, Eds. Boca Raton: CRC Press, pp. 327-387.

(Summary from bibliography***) A review of models measuring the rarity of fingerprint features from 1892 to 1999 is proposed. Out of the fifteen models presented, six are grouped together in a single class because they approach the question in the same way; 10 different approaches are therefore presented. All approaches are explained concerning the characteristics used and their description/measurement, as well as the probabilistic approaches employed and the results obtained. All of these approaches concern minutiae described using type and position or only one of these two elements. Furthermore, some of these models address only the rarity of a given configuration, while others include a consideration of the correspondence observed during a comparison, or rather, a consideration of differences that are allowed during the comparison. All of the models yield numbers that highlight the variability of fingerprints. After this description of the data and approach used, the models are critically discussed. Finally, the author's conclusion highlights the weakness of the scientific foundation of fingerprint individuality, based on the fact that while a number of models had been proposed, these were in general established on the basis of little data (with one exception), and in 2001, none of the models had been tested.

40. Su C. and Srihari S. (2009). Probability of Random Correspondence for Fingerprints, in *Computational Forensics: Third International Workshop, IWCF 2009*. LNCS 5718, ZJM Geradts, KY Franke and CJ Veenman, Eds. Berlin Heidelberg: Springer Verlag, pp. 55-66.

Individuality of fingerprints can be quantified by computing the probabilistic metrics for measuring the degree of fingerprint individuality. In this paper, we present a novel individuality evaluation approach to estimate the probability of random correspondence (PRC). Three generative models are developed, respectively, to represent the distribution of fingerprint features: ridge flow, minutiae, and minutiae together with ridge points. A mathematical model that computes the PRCs are derived based on the generative models. Three metrics are discussed in this paper: (i) PRC of two samples, (ii) PRC among a random set of n samples (n PRC) and (iii) PRC between a specific sample among n others (specific n PRC). Experimental results show that the theoretical estimates of fingerprint individuality using our model consistently follow the empirical values based on the NIST4 database.

****Zhu Y., Dass S.C. and Jain A.K. (2007). Statistical Models for Assessing the Individuality of Fingerprints. *IEEE Transactions on Information Forensics and Security*, Vol. 2, No. 3, pp. 391-401.**

Following the Daubert ruling in 1993, forensic evidence based on fingerprints was first challenged in the 1999 case of the U.S. versus Byron C. Mitchell and, subsequently, in 20 other cases involving fingerprint evidence. The main concern with the admissibility of fingerprint evidence is the problem of individualization, namely, that the fundamental premise for asserting the uniqueness of fingerprints has not been objectively tested and matching error rates are unknown. In order to assess the error rates, we require quantifying the variability of fingerprint features, namely, minutiae in the target population. A family of finite mixture models has been developed in this paper to represent the distribution of minutiae in fingerprint images, including minutiae clustering tendencies and dependencies in different regions of the fingerprint image domain. A mathematical model that computes the probability of a random correspondence (PRC) is derived based on the mixture models. A PRC of 2.25×10^{-6} corresponding to 12 minutiae matches was computed for the NIST4 Special Database, when the numbers of query and template minutiae both equal 46. This is also the estimate of the PRC for a target population with a similar composition as that of NIST4.

FINGERPRINT QUALITY

V. What scientific literature describes how latent print quality affects the probability of correctly matching (e.g. individualization or identification) or excluding fingerprints? (How is the quality currently assessed? Are there methods that enhance images without introducing distortions? What are the effects of aging or other environmental factors on quality? What are the effects of print development/lifting techniques on print quality? What are the effects of reverse or positive/negative prints?)

41. Alonso-Fernandez F., Fierrez J., Ortega-Garcia J., et al. (2007). A Comparative Study of Fingerprint Image-Quality Estimation Methods. *IEEE Transactions on Information Forensics and Security*, Vol. 2, No. 4, pp. 734-743.

One of the open issues in fingerprint verification is the lack of robustness against image-quality degradation. Poor-quality images result in spurious and missing features, thus degrading the performance of the overall system. Therefore, it is important for a fingerprint recognition system to estimate the quality and validity of the captured fingerprint images. In this work, we review existing approaches for fingerprint image-quality estimation, including the rationale behind the published measures and visual examples, showing their behavior under different quality conditions. We have also tested a selection of fingerprint image-quality estimation algorithms. For the experiments, we employ the BioSec multimodal baseline corpus, which includes 19 200 fingerprint images from 200 individuals, acquired in two sessions with three different sensors. The behavior of the selected quality

measures is compared, showing high correlation between them in most cases. The effect of low-quality samples in the verification performance is also studied for a widely available minutiae-based fingerprint matching system.

42. Alonso-Fernandez F., Roli F., Marcialis G.L., et al. (2007). Comparison of Fingerprint Quality Measures Using an Optical and a Capacitive Sensor, in *Biometrics: Theory, Applications and Systems. First IEEE International Conference on BTAS*, pp. 1-6.

Although several image quality measures have been proposed for fingerprints, no work has taken into account the differences among capture devices, and how these differences affect the image quality. In this paper, several representative measures for assessing the quality fingerprint images are compared using an optical and a capacitive sensor. The capability to discriminate between images of different quality and its relationship with the verification performance is studied. We report differences depending on the sensor, and interesting relationships between sensor technology and features used for quality assessment are also pointed out.

***43. Amorós González B. and de Puit M. (2014). A Model Study into the Effects of Light and Temperature on the Degradation of Fingerprint Constituents. *Science and Justice*, Vol. 54, No. 5, pp. 346-350.**

Determining the order of events in a criminal investigation can be crucial in many ways. Although the use of fingermarks is one of the most important forensic tools as a modality of individualization, placing them in time is a long-lasting difficulty. In the study presented in this paper, we describe the effects of light and the temperature on depositions of standard solutions containing squalene and cholesterol on paper. The results obtained give a clear view on what type of processes take place during a temperature- and a light-influenced degradation of the two materials in this study. Using well known methods for the determination of reaction constants, it was determined that the degradation mechanism of cholesterol is following a radical route, where the decay of squalene is more affected by temperature.

44. Anthonioz A., Egli N., Champod C., et al. (2011). Investigation of the Reproducibility of Third-Level Characteristics. *Journal of Forensic Identification*, Vol. 61, No. 2, pp. 171-192.

The process of comparing fingermarks recovered from a crime scene with the fingerprint taken from a known individual involves the characterization and comparison of different ridge details on both the mark and the print. Fingerprint examiners commonly classify these characteristics into three different groups, depending on their level of discriminating power. It is commonly considered that the general pattern of the ridge flow constitutes first level detail, specific ridge flow and minutiae (e.g., ending ridges, bifurcations) constitutes second-level detail, and fine ridge details (e.g., pore positions and shapes) are described as third-level detail.

***45. Barros R.M., Faria B.E.F. and Kuckelhaus S.A.S. (2013). Morphometry of Latent Palmprints as a Function of Time. *Science and Justice*, Vol. 53, No. 1, pp. 402-208.**

In many crimes, the elapsed time between production and collecting fingermark traces is crucial. and a method able to detect the aging of latent prints would represent an improvement in forensic procedures. Considering that as the latent print gets older, substantial changes in the relative proportion of individual components secreted by skin glands could affect the morphology of ridges, morphometry could be a potential tool to assess the aging of latent fingermarks. Then, considering the very limited research in the field, the present work aims to evaluate the morphometry of latent palmprint ridges, as a function of time in order to identify an aging pattern. The latent marks were deposited by 20 donors on glass microscope slides considering pressure and contact angle, and then were maintained under controlled environmental conditions. The morphometric study was conducted on marks developed with magnetic powder in 7 different time intervals after deposition (0, 5, 10, 15, 20, 25 or 30 days); 60 ridges were evaluated for each developed mark. The results showed that: 1) the method for the replacement and mixing of skin secretions on the palm was appropriate to ensure reproducibility of latent prints, and 2) considering

the studied group, there was a time-dependent reduction in the width of ridges and on the percentage of visible ridges over 30 days. Results suggest the possibility of using the morphometric method to determine an aging profile of latent palmprints on glass surface for forensic purposes.

46. Blomeke C.R., Modi S.K. and Elliott S.J. (2008). Investigating the Relationship between Fingerprint Image Quality and Skin Characteristics, in 42nd Annual IEEE International Carnahan Conference on Security Technology, pp. 158-161.

This paper reports the correlations between skin characteristics, such as moisture, oiliness, elasticity, and temperature of the skin, and fingerprint image quality across three sensing technologies. Fingerprint images from the index finger of the dominant hand of 190 individuals were collected on nine different fingerprint sensors. The sensors included four capacitance sensors, four optical sensors, and one thermal fingerprint sensor. Skin characteristics including temperature, moisture, oiliness and elasticity, were measured prior to the initial interaction with each of the individual sensors. The analysis of the full dataset indicated that the sensing technology and interaction type (swipe or touch) were moderately and weakly correlated, respectively, with image quality scores. Correlation analysis between image quality scores and the skin characteristics were also made on subsets of data, divided by the sensing technology. The results did not identify any significant correlations. This indicates that further work is necessary to determine the type of relationship between the variables, and how they affect image quality and matching performance.

47. Changlong J., Hakil K., Xuenan C., et al. (2009). Comparative Assessment of Fingerprint Sample Quality Measures Based on Minutiae-Based Matching Performance, in *Electronic Commerce and Security. Second International Symposium on ISECS '09*.

Fingerprint sample quality is one of the major factors influencing the matching performance of fingerprint recognition systems. The error rates of fingerprint recognition systems can be decreased significantly by removing poor quality fingerprints. The purpose of this paper is to assess the effectiveness of individual sample quality measures on the performance of minutiae-based fingerprint recognition algorithms. Initially, the authors examined the various factors that influenced the matching performance of the minutiae-based fingerprint recognition algorithms. Then, the existing measures for fingerprint sample quality were studied, and the more effective quality measures were selected and compared with two image quality software packages, (NFIQ from NIST, and QualityCheck from Aware Inc.) in terms of matching performance of a commercial fingerprint matcher (Verifinger 5.0 from Neurotechnologija). The experimental results over various fingerprint verification competition (FVC) datasets show that even a single sample quality measure can enhance the matching performance effectively.

***48. Cohen Y., Rozen E., Azoury, M., et al. (2012). Survivability of Latent Fingerprints Part 1: Adhesion of Latent Fingerprints to Smooth Surfaces. *Journal of Forensic Identification*, Vol. 62, No. 1, pp. 47-53.**

A latent print was developed on an aluminum window frame more than two years after it had been deposited. The ability to develop a fingerprint after such a long time is probably due to a "fixation" phenomenon to the metal frame. To understand this unusual case, we simulated the event in the laboratory.

49. Dass S.C. (2010). Assessing Fingerprint Individuality in Presence of Noisy Minutiae. *IEEE Transactions on Information Forensics and Security*, Vol. 5, No. 1, pp. 62-70.

Fingerprint image quality is an important source of intraclass variability. When the underlying image quality is poor, human experts as well as automatic systems are more likely to make errors in minutiae detection and matching by either missing true features or detecting spurious ones. As a consequence, fingerprint individuality estimates change depending on the quality of the underlying images. The goal of this paper is to quantitatively study the effect of noise in minutiae detection and localization resulting from varying image quality on fingerprint individuality. The measure of fingerprint individuality is modeled as a function of image quality via a random

effects model and methodology for the estimation of unknown parameters is developed in a Bayesian framework. Empirical results on two databases, one in-house and another publicly available, demonstrate how the measure of fingerprint individuality increases as image quality becomes poor. The measure corresponding to the “12-point match” with 26 observed minutiae in the query and template fingerprints increases by several orders of magnitude when the fingerprint quality degrades from “best” to “poor.”

***50. De Alcaraz-Fossoul J., Mestres Patris C., Balaciart Muntaner, et al. (2013). Determination of Latent Fingerprint Degradation Patterns- A Real Fieldwork Study. *International Journal of Legal Medicine*, Vol. 127, No. 1, pp. 857-870.**

For over a century, law enforcement agencies, forensic laboratories, and penal courts worldwide have used fingerprint impressions as reliable and conclusive evidence to identify perpetrators of criminal activity. Although fingerprint identification has been repeatedly proven as one of the most robust and definite forensic techniques, a measure of the rate at which latent fingerprints degrade over time has not been established effectively. Ideally, criminal investigators should be able not only to place any given individual at a crime scene but also be able to date the moment any latent fingerprints were deposited at the location. The present report aims to determine particular visual patterns of degradation of latent fingerprints exposed to certain monitored laboratory conditions simulating those in the field. Factors considered include temperature, relative humidity, air currents, composition of fingerprint depositions (sebaceous and eccrine), various exposures to daylight (direct, penumbra and darkness), and type of physical substrate (glass and plastic) over a period of 6 months. The study employs a titanium dioxide-based powder as developer. Our results indicate that, contrary to common belief, certain latent fingerprints exposed to direct sunlight indoors degrade similarly to those in the dark, where environmental conditions are more constant. While all sebaceous latent fingerprints on glass are still useful for identification after 6 months, diverse results are obtained with impressions on plastic; these demonstrate a much higher and faster degree of decay, making identification difficult or impossible, especially for eccrine depositions.

51. Gupta A., Buckley K. and Sutton R. (2008). Latent Fingermark Pore Area Reproducibility. *Forensic Science International*, Vol. 179, No. 2-3, pp. 172-175.

The study of the reproducibility of friction ridge pore detail in fingermarks is a measure of their usefulness in personal identification. Pore area in latent prints developed using cyanoacrylate and ninhydrin were examined and measured by photomicrography using appropriate software tools. The data were analyzed statistically and the results showed that pore area is not reproducible in developed latent prints, using either of the development techniques. The results add further support to the lack of reliability of pore area in personal identification.

52. Hicklin R.A. et al. (2011). Latent Fingerprint Quality: A Survey of Examiners. *Journal of Forensic Identification*, Vol. 61, No. 4, pp. 385-418.

A survey of latent print examiners was conducted to determine how they assess fingerprint quality. Participating examiners performed detailed anonymous assessments of both the local and overall quality characteristics of latent and exemplar fingerprint images, using a custom-designed software application. Eighty-six latent print examiners from federal, state, local, international, and private sector laboratories each spent 8 to 12 hours assessing the quality of approximately 70 fingerprint images. The fingerprints were overlapping subsets of 1,090 latent and exemplar fingerprint images derived from the National Institute of Standards and Technology (NIST) Special Database 27 and a Federal Bureau of Investigation (FBI) Laboratory dataset of images. An analysis of the results shows the extent of consistency between examiners in value determinations; the relationships between the overall perceived quality of a print and the size of clear ridge detail; and the relationships between quality, size, and correct pattern classification. An analysis of the examiners' subjective assessments of fingerprint quality revealed information useful for the development of guidelines, metrics, and software tools for assessing fingerprint quality.

***53. Hicklin R.A., Buscaglia J. and Roberts M.A. (2013). Assessing the Clarity of Friction Ridge Impressions. *Forensic Science International*, Vol. 226, pp. 106-117.**

The ability of friction ridge examiners to correctly discern and make use of the ridges and associated features in finger or palm impressions is limited by clarity. The clarity of an impression relates to the examiner's confidence that the presence, absence, and attributes of features can be correctly discerned. Despite the importance of clarity in the examination process, there have not previously been standard methods for assessing clarity in friction ridge impressions. We introduce a process for annotation, analysis, and interchange of friction ridge clarity information that can be applied to latent or exemplar impressions. This paper (1) describes a method for evaluating the clarity of friction ridge impressions by using color-coded annotations that can be used by examiners or automated systems; (2) discusses algorithms for overall clarity metrics based on manual or automated clarity annotation; and (3) defines a method of quantifying the correspondence of clarity when comparing a pair of friction ridge images, based on clarity annotation and resulting metrics. Different uses of this approach include examiner interchange of data, quality assurance, metrics as an aid in automated fingerprint matching.

***54. Holt D. (2013). Determining the Quality and Sustainability of Friction Ridge Deposits on Envelopes Sent Through the Postal System. *Journal of Forensic Identification*, Vol. 63, No. 3, pp. 247- 253.**

This study explored the quality and sustainability of test friction ridge deposits on standard envelopes that were sent through the postal system. The test envelopes were collected and chemically treated using 1,2-indanedione with ZnCl₂ (IND-Zn) to develop latent fingerprint impressions. The test envelopes were assessed to determine the extent to which the deposit was present, the level of friction ridge detail, and whether any foreign superimposed fingerprints deposited during the distribution process had affected the quality of the deposit. The research provided a statistical overview, whereby the greater number of deposits were strongly present (sustainable), and the majority of these deposits exhibited friction ridge detail that was suitable for comparison and identification purposes (quality). Only a relatively small number of deposits were affected by the physical handling of test envelopes.

****Kalka N.D. and Hicklin R.A. (2014). On Relative Distortion in Fingerprint Comparison. *Forensic Science International*, Vol. 244, pp. 78-84.**

When fingerprints are deposited, non-uniform pressure in conjunction with the inherent elasticity of friction ridge skin often causes linear and non-linear distortions in the ridge and valley structure. The effects of these distortions must be considered during analysis of fingerprint images. Even when individual prints are not notably distorted, relative distortion between two prints can have a serious impact on comparison. In this paper we discuss several metrics for quantifying and visualizing linear and non-linear fingerprint deformations, and software tools to assist examiners in accounting for distortion in fingerprint comparisons.

****Kellman P.J., et al. (2014). Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates through Understanding and Predicting Difficulty. *PLoS ONE*, Vol. 9, No. 5, pp. 1-14 (Same as #78).**

Latent fingerprint examination is a complex task that, despite advances in image processing, still fundamentally depends on the visual judgments of highly trained human examiners. Fingerprints collected from crime scenes typically contain less information than fingerprints collected under controlled conditions. Specifically, they are often noisy and distorted and may contain only a portion of the total fingerprint area. Expertise in fingerprint comparison, like other forms of perceptual expertise, such as face recognition or aircraft identification, depends on perceptual learning processes that lead to the discovery of features and relations that matter in comparing prints. Relatively little is known about the perceptual processes involved in making comparisons, and even less is known about what characteristics of fingerprint pairs make particular comparisons easy or difficult. We measured expert examiner performance and judgments of difficulty and confidence on a new fingerprint database. We developed a

number of quantitative measures of image characteristics and used multiple regression techniques to discover objective predictors of error as well as perceived difficulty and confidence. A number of useful predictors emerged, and these included variables related to image quality metrics, such as intensity and contrast information, as well as measures of information quantity, such as the total fingerprint area. Also included were configural features that fingerprint experts have noted, such as the presence and clarity of global features and fingerprint ridges. Within the constraints of the overall low error rates of experts, a regression model incorporating the derived predictors demonstrated reasonable success in predicting objective difficulty for print pairs, as shown both in goodness of fit measures to the original data set and in a cross validation test. The results indicate the plausibility of using objective image metrics to predict expert performance and subjective assessment of difficulty in fingerprint comparisons.

55. Lee H.C. and Gaensslen R.E. (2001). Methods of Latent Fingerprint Development, in *Advances in Fingerprint Technology*, 2nd ed. HC Lee and RE Gaensslen, Eds. Boca Raton: CRC Press, 2001, pp. 105-107.

The authors discuss, in chapter four, the various physical, chemical, and illumination methods that are used for the detection of latent fingerprints. They explain which fingerprint detection method can be used on which type of surface to obtain the best results. The quality of the obtained prints is thus not directly dependent on the print development methods, because those methods are chosen to obtain the best possible results. The quality of the detected prints is thus rather dependent on the latent fingerprint itself, on the type of surface, and on the altering of the latent fingerprint due to environmental factors.

56. Li Z., Han Z. and Fu B. (2009). A Novel Method for the Fingerprint Image Quality Evaluation, in *Computational Intelligence and Software Engineering*. International Conference on CISE, pp. 1-4.

The performance of an automatic fingerprint identification system relies heavily on the quality of the captured fingerprint images. A novel method for fingerprint image quality analysis has been presented, which overcomes the shortcomings of most of the existing methods, considering the correlation of each quality feature as linear and paying no attention to the clarity of local texture. In this paper, ten features are extracted from the fingerprint image and then Fuzzy Relation Classifier is trained to classify the fingerprint images, which includes the unsupervised clustering and supervised classification to care more about the revelation of the data structure than other classifiers. Experimental results show that the proposed method has a good performance in evaluating the quality of the fingerprint images.

****Maceo A.V. (2009). Qualitative Assessment of Skin Deformation: A Pilot Study. *Journal of Forensic Identification*, Vol. 59, No. 4, pp. 390-440.**

Friction ridge skin deforms each time it contacts a surface. The primary factors determining the limits of skin deformation under applied stress to a fixed surface are the elastic nature of the friction ridge skin and the structure of the hand or foot area contacting the surface. This pilot study explored the flexibility of the distal phalanx of two index fingers of a single donor when compressive stress (deposition pressure) and tangential stresses (vertical sheering stress, horizontal sheering stress, and torque) were applied to a smooth, flat surface. The flexibility of the skin was found to be dependent upon the amount of compressive stress applied, the direction of tangential stress, and ridge flows in the fingerprint pattern. In addition to exploring the limits of skin flexibility, the effects of these different stresses were studied in latent prints generated under these conditions. The latent prints displayed robust clues that permit interpretation of the skin deformation by properly trained specialists.

***57. McMullen L. and Beaudoin A. (2013). Application of Oil Red O Following DFO and Ninhydrin Sequential Treatment: Enhancing Latent Fingerprints on Dry, Porous Surfaces. *Journal of Forensic Identification*, Vol. 63, No. 4, pp.387-423.**

The recovery of fingerprints from porous surfaces is often problematic, because fingerprints cannot usually be directly lifted from such objects. Also, the fingerprints are often not visible to the naked eye. 1,8-Diazafluoren-9-one (DFO) and ninhydrin (NIN) are amino acid-specific chemicals and are widely used to visualize latent prints on such surfaces. When these two fingerprint reagents are used consecutively, more fingerprints are able to be identified. Because Oil Red O (ORO) targets lipids, the strategy is to add this reagent to the sequence DFO → NIN to enhance previously undetected latent prints on dry, porous surfaces (e.g., paper). Targeting lipids can be a valuable asset to enhance prints that contain fewer amino acids or prints that have been exposed to a humid environment. In this study, an assessment of the usefulness of ORO in the DFO → NIN sequence for dry, porous surface was conducted. The usefulness of the addition of ORO in the sequence was assessed based on its sensitivity as well as the contrast, the quality of the recovered fingerprints, and the ability of ORO to produce additional fingerprints on various paper matrices. This research demonstrated that (1) the pretreatment of evidence with DFO → NIN did influence the ORO result, but did not prevent development of useful fingerprints with the sequential process, (2) the ORO sequential treatment did present lower contrast than ORO alone, but this lower contrast did not limit the ability of the fingerprint examiner to use the print, and finally, (3) the addition of ORO following the DFO → NIN sequence enhanced fingerprints already developed with those two amino acid reagents and even developed previously undetected fingerprints. This research supports using ORO in laboratories to visualize or even locate previously undetected prints on dry, porous surfaces.

58. Modi S.K., Elliott S.J., Whetsone J. et al. (2007). Impact of Age Groups on Fingerprint Recognition Performance, in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 19-23.

Ever since introduction of automated fingerprint recognition in law enforcement in the 1970s it has been used in applications ranging from personal authentication to civilian border control. The increasing use of automated fingerprint recognition raises a challenge of processing a diverse range of fingerprints. The quality control module is important to this process because it supports consistent fingerprint detail extraction, which helps in identification / verification. Inherent feature issues, such as poor ridge flow, and interaction issues, such as inconsistent finger placement, have an impact on captured fingerprint quality, which eventually affects overall system performance. Aging results in loss of collagen; compared to younger skin, aging skin is loose and dry. Decreased skin firmness directly affects the quality of fingerprints acquired by sensors. Medical conditions, such as arthritis, may affect the user's ability to interact with the sensor, further reducing fingerprint quality. Because quality of fingerprints varies according to the user population's ages, and fingerprint quality has an impact on overall system performance, it is important to understand the significance of fingerprint samples from different age groups. This research examines the effects of fingerprints from different age groups on quality levels, minutiae count, and performance of a minutiae-based matcher. The results show a difference in fingerprint image quality across age groups, most pronounced in the 62-and-older age group, confirming the work of [7].

59. Popa G., Potorac R. and Preda N. (2010). Method for Fingerprints Age Determination. *Romanian Journal of Legal Medicine*, Vol. 18, No. 2, pp. 149-154.

The aim of this study was to establish an estimation relationship of the age of fingerprints left on surfaces, by morphological, structural, and macro- and microscopic examinations, together with biochemical and titration DNA tests in order to confirm the rate of biological degradation during a certain period. The capacity of counting the age of a fingerprint leads to the possibility of placing it in time and to correlate it with the time of commission of the criminal act, bringing us information about the presence of a person in a certain place and period. As research methods, we used forensic techniques for fingerprints, as well as cytology and molecular biological methods (DNA analysis, DNA quantification with TaqMan using Real Time PCR). The estimation of the age of fingerprints using these methods offers us the advantages of standardization based on relationships between morphological or/and

biochemical characteristics depending on time, as well as the possibility to assign a rough guide to an individual's blood type.

60. Reis G. (2007). *Photoshop® CS3 for Forensics Professionals*. Indianapolis, USA: Wiley Publishing, Inc.

(Summary from bibliography***) George Reis presents in the first chapter of his book the principles and fundamentals of image processing. These principles ensure that images are processed according to the rules of evidence and avoid the risk of bias in image processing. In chapter five of the book, the author presents a list of valid image enhancement methods, which do not risk bias in the results. Finally, in the sixth chapter, he describes how the metadata of the image can be used and how to keep track of all the image enhancement methods that were applied.

***61. Schwarz L. and Hermanowski M.L. (2012). The Effect of Humidity on Long-Term Storage of Evidence Prior to Using Cyanoacrylate Fuming for the Detection of Latent Fingerprints. *Journal of Forensic Identification*, Vol. 62, No. 3, pp. 227-233.**

In Germany, cyanoacrylate fuming is the most popular method used for detecting latent fingerprints on nonporous surfaces. Many articles have been written about cyanoacrylate and fingerprint detection, but it is difficult to find information about the influence of relative humidity on the quality of developed prints while storing items until fuming. The influence of humidity (30%, 54% and 80%) while storing items at room temperature for a period of up to six months before fuming was tested. The results indicate that the influence measured is negligible.

****Sheets H.D., et al. (2014). Distortion in Fingerprints: A Statistical Investigation Using Shape Measurement Tools. *Journal of Forensic Sciences*, Vol. 59, No. 4, pp. 1113-1120.**

Friction ridge impression appearance can be affected due to the type of surface touched and pressure exerted during deposition. Understanding the magnitude of alterations, regions affected, and systematic/detectable changes occurring would provide useful information. Geometric morphometric techniques were used to statistically characterize these changes. One hundred and fourteen prints were obtained from a single volunteer and impressed with heavy, normal, and light pressure on computer paper, soft gloss paper, 10-print card stock, and retabs. Six hundred prints from 10 volunteers were rolled with heavy, normal, and light pressure on soft gloss paper and 10-print card stock. Results indicate that while different substrates/pressure levels produced small systematic changes in fingerprints, the changes were small in magnitude: roughly the width of one ridge. There were no detectable changes in the degree of random variability of prints associated with either pressure or substrate. In conclusion, the prints transferred reliably regardless of pressure or substrate

****Ulery B.T., et al. (2013). Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations. *Forensic Science International*, Vol. 230, pp. 99-106.**

A latent print examiner's assessment of the value, or suitability, of a latent impression is the process of determining whether the impression has sufficient information to make a comparison. A "no value" determination preemptively states that no individualization or exclusion determination could be made using the impression, regardless of quality of the comparison prints. Factors contributing to a value determination include clarity and the types, quantity, and relationships of features. These assessments are made subjectively by individual examiners and may vary among examiners. We modeled the relationships between value determinations and feature annotations made by 21 certified latent print examiners on 1850 latent impressions. Minutia count was strongly associated with value determinations. None of the models resulted in a stronger intraexaminer association with "value for individualization" determinations than minutia count alone. The association between examiner annotation and value determinations is greatly limited by the lack of reproducibility of both annotation and value determinations.

****Yoon S., et al. (2013). LFIQ: Latent Fingerprint Image Quality. *Biometrics: Theory, Applications and Systems (BTAS)*, IEEE Sixth International Conference on Biometrics Compendium, pp. 1-8.**

Latent fingerprint images are typically obtained under non-ideal acquisition conditions, resulting in incomplete or distorted impression of a finger, and ridge structure corrupted by background noise. This necessitates involving latent experts in latent fingerprint examination, including assessing the value of a latent print as forensic evidence. However, it is now generally agreed that human factors (e.g., human visual perception, expertise of latent examiners, workload, etc.) can significantly affect the reliability and consistency of the value determinations made by latent examiners. We propose an objective quality measure for latent fingerprints, called Latent Fingerprint Image Quality (LFIQ), that can be effectively used to distinguish latent fingerprints of good quality, which do not require any human intervention, and to compensate for the subjective nature of value determination by latent examiners. We investigate several factors that determine the latent quality: (i) ridge quality based on ridge clarity and connectivity of good ridge structures, (ii) minutiae reliability based on a minutiae dictionary learnt from high quality minutia patches, and (iii) position of the finger by detecting a reference point. The proposed LFIQ metric is based on triangulation of minutiae incorporating the above three factors. Experimental results show that (i) the proposed LFIQ is a good predictor of the latent matching performance by AFIS and (ii) it is also correlated with value determination by latent examiners.

FINGERPRINT MATCHING

VI. What scientific literature describes and justifies the role of automatic software algorithms and other mathematical/data-based models in the fingerprint matching process? (What is the ability of algorithms to operate as a stand-alone tool? What is the effect of applying algorithms to aid a human analyst? What is the confidence in match results with automatic algorithms?)

****Abraham J., Champod C., Lennard C. and Roux C. (2013). Spatial Analysis of Corresponding Fingerprint Features from Match and Close Non-Match Populations. *Forensic Science International*, Vol. 230, pp. 87-98.**

The development of statistical models for forensic fingerprint identification purposes has been the subject of increasing research attention in recent years. This can be partly seen as a response to a number of commentators who claim that the scientific basis for fingerprint identification has not been adequately demonstrated. In addition, key forensic identification bodies such as ENFSI [1] and IAI [2] have recently endorsed and acknowledged the potential benefits of using statistical models as an important tool in support of the fingerprint identification process within the ACE-V framework. In this paper, we introduce a new Likelihood Ratio (LR) model based on Support Vector Machines (SVMs) trained with features discovered via morphometric and spatial analyses of corresponding minutiae configurations for both match and close non-match populations often found in AFIS candidate lists. Computed LR values are derived from a probabilistic framework based on SVMs that discover the intrinsic spatial differences of match and close non-match populations. Lastly, experimentation performed on a set of over 120,000 publicly available fingerprint images (mostly sourced from the National Institute of Standards and Technology (NIST) datasets) and a distortion set of approximately 40,000 images, is presented, illustrating that the proposed LR model is reliably guiding towards the right proposition in the identification assessment of match and close non-match populations. Results further indicate that the proposed model is a promising tool for fingerprint practitioners to use for analyzing the spatial consistency of corresponding minutiae configurations.

****Abraham J., Champod C., Lennard C. and Roux C. (2013). Modern Statistical Models for Forensic Fingerprint Examinations: A Critical Review. *Forensic Science International*, Vol. 232, pp. 131-150.**

Over the last decade, the development of statistical models in support of forensic fingerprint identification has been the subject of increasing research attention, spurned on recently by commentators who claim that the scientific basis for fingerprint identification has not been adequately demonstrated. Such models are increasingly seen as useful tools in support of the fingerprint identification process within or in addition to the ACE-V framework. This paper provides a critical review of recent statistical models from both a practical and theoretical perspective. This includes analysis of models of two different methodologies: Probability of Random Correspondence (PRC) models that focus on calculating probabilities of the occurrence of fingerprint configurations for a given population, and Likelihood Ratio (LR) models which use analysis of corresponding features of fingerprints to derive a likelihood value representing the evidential weighting for a potential source.

****Alberink I., de Jongh A. and Rodriguez C. (2014). Fingermark Evidence Evaluation Based on Automated Fingerprint Identification System Matching Scores: The Effect of Different Types of Conditioning on Likelihood Ratios. *Journal of Forensic Sciences*, Vol. 59, No. 1, pp. 70-81.**

In recent studies, the evidential value of the similarity of minutiae configurations of fingermarks and fingerprints, for example expressed by automated fingerprint identification systems (AFIS), is determined by likelihood ratios (LRs). The paper explores whether there is an effect on LRs if conditioning takes place on specified fingers, fingerprints, or fingermarks under competing hypotheses: In addition, an approach is explored where conditioning is asymmetric. Comparisons between fingerprints and simulated fingermarks with eight minutiae are performed to produce similarity score distributions for each type of conditioning, given a fixed AFIS matching algorithm. Both similarity scores and LRs are significantly different if the conditioning changes. Given a common-source scenario, "LRs" resulting from asymmetric conditioning are on average higher. The difference may reach a factor of 2000. As conditioning on a suspect's finger(print) is labor-intensive and requires a cooperating suspect, it is recommended to just condition on the number of minutiae in the fingermark.

***62. Anthonioz N.M.E and Champod C. (2014). Evidence Evaluation in Fingerprint Comparison and Automated Fingerprint Identification Systems- Modeling Between Finger Variability. *Forensic Science International*, Vol. 235, No. 1, pp. 86-101.**

In the context of the investigation of the use of automated fingerprint identification systems (AFIS) for the evaluation of fingerprint evidence, the current study presents investigations into the variability of scores from an AFIS system when fingermarks from a known donor are compared to fingerprints that are not from the same source. The ultimate goal is to propose a model, based on likelihood ratios, that allows the evaluation of mark-to-print comparisons. In particular, this model, through its use of AFIS technology, benefits from the possibility of using a large amount of data, as well as from an already built-in proximity measure, the AFIS score. More precisely, the numerator of the LR is obtained from scores issued from comparisons between impressions from the same source and showing the same minutia configuration. The denominator of the LR is obtained by extracting scores from comparisons of the questioned mark with a database of non-matching sources. This paper focuses solely on the assignment of the denominator of the LR. We refer to it by the generic term of between-finger variability. The issues addressed in this paper in relation to between-finger variability are the required sample size, the influence of the finger number and general pattern, as well as that of the number of minutiae included and their configuration on a given finger. Results show that reliable estimation of between-finger variability is feasible with 10,000 scores. These scores should come from the appropriate finger number/ general pattern combination as defined by the mark. Furthermore, strategies of obtaining between-finger variability when these elements cannot be conclusively seen on the mark (and its position with respect to other marks for finger number) have been presented. These results immediately allow case-by-case estimation of the between-finger variability in an operational setting.

****Arora S.S., et al. (2014). Latent Fingerprint Matching: Performance Gain via Feedback from Exemplar Prints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 12, pp. 2452-2465.**

Latent fingerprints serve as an important source of forensic evidence in a court of law. Automatic matching of latent fingerprints to rolled/plain (exemplar) fingerprints with high accuracy is quite vital for such applications. However, latent impressions are typically of poor quality with complex background noise which makes feature extraction and matching of latents a significantly challenging problem. We propose incorporating top-down information or feedback from an exemplar to refine the features extracted from a latent for improving latent matching accuracy. The refined latent features (e.g. ridge orientation and frequency), after feedback, are used to re-match the latent to the top K candidate exemplars returned by the baseline matcher and resort the candidate list. The contributions of this research include: (i) devising systemic ways to use information in exemplars for latent feature refinement, (ii) developing a feedback paradigm which can be wrapped around any latent matcher for improving its matching performance, and (iii) determining when feedback is actually necessary to improve latent matching accuracy. Experimental results show that integrating the proposed feedback paradigm with a state-of-the-art latent matcher improves its identification accuracy by 0.5-3.5 percent for NIST SD27 and WVU latent databases against a background database of 100k exemplars.

63. Bolle R.M., Connell J.H., Pankanti S., et al. (2004). *Guide to Biometrics*. New York: Springer-Verlag.

(Summary from bibliography***) This book presents the main aspects related to biometric systems and recognition. Two chapters are dedicated to system errors. The two principal types of errors performed in biometric systems (called false acceptance and false rejection) are presented and discussed in the context both of verification and identification purposes. Since these two types of errors are highly correlated, the choice of the operating point is crucial and application related. This specificity is discussed in this book and various ways of presenting error probabilities are presented.

****Busey T., Silapiruti A. and Vanderkolk J. (2014). *The Relation Between Sensitivity, Similar Non-Matches and Database Size in Fingerprint Database Searches*. *Law, Probability and Risk*, Vol. 13, pp. 151-168.**

Searching against larger Automated Fingerprint Identification System (AFIS) databases may increase the likelihood of finding a suspect in the database. However, Dror and Mnookin (2010) have argued that this also leads to an increase in the number of similar non-matching prints, which could lead to an erroneous identification. Using simulations, we explore the relation between database size and two outcome factors: close non-matching prints and overall database sensitivity, which is a measure of discriminability between true matches and close non-matches. We find that larger databases tend to increase both the likelihood of finding the suspect in the database as well as the number of close nonmatching prints. However, the former tends to asymptote while the latter increases without bound and this leads to an initial increase and then a decrease in the sensitivity of the database as more prints are added. This suggests the existence of an optimal database size, and that caution should be observed when interpreting results from larger databases. Quantitative evidentiary techniques such as likelihood ratios have the potential to address some of these concerns, although they too must consider the database size when calculating the likelihood ratio. Implications for practitioners are discussed.

****Cao K., Liu E. and Jain A.K. (2014). *Segmentation and Enhancement of Latent Fingerprints: A Coarse to Fine Ridge Structure Dictionary*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 9, pp. 1847-1859.**

Latent fingerprint matching has played a critical role in identifying suspects and criminals. However, compared to rolled and plain fingerprint matching, latent identification accuracy is significantly lower due to complex background noise, poor ridge quality and overlapping structured noise in latent images. Accordingly, manual markup of various features (e.g., region of interest, singular points and minutiae) is typically necessary to extract reliable features from latents. To reduce this markup cost and to improve the consistency in feature markup, fully automatic and highly accurate ("lights-out" capability) latent matching algorithms are needed. In this paper, a dictionary-based approach is proposed for automatic latent segmentation and enhancement towards the goal of achieving "lights-out" latent identification systems. Given a latent fingerprint image, a total variation (TV)

decomposition model with L1 fidelity regularization is used to remove piecewise-smooth background noise. The texture component image obtained from the decomposition of latent image is divided into overlapping patches. Ridge structure dictionary, which is learnt from a set of high quality ridge patches, is then used to restore ridge structure in these latent patches. The ridge quality of a patch, which is used for latent segmentation, is defined as the structural similarity between the patch and its reconstruction. Orientation and frequency fields, which are used for latent enhancement, are then extracted from the reconstructed patch. To balance robustness and accuracy, a coarse to fine strategy is proposed. Experimental results on two latent fingerprint databases (i.e., NIST SD27 and WVUDB) show that the proposed algorithm outperforms the state-of-the-art segmentation and enhancement algorithms and boosts the performance of a state-of-the-art commercial latent matcher.

64. Dvornychenko V.N. and Garris M.D. (2006). Summary of NIST Latent Fingerprint Testing Workshop. US Department of Commerce, National Institute of Standards and Technology, NISTIR 7377.

The central focus of this paper is on the Latent Testing Workshop held on the National Institute of Standards and Technology (NIST) Campus April 5 to 6, 2006 and the lessons that were learned from it. The primary goal of the workshop was to gather information for the creation of a “Latent Challenge,” whose purpose is to stimulate Latent Automated Fingerprint Identification System (AFIS) vendors to submit their AFIS prototype systems for evaluation. To lay the foundations for such testing, it is necessary that 1) suitable test sets be identified and prepared; 2) the Application Programming Interface (API) is defined; and 3) the effective methods of performance scoring be defined. To provide background and context, past and present latent fingerprint activity at NIST is discussed. While the primary focus is on latent fingerprints, the paper also surveys relevant general biometrics activity.

***65. Guan H., Dienstfrey A., Theofanos M. et al. (2014). A Measurement Metric for Forensic Latent Fingerprint PreProcessing. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 8017.**

Although fingerprint mark-up and identification are well-studied fields, forensic fingerprint image preprocessing is still a relatively new domain in need of further scientific study and development of best practice guidance. Latent fingerprint image preprocessing is a common step in the forensic analysis workflow that is performed to improve image quality for subsequent identification analysis while simultaneously ensuring data integrity. Due to the low quality of the latent fingerprint images, preprocessing is especially crucial to the success of the final fingerprint identification in the forensic fingerprint image examination. In this report, we isolate the forensic fingerprint image preprocessing step for more detailed analysis. First we provide a brief review of latent fingerprint image preprocessing. We then turn to the problem of defining an image-based quality metric suitable for analysis of forensic latent fingerprint preprocessing. More precisely, we propose to extend Spectral Image Validation and Verification (SIVV) [1] to serve as a metric for latent fingerprint image quality measurement. SIVV analysis was originally developed to differentiate ten-print or rolled fingerprint images from other non-fingerprint images such as face or iris images. Several modifications are required to extend SIVV analysis to the latent space. We implement and test this new SIVV-based metric for latent fingerprint image quality and use it to measure the performance of the forensic latent fingerprint preprocessing step. Preliminary results show that the new metric can provide positive indications of both latent fingerprint image quality and the performance of the fingerprint preprocessing.

66. Indovina M., et al. (2011). ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #1]. U.S. Department of Commerce, National Institute of Standards and Technology, Washington DC, NISTIR 7775.

The National Institute of Standards and Technology (NIST) Evaluation of Latent Fingerprint Technologies - Extended Feature Sets (ELFT-EFS) consists of multiple ongoing latent algorithm evaluations. This report describes the design, process, results, and conclusions of ELFT-EFS Evaluation #1; an accuracy test of latent fingerprint searches using features marked by experienced human latent fingerprint examiners, in addition to automatic

feature extraction and matching (AFEM). There has never previously been an evaluation of latent fingerprint matchers of this scale in which systems from different vendors used a common, standardized feature set. The results show that searches using images plus manually marked Extended Features (EFS) demonstrated effectiveness as an interoperable feature set. The four most accurate matchers demonstrated a benefit from manually marked features when provided along with the latent image. The latent image itself was shown to be the single most effective search component for improving accuracy and was superior to features alone in most cases. For most matchers, the addition of new EFS features provided an improvement in accuracy. In several cases, some of the algorithms provided counterintuitive results that may be indicative of implementation issues; therefore, these results are preliminary, and broad conclusions on the efficacy of these features in improving performance should await the subsequent results from Evaluation #2, in which some known software issues are being corrected by the participants. The accuracy when searching with EFS features is promising, considering the results are derived from early-development, first-generation matchers. Further studies using next-generation matchers are warranted (and underway) to determine the performance gains possible with EFS.

****Indovina M., et al. (2012). ELFT- EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #2]. U.S. Department of Commerce, National Institute of Standards and Technology, Washington DC, NISTIR 7859.**

The National Institute of Standards and Technology (NIST) Evaluation of Latent Fingerprint Technologies – Extended Feature Sets (ELFT-EFS) consists of multiple ongoing latent algorithm evaluations. This report describes the results and conclusions of ELFT-EFS Evaluation #2; an accuracy test of latent fingerprint searches using features marked by experienced human latent fingerprint examiners, in addition to automatic feature extraction and matching (AFEM). ELFT-EFS Evaluation #1 was the first evaluation of latent fingerprint matchers in which systems from different vendors used a common, standardized feature set. ELFT-EFS Evaluation #2 repeats the same tests using updated matchers. The results show that in most cases there were measurable improvements in the ability of matchers to use images plus manually marked Extended Features (EFS) as an effective and interoperable feature set over the Evaluation #1 results.

The accuracy when searching with EFS features is promising considering the results are derived from early-development matchers. Further studies using next-generation matchers are warranted to determine the performance gains possible with EFS.

67. International Association for Identification, et al. (2011). Chapter 6: Automatic Fingerprint Identification Systems (AFIS), in *The Fingerprint Sourcebook*. NCJ 225320.

The benefits of the use of automated fingerprint identification systems (AFIS) as an essential tool for law enforcement agencies are presented. A complete description of the system's working principles is given, including AFIS functions and capabilities, system accuracy, and system and image requirements. The algorithm requirements for a successful data acquisition, data enrollment, feature extraction, and matching process are also presented. The difference in performance between experts and automated systems is also discussed in this chapter. The authors highlight that "automatic fingerprint-matching algorithms are significantly less accurate than a well-trained forensic expert," but even so they can significantly reduce the work for forensic experts.

****Kellman P.J., et al. (2014). Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates through Understanding and Predicting Difficulty. *PLoS ONE*, Vol. 9, No. 5, pp. 1-14.**

Latent fingerprint examination is a complex task that, despite advances in image processing, still fundamentally depends on the visual judgments of highly trained human examiners. Fingerprints collected from crime scenes typically contain less information than fingerprints collected under controlled conditions. Specifically, they are often noisy and distorted and may contain only a portion of the total fingerprint area. Expertise in fingerprint

comparison, like other forms of perceptual expertise, such as face recognition or aircraft identification, depends on perceptual learning processes that lead to the discovery of features and relations that matter in comparing prints. Relatively little is known about the perceptual processes involved in making comparisons, and even less is known about what characteristics of fingerprint pairs make particular comparisons easy or difficult. We measured expert examiner performance and judgments of difficulty and confidence on a new fingerprint database. We developed a number of quantitative measures of image characteristics and used multiple regression techniques to discover objective predictors of error as well as perceived difficulty and confidence. A number of useful predictors emerged, and these included variables related to image quality metrics, such as intensity and contrast information, as well as measures of information quantity, such as the total fingerprint area. Also included were configural features that fingerprint experts have noted, such as the presence and clarity of global features and fingerprint ridges. Within the constraints of the overall low error rates of experts, a regression model incorporating the derived predictors demonstrated reasonable success in predicting objective difficulty for print pairs, as shown both in goodness of fit measures to the original data set and in a cross validation test. The results indicate the plausibility of using objective image metrics to predict expert performance and subjective assessment of difficulty in fingerprint comparisons.

***68. Liu E., Jain A.K. and Tian J. (2013). A Coarse to Find Minutiae-Based Latent Palmprint Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 10, pp. 2307- 2322.**

With the availability of live-scan palmprint technology, high resolution palmprint recognition has started to receive significant attention in forensics and law enforcement. In forensic applications, latent palmprints provide critical evidence, as it is estimated that about 30 percent of the latents recovered at crime scenes are those of palms. Most of the available high-resolution palmprint matching algorithms essentially follow the minutiae-based fingerprint matching strategy. Considering the large number of minutiae (about 1,000 minutiae in a full palmprint compared to about 100 minutiae in a rolled fingerprint) and large area of foreground region in full palmprints, novel strategies need to be developed for efficient and robust latent palmprint matching. In this paper, a coarse to fine matching strategy based on minutiae clustering and minutiae match propagation is designed specifically for palmprint matching. To deal with the large number of minutiae, a local feature-based minutiae clustering algorithm is designed to cluster minutiae into several groups such that minutiae belonging to the same group have similar local characteristics. The coarse matching is then performed within each cluster to establish initial minutiae correspondences between two palmprints. Starting with each initial correspondence, a minutia match propagation algorithm searches for mated minutiae in the full palmprint. The proposed palmprint matching algorithm has been evaluated on a latent-to-full palmprint database consisting of 446 latents and 12,489 background full prints. The matching results show a rank-1 identification accuracy of 79.4 percent, which is significantly higher than the 60.8 percent identification accuracy of a state-of-the-art latent palmprint matching algorithm on the same latent database. The average computation time of our algorithm for a single latent-to-full match is about 141 ms for genuine match and 50 ms for impostor match, on a Windows XP desktop system with 2.2-GHz CPU and 1.00-GB RAM. The computation time of our algorithm is an order of magnitude faster than a previously published state-of-the-art-algorithm.

****Nagar A., Choi H. and Jain A.K. (2012). Evidential Value of Automated Latent Fingerprint Comparison: An Empirical Approach. *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 6, pp. 1752-1765.**

Latent prints are routinely recovered from crime scenes and are compared with available databases of known fingerprints for identifying criminals. However, current procedures to compare latent prints to large databases of exemplar (rolled or plain) prints are prone to errors. This suggests caution in making conclusions about a suspect's identity based on a latent fingerprint comparison. A number of attempts have been made to statistically model the utility of a fingerprint comparison in making a correct accept/reject decision or its evidential value. These approaches, however, either make unrealistic assumptions about the model or they lack simple interpretation. We argue that the posterior probability of two fingerprints belonging to different fingers given their match score, referred to as the nonmatch probability (NMP), effectively captures any implicating evidence of the comparison.

NMP is computed using state-of-the-art matchers and is easy to interpret. To incorporate the effect of image quality, number of minutiae, and size of the latent on NMP value, we compute the NMP vs. match score plots separately for image pairs (latent and exemplar prints) with different characteristics. Given the paucity of latent fingerprint databases in public domain, we simulate latent prints using two exemplar print databases (NIST SD-14 and Michigan State Police) by cropping regions of three different sizes. We appropriately validate this simulation using four latent databases (NIST SD-27 and three proprietary latent databases) and two state-of-the-art fingerprint matchers to compute their respective match scores. We also discuss a practical scenario where a latent examiner uses the proposed framework to compute the evidential value of a latent-exemplar print pair comparison.

****Neumann C., et al. (2007). Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiae. *Journal of Forensic Sciences*, Vol. 52, No. 1, pp. 54-64.**

Recent court challenges have highlighted the need for statistical research on fingerprint identification. This paper proposes a model for computing likelihood ratios (LRs) to assess the evidential value of comparisons with any number of minutiae. The model considers minutiae type, direction and relative spatial relationships. It expands on previous work on three minutiae by adopting a spatial modeling using radial triangulation and a probabilistic distortion model for assessing the numerator of the LR. The model has been tested on a sample of 686 ulnar loops and 204 arches. Features vectors used for statistical analysis have been obtained following a preprocessing step based on Gabor filtering and image processing to extract minutiae data. The metric used to assess similarity between two feature vectors is based on an Euclidean distance measure. Tippett plots and rates of misleading evidence have been used as performance indicators of the model. The model has shown encouraging behavior with low rates of misleading evidence and a LR power of the model increasing significantly with the number of minutiae. The LRs that it provides are highly indicative of identity of source on a significant proportion of cases, even when considering configurations with few minutiae. In contrast with previous research, the model, in addition to minutia type and direction, incorporates spatial relationships of minutiae without introducing probabilistic independence assumptions. The model also accounts for finger distortion.

****Neumann C., et al. (2013). Improving the Understanding and the Reliability of the Concept of “Sufficiency” in Friction Ridge Examination. NIJ Document #244231, Award #2010-DN-BX-K267.**

The purpose of this research project is to gather data informing on the robustness and transparency of fingerprint examination and to identify areas of improvement for preventing divergent decisions between two examiners considering the same latent print. The objective of this project is also to provide the fingerprint community with a body of research, tools and data allowing examiners to better understand the concept of *sufficiency*, in order to define better protocols for expressing and supporting the conclusions of fingerprint examinations. More specifically, the project has been designed to study the relationships between the observations made by examiners on pairs of latent/control prints and the decisions reached at the end of the different phases of the examination of those prints. A web-based system (called PiAnoS) has been used to capture the observations and the decisions made by a group of examiners on a set of paired latent/ control prints (section 5). The observations were summarized using different types of variables, some derived directly from the web-based system (section 6) and some assigned by a statistical model quantifying the weight of fingerprint evidence (section 7). A statistical analysis was conducted to measure the respective importance of the different variables in the decision-making process (sections 8 and 9). Finally, a series of recommendations were derived from our findings (section 10).

****Neumann C., et al. (2015). Quantifying the Weight of Fingerprint Evidence through the Spatial Relationship, Directions and Types of Minutiae Observed on Fingermarks. *Forensic Science International*, Vol. 248, pp. 154-171.**

This paper presents a statistical model for the quantification of the weight of fingerprint evidence. Contrarily to previous models (generative and score-based models), our model proposes to estimate the probability distributions of spatial relationships, directions and types of minutiae observed on fingerprints for any given

fingermark. Our model is relying on an AFIS algorithm provided by 3M Cogent and on a dataset of more than 4,000,000 fingerprints to represent a sample from a relevant population of potential sources. The performance of our model was tested using several hundreds of minutiae configurations observed on a set of 565 fingermarks. In particular, the effects of various sub-populations of fingers (i.e., finger number, finger general pattern) on the expected evidential value of our test configurations were investigated. The performance of our model indicates that the spatial relationship between minutiae carries more evidential weight than their type or direction. Our results also indicate that the AFIS component of our model directly enables us to assign weight to fingerprint evidence without the need for the additional layer of complex statistical modeling involved by the estimation of the probability distributions of fingerprint features. In fact, it seems that the AFIS component is more sensitive to the sub-population effects than the other components of the model. Overall, the data generated during this research project contributes to support the idea that fingerprint evidence is a valuable forensic tool for the identification of individuals.

***69. Orandi S., Ko K., Wood S.S., Grantham J.D., et al. (2014). Examination of the Impact of Fingerprint Spatial Area Loss on Matcher Performance in Various Mobile Identification Scenarios. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 7950.**

Background: NIST conducted a study of the FBI Repository for Individuals of Special Concern (RISC) system using various gallery and Mobile ID [MOBID] acquisition profile combinations to examine performance characteristics of the various profiles in terms of matching effectiveness and throughput.

Results: The predominant RISC operational case of Mobile ID FAP10 (fingerprint acquisition profile 10) using the left and right index fingers is at a marked disadvantage in terms of matcher performance compared to the larger FAP20 and FAP30 cases using the same fingers. In terms of false non-identification rate (FNIR), FAP10 submissions fail to identify their target approximately twice as often as FAP20 or FAP30. FAP30 appears to be best/optimal and its performance is on a par with (if not slightly better) than the uncropped control case. False positive identification rates remain relatively flat across all the cases examined.

Conclusion: System false non-identification rates suffer a significant performance penalty in the typical operational case of FAP10 two index finger (2,7) capture. FNIR performance can be markedly improved by either adopting FAP20 or FAP30 capture. To preserve legacy FAP10 equipment, it may be possible to use four finger capture (i.e., 2,3,7,8) with the smaller FAP10 to help mitigate some of the large performance penalty incurred with FAP10 in two finger operation, but this may increase the risk of sequencing errors from additional fingers having to be captured. Using additional fingers or larger FAPs can also help mitigate performance penalties when a database of only-rolled or only-flat fingerprint impressions is used rather than one populated with complete FD-249 records.

70. Paulino A.A, Jain A.K. and Jianjiang F. (2010). Latent Fingerprint Matching: Fusion of Manually Marked and Derived Minutiae, in 23rd SIBGRAPI Conference on Graphics, Patterns and Images, pp. 63-70.

Matching unknown latent fingerprints lifted from crime scenes to full (rolled or plain) fingerprints in law enforcement databases is of critical importance for combating crime and fighting terrorism. Compared to good quality full fingerprints acquired using live-scan or inking methods during enrollment, latent fingerprints are often smudgy and blurred, capture only a small finger area, and have large nonlinear distortion. For this reason, features (minutiae and singular points) in latents are typically manually marked by trained latent examiners. However, this introduces an undesired interoperability problems between latent examiners and automatic fingerprint identification systems (AFIS); the features marked by examiners are not always compatible with those automatically extracted by AFIS, resulting in reduced matching accuracy. While the use of automatically extracted minutiae from latents can avoid interoperability problem, such minutiae tend to be very unreliable because of the poor quality of latents. In this paper, we improve latent to full fingerprint matching accuracy by combining manually marked (ground truth) minutiae with automatically extracted minutiae. Experimental results on a public domain database, NIST SD27, demonstrate the effectiveness of the proposed algorithm.

****Paulino A.A., Feng J. and Jain A.K. (2013). Latent Fingerprint Matching Using Descriptor-Based Hough Transform. *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 1, pp. 31-45.**

Identifying suspects based on impressions of fingers lifted from crime scenes (latent prints) is a routine procedure that is extremely important to forensics and law enforcement agencies. Latents are partial fingerprints that are usually smudgy, with small area and containing large distortion. Due to these characteristics, latents have a significantly smaller number of minutiae points compared to full (rolled or plain) fingerprints. The small number of minutiae and the noise characteristic of latents make it extremely difficult to automatically match latents to their mated full prints that are stored in law enforcement databases. Although a number of algorithms for matching full-to-full fingerprints have been published in the literature, they do not perform well on the latent to full matching problem. Further, they often rely on features that are not easy to extract from poor quality latents. In this paper, we propose a new fingerprint matching algorithm which is especially designed for matching latents. The proposed algorithm uses a robust alignment algorithm (descriptor-based Hough transform) to align fingerprints and measures similarity between fingerprints by considering both minutiae and orientation field information. To be consistent with the common practice in latent matching (i.e., only minutiae are marked by latent examiners), the orientation field is reconstructed from minutiae. Since the proposed algorithm relies only on manually marked minutiae, it can be easily used in law enforcement applications. Experimental results on two different latent databases (NIST SD27 and WVU latent databases) show that the proposed algorithm outperforms two well optimized commercial fingerprint matchers. Further, a fusion of the proposed algorithm and commercial fingerprint matchers leads to improved matching accuracy.

****Watson C., et al. (2014). Fingerprint Vendor Technology Evaluation. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 8034.**

FpVTE was conducted primarily to assess the current capabilities of fingerprint matching algorithms using operational datasets containing several million subjects. There were three classes of participation that examined one-to-many identification using various finger combinations from single finger up to ten fingers. Class A used single-index finger capture data and evaluated single index finger (right or left) and two index finger (right and left) identification. Class B used identification flat (IDFlat) captures (4-4-2; left slap, right slap, and two thumbs simultaneously) and evaluated ten-finger, eight-finger (right and left slap), and four-finger (right or left slap) identification. Class C used rolled and plain impression (4-4-1-1; left slap, right slap, left thumb, and right thumb) captures and evaluated ten-finger rolled-to-rolled, ten-finger plain-to-plain, and ten-finger plain-to-rolled identification. Enrollment sets used for one-to-many identification varied in size from 5 000 up to 5 000 000 enrolled subjects. Any segmentation of four-finger slap images or two-thumb captures was performed by the submitted software. All data used was sequestered operational data that was not shared with any of the participants. The evaluation allowed each participant to make two submissions per class (A, B, and C) of participation over three rounds. After each of the first two rounds of submissions, feedback was provided to the participants and they were allowed to evaluate their performance, make adjustments to their submissions, and resubmit for the next round. The results of the third and final round of submissions are reported in this document. The evaluation was conducted at the National Institute of Standards and Technology (NIST) using commodity NIST owned hardware. Participant submissions were compliant to the testing Application Programming Interface (API), which were linked to a NIST-developed test driver and run by NIST employees. All submissions went through validation testing to ensure that results generated on NIST's hardware matched results participants generated on their own hardware. This was the first large-scale one-to-many fingerprint evaluation since FpVTE 2003. In 2003, participants brought their own hardware to NIST to process the evaluation data. The datasets in 2003 had approximately 25 000 subjects and required millions of single subject-to-subject matches. The current FpVTE used a testing model closer to real one-to-many identification systems by allowing the submitted software to control how it does the one-to-many search and return a candidate list of potential matches. The number of subjects used was also significantly higher, as the current FpVTE had ≈ 10 million subjects in the testing datasets. The results in this report are based on 30 000 (10 000 mates and 20 000 nonmates) search subjects. There will be an additional report with results (lower errors rates) using 350 000 (50 000 mates and 300 000 nonmates) search subjects

****Wilson C., et al. (2004). Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report. U.S. Department of Commerce, National Institute of Standards and Technology, NISTIR 7123.**

The Fingerprint Vendor Technology Evaluation (FpVTE) 2003 was conducted to evaluate the accuracy of fingerprint matching, identification, and verification systems. The FpVTE is one of the tests that NIST has conducted in order to fulfill part of its PATRIOT Act mandate. Additional evaluations include the testing of the FBI IAFIS system, the US-VISIT IDENT system and SDKs (Software Development Kits) from several vendors. Eighteen different companies competed in FpVTE, and 34 systems were evaluated. Different subtests measured accuracy for various numbers and types of fingerprints, using operational fingerprint data from a variety of U.S. Government sources. The most accurate systems were found to have consistently very low error rates across a variety of data sets. The variables that had the clearest effect on system accuracy were the number of fingers used and fingerprint quality. An increased number of fingers resulted in higher accuracy: the accuracy of searches using four or more fingers was better than the accuracy of two-finger searches, which was better than the accuracy of single-finger searches. The test also shows that the most accurate fingerprint systems are more accurate than the most accurate facial recognition systems, even when comparing the performance of operational quality single fingerprints to high-quality face images.

****Yang X., Feng J. and Zhou J. (2014). Localized Dictionaries Based Orientation Field Estimation for Latent Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 5, pp. 955-969.**

Dictionary based orientation field estimation approach has shown promising performance for latent fingerprints. In this paper, we seek to exploit stronger prior knowledge of fingerprints in order to further improve the performance. Realizing that ridge orientations at different locations of fingerprints have different characteristics, we propose a localized dictionaries-based orientation field estimation algorithm, in which noisy orientation patch at a location output by a local estimation approach is replaced by real orientation patch in the local dictionary at the same location. The precondition of applying localized dictionaries is that the pose of the latent fingerprint needs to be estimated. We propose a Hough transform-based fingerprint pose estimation algorithm, in which the predictions about fingerprint pose made by all orientation patches in the latent fingerprint are accumulated. Experimental results on challenging latent fingerprint datasets show the proposed method outperforms previous ones markedly.

VII. What scientific literature describes and justifies standards regarding formulation and reporting of “inclusion/identification, exclusion and inconclusive” (ACE-V) opinions?

****Expert Working Group on Human Factors in Latent Print Analysis. (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.**

Fingerprints have provided a valuable method of personal identification in forensic science and criminal investigations for more than 100 years. Fingerprints left at crime scenes generally are latent prints—unintentional reproductions of the arrangement of ridges on the skin made by the transfer of materials (such as amino acids, proteins, polypeptides, and salts) to a surface. Palms and the soles of feet also have friction ridge skin that can leave latent prints. The examination of a latent print consists of a series of steps involving a comparison of the latent print to a known (or exemplar) print. Courts have accepted latent print evidence for the past century. However, several high-profile cases in the United States and abroad have highlighted the fact that human errors can occur, and litigation and expressions of concern over the evidentiary reliability of latent print examinations and other forensic identification procedures has increased in the past decade. “Human factors” issues can arise in any experience- and judgment-based analytical process, such as latent print examination. Inadequate training,

extraneous knowledge about the suspects in the case or other matters, poor judgment, health problems, limitations of vision, complex technology, and stress are but a few factors that can contribute to errors. A lack of standards or quality control, poor management, insufficient resources, and substandard working conditions constitute other potentially contributing factors. In addition to reaching correct conclusions in the matching process, latent print examiners are expected to produce records of the examination and, in some cases, to present their conclusions and the reasoning behind them in the courtroom. Human factors issues related to the documentation and communication of an examiner's work and findings, therefore, merit attention as well. The study of human factors focuses on the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered at work and in daily living. Human factors analysis can advance our understanding of the nature of errors in complex work settings. Most preventable, adverse events are not just the result of isolated or idiosyncratic behavior but are in part caused by systemic factors. The forensic science community can benefit from the application of human factors research to enhance quality and productivity in friction ridge examinations and to reduce the likelihood and consequences of human error at various stages in the interpretation of evidence. To further this effort, the National Institute of Justice (NIJ) Office of Investigative and Forensic Sciences (OFIS) within the U.S. Department of Justice and the National Institute of Standards and Technology's (NIST's) Law Enforcement Standards Office (OLES) sponsored the work of this expert panel to examine human factors in latent print analysis and to develop recommendations to reduce the risk of error and improve the practice of latent print analysis.

****Langenburg G., Champod C. and Genessay T. (2012). Informing the Judgements of Fingerprint Analysts Using Quality Metric and Statistical Assessment Tools. *Forensic Science International*, Vol. 219, pp. 183-198.**

The aim of this research was to evaluate how fingerprint analysts would incorporate information from newly developed tools into their decision making processes. Specifically, we assessed effects using the following: (1) a quality tool to aid in the assessment of the clarity of the friction ridge details, (2) a statistical tool to provide likelihood ratios representing the strength of the corresponding features between compared fingerprints, and (3) consensus information from a group of trained fingerprint experts. The measured variables for the effect on examiner performance were the accuracy and reproducibility of the conclusions against the ground truth (including the impact on error rates) and the analyst accuracy and variation for feature selection and comparison. The results showed that participants using the consensus information from other fingerprint experts demonstrated more consistency and accuracy in minutiae selection. They also demonstrated higher accuracy, sensitivity, and specificity in the decisions reported. The quality tool also affected minutiae selection (which, in turn, had limited influence on the reported decisions); the statistical tool did not appear to influence the reported decisions.

****Neumann C., et al. (2013). Improving the Understanding and the Reliability of the Concept of "Sufficiency" in Friction Ridge Examination. NIJ Document #244231, Award #2010-DN-BX-K267.**

The purpose of this research project is to gather data informing on the robustness and transparency of fingerprint examination and to identify areas of improvement for preventing divergent decisions between two examiners considering the same latent print. The objective of this project is also to provide the fingerprint community with a body of research, tools and data allowing examiners to better understand the concept of *sufficiency*, in order to define better protocols for expressing and supporting the conclusions of fingerprint examinations. More specifically, the project has been designed to study the relationships between the observations made by examiners on pairs of latent/control prints and the decisions reached at the end of the different phases of the examination of those prints. A web-based system (called PiAnoS) has been used to capture the observations and the decisions made by a group of examiners on a set of paired latent/ control prints (section 5). The observations were summarized using different types of variables, some derived directly from the web-based system (section 6) and some assigned by a statistical model quantifying the weight of fingerprint evidence (section 7). A statistical analysis was conducted to measure the respective importance of the different variables in the decision-making process (sections 8 and 9). Finally, a series of recommendations were derived from our findings (section 10).

****Ray E. and Dechant P.J. (2013). Sufficiency and Standards for Exclusion Decisions. *Journal of Forensic Identification*, Vol. 63, No. 6, pp. 675-697.**

Current research into latent fingerprint examiner decisions shows that erroneous exclusions are common and inevitable. These errors may be dramatically reduced by establishing clear standards for exclusion decisions and providing comprehensive training on exclusions to all latent print examiners. The first step in this process is to standardize verification for all exclusion decisions. Second, examiners must be able to use the inconclusive decision when it is appropriate. The inconclusive decision should be reached whenever there is insufficient detail in agreement to identify and when there is insufficient detail in disagreement to exclude. This decision gives the examiner an option that reduces the chance of erroneously excluding a print that was not located. The latent print unit at the Arizona Department of Public Safety has set its standard for exclusion to be Level 1 and Level 2 detail in disagreement. In other words, two or more target groups of minutiae near an anchor point such as a delta or core must be in disagreement for an exclusion. When these features are used in conjunction, examiners can be confident that they found sufficient disagreement to warrant an exclusion decision and reduce the chance of an erroneous exclusion. The latent print community should continue the discussion on a standard for exclusion to reduce the unacceptably high error rate on this decision and to clarify the appropriate use of the exclusion decision.

***71. Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST). (2013). Standards for Examining Friction Ridge Impressions and Resulting Conclusions, Ver. 2.0.**

Friction ridge impression examinations are conducted by examiners using the Analysis, Comparison, Evaluation, and Verification (ACE-V) methodology, which includes both qualitative and quantitative aspects. ACE is not generally applied as a strictly linear process because it may include a return to any previous phase. Application of ACE includes observations, measurements, assessments, decision-making, and documentation, which are enabled by the education, training, skill, and experience of the examiner. The examination of friction ridge impressions and the resulting conclusions are based on ridge flow and ridge paths; the location, direction, and spatial relationships of minutiae; and ridge structure. The analysis phase leads to the determination of *suitability*. Following comparison, the evaluation phase leads to the following conclusions: *individualization, exclusion, or inconclusive*. These conclusions are based on the following premises:

- Friction ridge skin bears an extremely complex, unique, and persistent morphological structure.
- Notwithstanding the pliability of friction ridge skin, the contingencies of touching a surface and the nature of the matrix, an impression of friction ridge structure may be left following contact with a surface.
- This impression may display features of varying quality (clarity of ridge features) and specificity (weighted values and rarity).
- Notwithstanding variations in clarity and specificity, the unique aspects of friction ridge skin may be represented as highly discriminative features in impressions.
- An impression that contains sufficient quality and quantity of friction ridge features can be individualized to, or excluded from, a source.
- The use of a fixed number of friction ridge features as a threshold for the establishment of an individualization is not sufficiently supported

****Taylor S.J., et al. (2012). Application of Spatial Statistics to Latent Print Identifications: Towards Improved Forensic Science Methodologies. NIJ Document #240590, Award #2009-DN-BX-K228.**

In 2010 we initiated a research project to address criticisms raised in a 2009 National Academy of Sciences (NAS) report regarding the presumption of fingerprint uniqueness and the reliability of latent print identifications using the ACE-V methodology (National Research Council 2009). This project addresses the question of fingerprint uniqueness (i.e., the discriminating value of the various fingerprint ridgeline features) by statistically evaluating the spatial distribution of these features. The purpose of the project was to review the latent print ACE-V comparison methodology to ascertain the fingerprint features considered during the comparison process and apply principles

of spatial analyses to calculate false-match probabilities. The objectives were to spatially analyze fingerprint features (e.g., minutiae and ridge lines) using Geographic Information Systems (GIS) techniques and empirically derive probabilities to provide a quantitative measure of the discriminating value of the various ridgeline features. The resultant probabilities are applicable for subsequent qualification of latent print comparison conclusions.

Project methods included spatial pattern characterization using GIS, geometric morphometric (GM) analysis, and the calculation of false-match probabilities using Monte Carlo (MC) simulations. A data set of digitized fingerprints from the Oregon population was compiled and spatially analyzed utilizing GIS software to place minutiae and ridge line features in a common Cartesian coordinate system. The parameters of these fingerprint features, including minutiae location, direction and minutiae ridgeline configurations, were evaluated. Geometric morphometrics was used to study shape variation between and among fingerprint pattern types. GIS-based procedures were established for the selection of landmarks and semi-landmarks, the superimposition of fingerprint images, the visualization of shape change, the ordination of superimposition data, and the application of multivariate statistics. Using MC simulations, random-match probabilities were calculated to evaluate the spatial configurations of minutiae within and between pattern types to quantitatively evaluate the discriminating value of fingerprints features; that is, do two fingerprints or two regions of different fingerprints have the same spatial distribution of minutiae and ridgelines? MC simulations were performed using 3, 5, 7 and 9 minutiae with other minutiae attributes chosen for additional match criteria.

GIS results showed there was a greater density of minutiae and ridgelines below the core compared to above the core, regardless of pattern type. However, the distributions of bifurcations and ridge endings were more similar within any pattern type rather than among them. Also, pattern types with comparable ridge flow (e.g., right and left slant loops, and whorls and double loop whorls) had greater similarity between them when comparing various metrics such as axis dimensions and Thiessen polygon ratios. GM results demonstrated little shape variation among fingerprints of the same pattern type with the greatest shape variation associated with the deltas. Additional GM spatial analyses suggested a very high degree of shape consistency between left and right slant loops and between whorls and double loop whorls. MC simulations showed that the probability of random minutiae correspondence drastically decreased as the fingerprint attribute criteria (e.g., minutiae type and direction) increased. In addition, increasing the number of minutiae and fingerprint attributes applied in searches away from the core and delta regions yielded lower probabilities for a false match. However, results demonstrated that minutiae spatial distributions in regions around and below the core were not always unique.

Fingerprint characterization of ridgeline minutiae configurations and establishing random-match probabilities when using specified features quantitatively describe the discriminating value of these fingerprint ridgeline features. As such, random-match probabilities will allow the latent print examiner to qualify their comparison conclusions.

****Triplett M. and Cooney L. (2006). The Etiology of ACE-V and its Proper Use: An Exploration of the Relationship Between ACE-V and the Scientific Method of Hypothesis Testing. *Journal of Forensic Identification*, Vol. 56, No. 3, pp. 345-355.**

ACE-V is commonly described as the scientific methodology that fingerprint practitioners use to individualize friction skin impressions, including both ten-print and latent print examinations. This paper looks at the history of ACE-V, analyzes whether a clear understanding of ACE-V exists, gives a brief description of how ACE-V should be used, and looks at the repercussions of incorrectly using ACE-V. Recognizing the misconceptions about ACE-V is the first step in establishing a comprehensive grasp of this process, which in turn will result in practitioners reaching the best possible conclusions.

***72. Ulery B.T., Hicklin R.A., Kiebusinski G.I., et al. (2013). Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations. *Forensic Science International*, Vol. 230, No. 1, pp. 99-106.**

A latent print examiner's assessment of the value, or suitability, of a latent impression is the process of determining whether the impression has sufficient information to make a comparison. A "no value" determination preemptively states that no individualization or exclusion determination could be made using the impression, regardless of quality of the comparison prints. Factors contributing to a value determination include clarity and the types, quantity, and relationships of features. These assessments are made subjectively by individual examiners and may vary among examiners. We modeled the relationships between value determinations and feature annotations made by 21 certified latent print examiners on 1850 latent impressions. Minutia count was strongly associated with value determinations. None of the models resulted in a stronger intra-examiner association with "value for individualization" determinations than minutia count alone. The association between examiner annotation and value determinations is greatly limited by the lack of reproducibility of both annotation and value determinations.

VIII. How much precision/uncertainty is associated with probability estimates? (How much similarity is needed to obtain a high probability of match or confidence in mismatch? How are the three levels of detail accommodated in the analysis?)

73. Champod C. (2000). Fingerprints (Dactyloscopy): Standard of Proof, in *Encyclopedia of Forensic Sciences*. J Siegel, P Saukko and G Knupfer, Eds. London: Academic Press, 2000, pp. 884-890.

(Summary from bibliography***) This paper presents the two existing standards of proof, one using a predetermined minimum number of minutiae and one with no numerical standard. The author recommends the use of the latter, because the identification process is a global assessment that balances both quantitative and qualitative aspects, rather than being a concept reduced to counting fingerprint minutiae. This statement is supported by the 1995 Ne'Urim declaration, stating that "no scientific basis exists for requiring that a predetermined minimum number of friction ridge features must be present in two impressions in order to establish a positive identification." Moreover, arguments are given in favor of a probabilistic approach and the possibility of giving a qualified opinion based on dactyloscopic evidence in the decision-making process, rather than the use of positive identification. The presented way forward is the adoption of a scheme of total quality management dealing with various aspects of identification and training, laboratory procedures, and audits.

***Champod C. (2013). Friction Ridge Skin Impression Evidence – Standards of Proof, in *Encyclopedia of Forensic Sciences*, 2nd edition. J Siegel and P Saukko, Eds. London: Academic Press, pp. 111-116.**

Papillary surfaces are covered in fine lines, or 'ridges' arranged generally in patterns such as loops, whorls, and arches. The ridges form characteristics called minutiae, such as bifurcations, ridge endings (or combined minutiae), and also smaller features (pores, ridge edges, and structures). These detailed features are the consequence of a morphogenesis, so sensitive to outside influence that they are difficult to predict, unlike general patterns, in their position and their form. The variability is such that even monozygotic twins have distinguishable friction ridge skin in this respect. Following a comparison between a mark (recovered for example in association with a crime) and a print (controlled impression associated with a suspect), suppose that the examiner observes agreement – within allowed tolerances given the flexibility of the skin, the distortion during the deposition process, etc. – between the mark and the print without any significant discrepancy. The question is then often expressed as follows: 'how many similarities are required to identify?' or 'what is the required sufficient agreement to conclude an individualization?'. Both terms 'identification' and 'individualization' are used here and hereinafter as synonymous. The aim of this article is to try to address these questions of the 'standard of proof', by reviewing international views and practices. The perspective adopted here is essentially the one of the forensic practitioners dealing with friction ridge skin and not the perspective of the court or judiciary.

74. International Association for Identification, et al. (2011). Chapter 14: Scientific Research Supporting the Foundations of Friction Ridge Examinations, in *The Fingerprint Sourcebook*. NCJ 225320.

Based on Popper's criteria, the author presents friction ridge principles (i.e. uniqueness and permanence) as appropriate scientific theories and laws, and friction ridge examination is presented as hypothesis testing. He also provides a state of the art (through 2007) regarding scientific work carried out in order to answer questions related to uniqueness and permanence of friction ridge skin. Moreover, a review of available models to assess matching probabilities of a specific configuration of minutiae is provided. Finally, he shows the results of studies performed to test the ACE-V method used by fingerprint examiners to compare fingerprint impressions, such as the benefit of training or the absence of bias during the analysis step.

75. Interpol European Expert Group on Fingerprint Identification II- IEEGFI II. (2004). *Part 2: Detailing the Method Using Common Terminology and through the Definition and Application of Shared Principles*. Interpol, Lyon.

(No abstract in original document***) The report recognizes and endorses the basic principles of fingerprint identification. It contains detailed guidelines for analysis, comparison, evaluation, validation and verification of fingerprint details, introduces common terminology, identifies certain areas of risk, and advises upon the application of general scientific principles and methodology that can be readily translated to the field of fingerprint identification. This leads to an entirely transparent, rigorous, reproducible, and verifiable methodology and process. Intentionally, no numerical standard for conclusions is proposed. Instead, some criteria are proposed to help assess the strength associated with the findings.

76. Jain A.K. (2011). *Automatic Fingerprint Matching Using Extended Feature Set*. National Institute of Justice, 235577.

Fingerprint friction ridge features are generally described in a hierarchical order at three different levels, namely, Level 1 (ridge flow), Level 2 (minutiae points) and Level 3 (pores and ridge shape, etc.). Current Automated Fingerprint Identification Systems (AFIS) generally rely only on a subset of Level 1 and Level 2 features (minutiae and core/delta) for matching. On the other hand, latent print examiners frequently take advantage of a much richer set of features naturally occurring in fingerprints. It is believed that this difference may be one of the reasons for the superior performance of fingerprint examiners over AFIS, particularly in case of difficult latent matches. Fingerprint features, other than minutiae and core/delta, are also referred to as the extended feature set (EFS). The goal of this study is to i) develop algorithms for encoding and matching extended features, ii) develop fusion algorithms to combine extended features with minutiae information to improve fingerprint matching accuracy, and iii) understand the contributions of various extended features in latent fingerprint matching. We study a number of extended features at all three levels, including ridge flow map, ridge wavelength map, ridge quality map, ridge skeleton, pores, dots, incipient ridges, and ridge edge protrusions. Feature extraction and matching algorithms are developed for each type of feature. Relative contribution of each feature towards the overall matching accuracy is evaluated by incrementally adding features to baseline features (minutiae and core/delta). The order of adding features is determined based on the amount of manual labor in feature marking and the estimated importance of features. Latent fingerprint databases, NIST SD27 and ELFT-EFS-PC, and several NIST rolled/plain fingerprint databases are used in our experiments. Based on extensive experiments, we report the following findings: i) almost all the extended features lead to some improvement in latent matching accuracy, ii) extended features at higher level are more effective in improving latent matching accuracy than those at lower level, iii) high image resolution (at least 1000 ppi) is necessary but not sufficient for reliably capturing Level 3 features. Based on our study, we would like to offer the following recommendations: i) extended features at Level 1 and Level 2 are strongly recommended to be incorporated into AFIS, ii) convenient GUI tools should be developed to help fingerprint examiners manually mark extended features (especially ridge skeleton) at Level 1

and Level 2 in latents, and iii) it is crucial to improve the quality of enrolled fingerprints (so that a sufficient number of Level 3 features can be extracted) before Level 3 features can play an important role in AFIS.

77. Jain A.K. and Feng J. (2011). Latent Fingerprint Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, pp. 88-100.

Latent fingerprint identification is of critical importance to law enforcement agencies in identifying suspects. Latent fingerprints are inadvertent impressions left by fingers on surfaces of objects. While tremendous progress has been made in plain and rolled fingerprint matching, latent fingerprint matching continues to be a difficult problem. Poor quality of ridge impressions, small finger area, and large nonlinear distortion are the main difficulties in latent fingerprint matching compared to plain or rolled fingerprint matching. We propose a system for matching latent fingerprints found at crime scenes to rolled fingerprints enrolled in law enforcement databases. In addition to minutiae, we also use extended features, including singularity, ridge quality map, ridge flow map, ridge wavelength map, and skeleton. We tested our system by matching 258 latents in the NIST SD27 database against a background database of 29,257 rolled fingerprints obtained by combining the NIST SD4, SD14, and SD27 databases. The minutiae-based baseline rank-1 identification rate of 34.9 percent was improved to 74 percent when extended features were used. In order to evaluate the relative importance of each extended feature, these features were incrementally used in the order of their cost in marking by latent experts. The experimental results indicate that singularity, ridge quality map, and ridge flow map are the most effective features in improving the matching accuracy.

***78. Kellman P.J., Mnookin J.L., Erlikhman G., et al. (2014). Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates Through Understanding and Predicting Difficulty. *PLoS ONE*, Vol. 9, No.5, pp. 1-15, e94617.**

Latent fingerprint examination is a complex task that, despite advances in image processing, still fundamentally depends on the visual judgments of highly trained human examiners. Fingerprints collected from crime scenes typically contain less information than fingerprints collected under controlled conditions. Specifically, they are often noisy and distorted and may contain only a portion of the total fingerprint area. Expertise in fingerprint comparison, like other forms of perceptual expertise, such as face recognition or aircraft identification, depends on perceptual learning processes that lead to the discovery of features and relations that matter in comparing prints. Relatively little is known about the perceptual processes involved in making comparisons, and even less is known about what characteristics of fingerprint pairs make particular comparisons easy or difficult. We measured expert examiner performance and judgments of difficulty and confidence on a new fingerprint database. We developed a number of quantitative measures of image characteristics and used multiple regression techniques to discover objective predictors of error as well as perceived difficulty and confidence. A number of useful predictors emerged, and these included variables related to image quality metrics, such as intensity and contrast information, as well as measures of information quantity, such as the total fingerprint area. Also included were configural features that fingerprint experts have noted, such as the presence and clarity of global features and fingerprint ridges. Within the constraints of the overall low error rates of experts, a regression model incorporating the derived predictors demonstrated reasonable success in predicting objective difficulty for print pairs, as shown both in goodness of fit measures to the original data set and in a cross validation test. The results indicate the plausibility of using objective image metrics to predict expert performance and subjective assessment of difficulty in fingerprint comparisons.

****Neumann C., et al. (2013). Improving the Understanding and the Reliability of the Concept of “Sufficiency” in Friction Ridge Examination. NIJ Document #244231, Award #2010-DN-BX-K267.**

The purpose of this research project is to gather data informing on the robustness and transparency of fingerprint examination and to identify areas of improvement for preventing divergent decisions between two examiners considering the same latent print. The objective of this project is also to provide the fingerprint community with a body of research, tools and data allowing examiners to better understand the concept of *sufficiency*, in order to

define better protocols for expressing and supporting the conclusions of fingerprint examinations. More specifically, the project has been designed to study the relationships between the observations made by examiners on pairs of latent/control prints and the decisions reached at the end of the different phases of the examination of those prints. A web-based system (called PiAnoS) has been used to capture the observations and the decisions made by a group of examiners on a set of paired latent/ control prints (section 5). The observations were summarized using different types of variables, some derived directly from the web-based system (section 6) and some assigned by a statistical model quantifying the weight of fingerprint evidence (section 7). A statistical analysis was conducted to measure the respective importance of the different variables in the decision-making process (sections 8 and 9). Finally, a series of recommendations were derived from our findings (section 10).

79. Polski J., Smith R., Garrett R., et al. (2011). *The Report of the International Association for Identification, Standardization II Committee*. National Institute of Justice, 233980.

The main recommendation of the present report states that, “there currently exists no scientific basis for requiring a minimum amount of corresponding friction ridge detail information between two impressions to arrive at an opinion of single source attribution.” Another milestone in this report is the recommendation that allows the examiner to offer oral or written reports of testimony of probably or likely conclusions concerning source attribution of two friction ridge impressions being from the same source. Note the presence of appendix G referencing a bibliography of available research material.

****Ulery B.T., Hicklin R.A., Roberts M.A. and Buscaglia J. (2014). *Measuring What Latent Fingerprint Examiners Consider Sufficient Information for Individualization Determinations*. *PLoS ONE*, Vol. 9, No. 11, pp. 1-16.**

Latent print examiners use their expertise to determine whether the information present in a comparison of two fingerprints (or palmprints) is sufficient to conclude that the prints were from the same source (individualization). When fingerprint evidence is presented in court, it is the examiner’s determination—not an objective metric—that is presented. This study was designed to ascertain the factors that explain examiners’ determinations of sufficiency for individualization. Volunteer latent print examiners (n = 170) were each assigned 22 pairs of latent and exemplar prints for examination, and annotated features, correspondence of features, and clarity. The 320 image pairs were selected specifically to control clarity and quantity of features. The predominant factor differentiating annotations associated with individualization and inconclusive determinations is the count of corresponding minutiae; other factors such as clarity provided minimal additional discriminative value. Examiners’ counts of corresponding minutiae were strongly associated with their own determinations; however, due to substantial variation of both annotations and determinations among examiners, one examiner’s annotation and determination on a given comparison is a relatively weak predictor of whether another examiner would individualize. The extensive variability in annotations also means that we must treat any individual examiner’s minutia counts as interpretations of the (unknowable) information content of the prints: saying “the prints had N corresponding minutiae marked” is not the same as “the prints had N corresponding minutiae.” More consistency in annotations, which could be achieved through standardization and training, should lead to process improvements and provide greater transparency in casework.

ANALYST CONSIDERATIONS

IX. What scientific literature characterizes the effect of analyst qualifications/experience on fingerprint matching accuracy?

80. Busey T., Yu C., Wyatte D., et al. (2011). *Consistency and Variability Among Latent Print Examiners as Revealed by Eye Tracking Methodologies*. *Journal of Forensic Identification*, Vol. 61, No. 1, pp. 60-90.

We recorded the eye positions of 18 expert latent print examiners and 18 novice participants across two separate experiments that were designed to represent abbreviated latent print examinations. In the first experiment, participants completed self-paced latent and inked comparisons presented on a computer monitor while their eyes were tracked with a commercial eye tracker. The similarity of eye fixation patterns was computed for each group of subjects. We found greater variability under some conditions among the experts than the novices in terms of the locations visited. However, experts spent approximately 50% longer than novices inspecting the images, which may have led to differences in strategies adopted by the two groups. A second experiment used trials that always lasted 20 seconds and found that under these time-controlled circumstances, experts were more consistent as a group than novices. Experts also had higher accuracy, spent a greater proportion of time inspecting the latent prints, and had shorter saccades than novices. However, the two groups spent an equal time looking at regions that contained minutiae. The results are generally consistent with experts relying on a common set of features that they choose to move their gaze to under time-limited conditions.

81. Dror I.E., Champod C., Langenburg G., et al. (2011). Cognitive Issues in Fingerprint Analysis: Inter- and Intra-expert Consistency and the Effect of a 'Target' Comparison. *Forensic Science International*, Vol. 208, No. 1-3, pp. 10-17.

Deciding whether two fingerprint marks originate from the same source requires examination and comparison of their features. Many cognitive factors play a major role in such information processing. In this paper we examined the consistency (both between- and within-experts) in the analysis of latent marks, and whether the presence of a "target" comparison print affects this analysis. Our findings showed that the context of a comparison print affected analysis of the latent mark, possibly influencing allocation of attention, visual search, and threshold for determining a "signal". We also found that even without the context of the comparison print, there was still a lack of consistency in analyzing latent marks. Not only was this reflected by inconsistency between different experts, but the same experts at different times were inconsistent with their own analysis. However, the characterization of these inconsistencies depends on the standard and definition of what constitutes inconsistent. Furthermore, these effects were not uniform; the lack of consistency varied across fingerprints and experts. We propose solutions to mediate variability in the analysis of friction ridge skin.

****Evetts I.W. and Williams R.L. (2015). A Review of the Sixteen Points Fingerprint Standard in England and Wales. *Journal of Forensic Identification*, Vol. 65, No. 4, pp. 557-580.**

The fingerprint service of England and Wales works to the requirement that a fingerprint identification should be based on at least 16 points of comparison before evidence may be given in court. In 1988-89 the authors carried out a review of the need for this requirement. The review included: visits to bureaus in the U.K. and in various other countries; a study of the statistical aspects of fingerprint identification; a historical review; and a collaborative study in which fingerprint experts from many different bureaus at home and abroad examined ten sets of comparisons. This paper describes the conduct of the review and its conclusions.

***82. Expert Working Group on Human Factors in Latent Print Analysis. (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.**

Fingerprints have provided a valuable method of personal identification in forensic science and criminal investigations for more than 100 years. Fingerprints left at crime scenes generally are latent prints—unintentional reproductions of the arrangement of ridges on the skin made by the transfer of materials (such as amino acids, proteins, polypeptides, and salts) to a surface. Palms and the soles of feet also have friction ridge skin that can leave latent prints. The examination of a latent print consists of a series of steps involving a comparison of the latent print to a known (or exemplar) print. Courts have accepted latent print evidence for the past century. However, several high-profile cases in the United States and abroad have highlighted the fact that human errors can occur, and litigation and expressions of concern over the evidentiary reliability of latent print examinations and other forensic identification procedures has increased in the past decade. "Human factors" issues can arise in any

experience- and judgment-based analytical process, such as latent print examination. Inadequate training, extraneous knowledge about the suspects in the case or other matters, poor judgment, health problems, limitations of vision, complex technology, and stress are but a few factors that can contribute to errors. A lack of standards or quality control, poor management, insufficient resources, and substandard working conditions constitute other potentially contributing factors. In addition to reaching correct conclusions in the matching process, latent print examiners are expected to produce records of the examination and, in some cases, to present their conclusions and the reasoning behind them in the courtroom. Human factors issues related to the documentation and communication of an examiner's work and findings, therefore, merit attention as well. The study of human factors focuses on the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered at work and in daily living. Human factors analysis can advance our understanding of the nature of errors in complex work settings. Most preventable, adverse events are not just the result of isolated or idiosyncratic behavior but are in part caused by systemic factors. The forensic science community can benefit from the application of human factors research to enhance quality and productivity in friction ridge examinations and to reduce the likelihood and consequences of human error at various stages in the interpretation of evidence. To further this effort, the National Institute of Justice (NIJ) Office of Investigative and Forensic Sciences (OFIS) within the U.S. Department of Justice and the National Institute of Standards and Technology's (NIST's) Law Enforcement Standards Office (OLEs) sponsored the work of this expert panel to examine human factors in latent print analysis and to develop recommendations to reduce the risk of error and improve the practice of latent print analysis.

83. International Association for Identification, et al. (2011). Chapter 15: Special Abilities and Vulnerabilities in Forensic Expertise, in *The Fingerprint Sourcebook*. NCJ 225320.

This chapter is a combination of several studies that were conducted on the forensic expertise. It brings up the importance of experience and perceptual learning that improve the performance of the examiner of a latent fingerprint. It also talks about the configural processing used by the experts to unitize the individual features in a noisy latent print, allowing them to improve the quality of information extracted from this print. In addition to that, it recommends that training should be with clear images since it develops the processes of external noise filtering and the enhancement of weak stimuli, and afterwards generalizes this knowledge to noisy images. Furthermore, the examiners who rely on information that is not easy to verbalize should refine their learning by training on stimulus sets for which the ground truth is known and can be immediately verified. In fact, they create psychological dimensions of stimuli, and integrate and differentiate them depending on the nature of the task. Therefore, the correct working procedures should minimize the psychological and cognitive interferences in making fingerprint matching decisions.

84. Koehler J.J. (2008). Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter. *Hastings Law Journal*, Vol. 59, No. 5 pp. 1077-1098.

The author addresses several questions about current proficiency tests: are they really efficient? Do they measure what we expect? How can we improve them? Properly designed, proficiency tests may provide a reasonable estimation of the rate at which false discoveries, false positive errors, and false negative errors occur. One conclusion offered by the author is that for a proficiency test to provide reasonable estimates of error rates in real casework, neither the examiners nor their supervisors should know that they are working on a proficiency test. The reason is that an examiner's performance is likely to improve when they know they are being tested. Contextual blindness is required to ensure that examiners base their conclusions on the forensic evidence alone. Proficiency participants should be a representative sample of laboratories and examiners drawn from the population of those who provide fingerprint testimony in court.

***Pacheco I., Cerchiai B. and Stoiloff S. (2014). Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations. Miami Dade Police Department/National Institute of Justice, Final Technical Report, Document # 248534.**

This research reports on an empirical study that evaluated the reliability of the Analysis, Comparison, and Evaluation (ACE) and Analysis, Comparison, Evaluation, and Verification (ACE-V) methodologies in latent fingerprint examinations. The participants' performance was measured in terms of accuracy and precision, and was evaluated under both unbiased and biased conditions. Accuracy was measured in terms of the participant's ability to correctly identify or exclude a latent print to a known source(s) and precision was measured in terms of the participant's ability to reproduce and repeat the same conclusion. Reproducibility is defined as the ability of multiple participants to examine the same latent print and reach the same conclusion independently, while repeatability is defined as the participant's ability to provide the same conclusion upon re-evaluation of the same latent print. For the purpose of this research, bias was defined as the ability of a participant to reproduce and repeat a conclusion when presented with two previous conclusions and asked to conduct a second verification. The foundation of latent fingerprint identification is that friction ridge skin is unique and persistent. Through the examination of all of the qualitative and quantitative features available in friction ridge skin, impressions can be positively identified or excluded to the individual that produced it. This study reports the results of four categorical opinions: identification, exclusion, inconclusive, and no value decisions. In addition, sufficiency determinations and comparison decisions were evaluated based on a latent Strength of Value and Difficulty of Comparison rating scale that was designed for this research. Tests were assembled using 80 latent prints with varying quantity and quality of information from ten known sources and were distributed to 109 latent print examiners across the United States. Participants had at least one year of latent print examination experience and employed the ACE methodology when comparing unknown latent prints to known sources. Responses from the participants yielded 5,963 sufficiency determinations, 4,536 ACE decisions, 532 ACE-V decisions, 1,311 repeatability decisions, 326 ACE decisions under biased conditions, and 333 repeatability decisions under biased conditions. This study took into account inconclusive responses in determining error rates and established a False Positive Rate (FPR) of 3.0% and False Negative Rate (FNR) of 7.5% for ACE examinations, as well as a FPR of 0.0% and FNR of 2.9% for ACE-V examinations. Participants were able to reproduce a correct identification 94.2% of the time and not reproduce an erroneous identification 100% of the time. Participants repeated their previous correct identifications 94.6% of the time and did not repeat their previous erroneous exclusions 93.1% of the time. Under biased conditions, participants were able to reproduce a correct identification 73.0% of the time and not reproduce an erroneous identification 96.5% of the time. Additionally, under biased conditions, participants repeated their previous correct identifications 93.2% of the time and did not repeat their previous erroneous exclusions 85.2% of the time.

****Tangen J.M., Thompson M.B. and McCarthy D.J. (2011). Identifying Fingerprint Expertise. *Psychological Science*, Vol. 22, No. 8, pp. 995-997.**

"CSI"-style TV shows give the impression that fingerprint identification is fully automated. In reality, when a fingerprint is found at a crime scene, it is a human examiner who is faced with the task of identifying the person who left the print—a task that falls squarely in the domain of psychology. The difficulty is that no properly controlled experiments have been conducted on fingerprint examiners' accuracy in identifying perpetrators (Loftus & Cole, 2004), even though fingerprints have been used in criminal courts for more than 100 years. Examiners have even claimed to be infallible (Federal Bureau of Investigation, 1984). However, the U.S. National Academy of Sciences has recently condemned these claims as scientifically implausible, reporting that faulty analyses may be contributing to wrongful convictions of innocent people (National Research Council, Committee on Identifying the Needs of the Forensic Science Community, 2009). Proficiency tests of fingerprint examiners and previous studies of examiners' performance have not adequately addressed the issue of accuracy, and they have been heavily criticized for (among other things) failing to include large, counterbalanced samples of targets and distractors for which the ground truth is known (see Cole, 2008, and Vokey, Tangen, & Cole, 2009). Thus, it is not clear what these tests say about the proficiency of fingerprint examiners, if they say anything at all. Researchers at the National Academy of Sciences and elsewhere (e.g., Saks & Koehler, 2005; Spinney, 2010) have argued that there is an urgent need to develop objective measures of accuracy in fingerprint identification. Here we present such data.

***85. Thompson M.B., Tangen J.M. and McCarthy D.J. (2013). Expertise in Fingerprint Identification. *Journal of Forensic Sciences*, Vol. 58, No. 6, pp. 1519- 1530.**

Although fingerprint experts have presented evidence in criminal courts for more than a century, there have been few scientific investigations of the human capacity to discriminate these patterns. A recent latent print matching experiment shows that qualified, court-practicing fingerprint experts are exceedingly accurate (and more conservative) compared with novices, but they do make errors. Here, a rationale for the design of this experiment is provided. We argue that fidelity, generalizability, and control must be balanced to answer important research questions; that the proficiency and competence of fingerprint examiners are best determined when experiments include highly similar print pairs, in a signal detection paradigm, where the ground truth is known; and that inferring from this experiment the statement “The error rate of fingerprint identification is 0.68%” would be unjustified. In closing, the ramifications of these findings for the future psychological study of forensic expertise and the implications for expert testimony and public policy are considered.

***86. Thompson M.B., Tangen J.M. and McCarthy D.J. (2014). Human Matching Performance of Genuine Crime Scene Latent Fingerprints. *Law and Human Behavior*, Vol. 38, No. 1, pp. 84-93.**

There has been very little research into the nature and development of fingerprint matching expertise. Here we present the results of an experiment testing the claimed matching expertise of fingerprint examiners. Expert ($n = 37$), intermediate trainee ($n = 8$), new trainee ($n = 9$), and novice ($n = 37$) participants performed a fingerprint discrimination task involving genuine crime scene latent fingerprints, their matches, and highly similar distractors in a signal detection paradigm. Results show that qualified, court-practicing fingerprint experts were exceedingly accurate compared with novices. Experts showed a conservative response bias, tending to err on the side of caution by making more errors of the sort that could allow a guilty person to escape detection than errors of the sort that could falsely incriminate an innocent person. The superior performance of experts was not simply a function of their ability to match prints, *per se*, but a result of their ability to identify the highly similar but nonmatching fingerprints as such. Comparing these results with previous experiments, experts were even more conservative in their decision making when dealing with these genuine crime scene prints than when dealing with simulated crime scene prints, and this conservatism made them relatively less accurate overall. Intermediate trainees— despite their lack of qualification and average 3.5 years’ experience—performed about as accurately as qualified experts who had an average 17.5 years’ experience. New trainees— despite their 5-week, full-time training course or their 6 months’ experience—were not any better than novices at discriminating matching and similar nonmatching prints, they were just more conservative. Further research is required to determine the precise nature of fingerprint matching expertise and the factors that influence performance. The findings of this representative, lab-based experiment may have implications for the way fingerprint examiners testify in court, but what the findings mean for reasoning about expert performance in the wild is an open, empirical, and epistemological question.

****Thompson M.B. and Tangen J.M. (2014). The Nature of Expertise in Fingerprint Matching: Experts Can Do A Lot with a Little. *PLoS ONE*, Vol. 9, No. 12, pp. 1-23.**

Expert decision making often seems impressive, even miraculous. People with genuine expertise in a particular domain can perform quickly and accurately, and with little information. In the series of experiments presented here, we manipulate the amount of “information” available to a group of experts whose job it is to identify the source of crime scene fingerprints. In Experiment 1, we reduced the amount of information available to experts by inverting fingerprint pairs and adding visual noise. There was no evidence for an inversion effect—experts were just as accurate for inverted prints as they were for upright prints—but expert performance with artificially noisy prints was impressive. In Experiment 2, we separated matching and nonmatching print pairs in time. Experts were conservative, but they were still able to discriminate pairs of fingerprints that were separated by five seconds, even though the task was quite different from their everyday experience. In Experiment 3, we separated the print pairs further in time to test the long-term memory of experts compared to novices. Long-term recognition memory

for experts and novices was the same, with both performing around chance. In Experiment 4, we presented pairs of fingerprints quickly to experts and novices in a matching task. Experts were more accurate than novices, particularly for similar nonmatching pairs, and experts were generally more accurate when they had more time. It is clear that experts can match prints accurately when there is reduced visual information, reduced opportunity for direct comparison, and reduced time to engage in deliberate reasoning. These findings suggest that non-analytic processing accounts for a substantial portion of the variance in expert fingerprint matching accuracy. Our conclusion is at odds with general wisdom in fingerprint identification practice and formal training, and at odds with the claims and explanations that are offered in court during expert testimony.

87. Ulery B.T., Hicklin R.A., Buscaglia J. et al. (2011). Accuracy and Reliability of Forensic Latent Fingerprint Decisions. *Proceedings of the National Academy of Sciences, USA*, Vol. 108, No. 19, pp. 7733-7738.

The interpretation of forensic fingerprint evidence relies on the expertise of latent print examiners. The National Research Council of the National Academies and the legal and forensic sciences communities have called for research to measure the accuracy and reliability of latent print examiners' decisions, a challenging and complex problem in need of systematic analysis. Our research is focused on the development of empirical approaches to studying this problem. Here, we report on the first large-scale study of the accuracy and reliability of latent print examiners' decisions, in which 169 latent print examiners each compared approximately 100 pairs of latent and exemplar fingerprints from a pool of 744 pairs. The fingerprints were selected to include a range of attributes and quality encountered in forensic casework, and to be comparable to searches of an automated fingerprint identification system containing more than 58 million subjects. This study evaluated examiners on key decision points in the fingerprint examination process; procedures used operationally include additional safeguards designed to minimize errors. Five examiners made false positive errors for an overall false positive rate of 0.1%. Eighty-five percent of examiners made at least one false negative error for an overall false negative rate of 7.5%. Independent examination of the same comparisons by different participants (analogous to blind verification) was found to detect all false positive errors and the majority of false negative errors in this study. Examiners frequently differed on whether fingerprints were suitable for reaching a conclusion.

****Ulery B.T., Hicklin R.A., Buscaglia J., et al. (2012). Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. *PLoS ONE*, Vol. 7, No. 3, pp. 1-12.**

The interpretation of forensic fingerprint evidence relies on the expertise of latent print examiners. We tested latent print examiners on the extent to which they reached consistent decisions. This study assessed intra-examiner repeatability by retesting 72 examiners on comparisons of latent and exemplar fingerprints, after an interval of approximately seven months; each examiner was reassigned 25 image pairs for comparison, out of total pool of 744 image pairs. We compare these repeatability results with reproducibility (inter-examiner) results derived from our previous study. Examiners repeated 89.1% of their individualization decisions, and 90.1% of their exclusion decisions; most of the changed decisions resulted in inconclusive decisions. Repeatability of comparison decisions (individualization, exclusion, inconclusive) was 90.0% for mated pairs, and 85.9% for nonmated pairs. Repeatability and reproducibility were notably lower for comparisons assessed by the examiners as "difficult" than for "easy" or "moderate" comparisons, indicating that examiners' assessments of difficulty may be useful for quality assurance. No false positive errors were repeated (n=4); 30% of false negative errors were repeated. One percent of latent value decisions were completely reversed (no value even for exclusion vs. of value for individualization). Most of the inter- and intra-examiner variability concerned whether the examiners considered the information available to be sufficient to reach a conclusion; this variability was concentrated on specific image pairs such that repeatability and reproducibility were very high on some comparisons and very low on others. Much of the variability appears to be due to making categorical decisions in borderline cases.

88. Wertheim K., Langenburg G. and Moenssens A. (2006). A Report of Latent Print Examiner Accuracy During Comparison Training Exercises. *Journal of Forensic Identification*, Vol. 56, No. 1, pp. 55-93.

During comparison training exercises, data from 108 participants were collected. For each participant, the following were recorded: the number of comparisons performed, the number of correct individualizations made, the number of erroneous individualizations made, the number of clerical errors made, and the assessments of the latent prints regarding the quantity and quality of information present in the latent prints in the exercises. Additional information regarding the training and experience of the participant was also gathered in such a manner that preserved the anonymity of the participant. Because the training courses were open to participants of any skill level, including participants with no training and experience, the authors separated the data of participants with more than one year of experience from the data of participants with one year of experience or less. The 92 participants with more than one year of experience made 5861 individualizations (identifications) at the highest level of confidence. Fifty-eight hundred of these individualizations were correct and 61 of these individualizations were one of two types of error: 59 were clerical in nature and 2 were erroneous individualizations. This resulted in an erroneous individualization rate of 0.034% and a clerical error rate of 1.01% for the participants with more than one year of experience during these training exercises. A follow-up experiment was performed involving verification of the errors reported by previous participants. Sixteen participants with more than one year of experience acted as verifiers to previous participants' results. Each verifier was given a packet to verify containing the results of eight correct individualizations and two errors. These 16 independent reviewers did not verify any of the errors given to them in the verification packet exercises.

X. What scientific literature describes and justifies current quality control measures? (What are appropriate quality controls on latent print matching and what justification is there for the recommended controls?)

****Hicklin R.A., et al. (2011). Latent Fingerprint Quality: A Survey of Examiners. *Journal of Forensic Identification*, Vol. 61, No. 4, pp. 385-418.**

A survey of latent print examiners was conducted to determine how they assess fingerprint quality. Participating examiners performed detailed anonymous assessments of both the local and overall quality characteristics of latent and exemplar fingerprint images, using a custom-designed software application. Eighty-six latent print examiners from federal, state, local, international and private sector laboratories each spent 8 to 12 hours assessing the quality of approximately 70 fingerprint images. The fingerprints were overlapping subsets of 1,090 latent and exemplar fingerprint images derived from the National Institute of Standards and Technology (NIST) Special Database 27 and a Federal Bureau of Investigation (FBI) Laboratory dataset of images. An analysis of the results shows the extent of consistency between examiners in value determinations; the relationships between the overall perceived quality of a print and the size of clear ridge detail; and the relationships between quality, size and correct pattern classification. An analysis of the examiners' subjective assessments of fingerprint quality, revealed information useful for the development of guidelines, metrics and software tools for assessing fingerprint quality.

89. International Association for Identification, et al. (2011). Chapter 12: Quality Assurance, in *The Fingerprint Sourcebook*. NCJ 225320.

This chapter is a review specifically dedicated to the quality-assurance topic and proposes information about accreditation and certification organizations. Special reference is given to the SWGFAST standards.

****Ulery B.T., Hicklin R.A., Roberts M.A. and Buscaglia J. (2015). Changes in Latent Fingerprint Examiners' Markup Between Analysis and Comparison. *Forensic Science International*, Vol. 247, pp. 54-61.**

After the initial analysis of a latent print, an examiner will sometimes revise the assessment during comparison with an exemplar. Changes between analysis and comparison may indicate that the initial analysis of the latent was inadequate, or that confirmation bias may have affected the comparison. 170 volunteer latent print examiners, each randomly assigned 22 pairs of prints from a pool of 320 total pairs, provided detailed markup documenting their interpretations of the prints and the bases for their comparison conclusions. We describe changes in value assessments and markup of features and clarity. When examiners individualized, they almost always added or deleted minutiae (90.3% of individualizations); every examiner revised at least some markups. For inconclusive and exclusion determinations, changes were less common, and features were added more frequently when the image pair was mated (same source). Even when individualizations were based on eight or fewer corresponding minutiae, in most cases some of those minutiae had been added during comparison. One erroneous individualization was observed: the markup changes were notably extreme, and almost all of the corresponding minutiae had been added during comparison. Latents assessed to be of value for exclusion only (VEO) during analysis were often individualized when compared to a mated exemplar (26%); in our previous work, where examiners were not required to provide markup of features, VEO individualizations were much less common (1.8%).

XI. What scientific literature establishes the value of replication by a different examiner? (How consistent are results by different examiners? How much error is reduced by verification? How consistent are results of reexamination by the original examiner?)

****Black J.P. (2010). Friction Ridge Examination (Fingerprints): Evaluating the Extent and Scope of "Verification" in Analysis Comparison Evaluation and Verification (ACE-V), in *Wiley Encyclopedia of Forensic Science*, 1st edition. A Jamieson and A Moenssens, Eds. Chichester: John Wiley & Sons Ltd.**

Preliminary research has shown that nearly every agency is performing verifications on 100% of reported fingerprint identifications. However, exclusion, inconclusive, and no value decisions are not being verified as a matter of course. Surprisingly perhaps, it is not erroneous identifications that comprise the bulk of errors, but rather erroneous exclusions as well as incorrect "inconclusive" and "no value" decisions. Therefore, it is appropriate to examine both the extent and scope of the "verification" phase of the analysis, comparison, evaluation, and verification (ACE-V) methodology as utilized by fingerprint examiners.

***90. Black J.P. (2012). Is There a Need for 100% Verification (Review) of Latent Print Examination Conclusions? *Journal of Forensic Identification*, Vol. 62, No. 1, pp. 80-100.**

This research attempts to provide insight on the extent of verification as currently practiced within the latent fingerprint community. Ten questions were posed to this community regarding various aspects of verification; 56 agencies responded. The study results indicate that nearly every agency is performing verifications on 100% of reported fingerprint identifications. The study results also indicate that exclusion, inconclusive, and "no value" decisions are not being verified to the same extent. Interestingly, erroneous identifications constitute the minority of technical fingerprint errors, whereas erroneous exclusions, missed identifications, and inappropriate "inconclusive" and "no value" decisions are far more numerous.

***91. Bradford T.U., Hicklin R.A., Buscaglia J. et al. (2012). Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners. *PLOS ONE*, Vol. 7, No. 3, pp. 1-12.**

The interpretation of forensic fingerprint evidence relies on the expertise of latent print examiners. We tested latent print examiners on the extent to which they reached consistent decisions. This study assessed intra-

examiner repeatability by retesting 72 examiners on comparisons of latent and exemplar fingerprints, after an interval of approximately seven months; each examiner was reassigned 25 image pairs for comparison, out of total pool of 744 image pairs. We compare these repeatability results with reproducibility (inter-examiner) results derived from our previous study. Examiners repeated 89.1% of their individualization decisions, and 90.1% of their exclusion decisions; most of the changed decisions resulted in inconclusive decisions. Repeatability of comparison decisions (individualization, exclusion, inconclusive) was 90.0% for mated pairs, and 85.9% for non-mated pairs. Repeatability and reproducibility were notably lower for comparisons assessed by the examiners as “difficult” than for “easy” or “moderate” comparisons, indicating that examiners’ assessments of difficulty may be useful for quality assurance. No false positive errors were repeated (n = 4); 30% of false negative errors were repeated. One percent of latent value decisions were completely reversed (no value even for exclusion vs. of value for individualization). Most of the inter- and intra-examiner variability concerned whether the examiners considered the information available to be sufficient to reach a conclusion; this variability was concentrated on specific image pairs so that repeatability and reproducibility were very high on some comparisons and very low on others. Much of the variability appears to be due to making categorical decisions in borderline cases.

92. Langenburg G. (2009). A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability and Biasability of Conclusions Resulting from the ACE-V Process. *Journal of Forensic Identification*, Vol. 59, No. 2, pp. 219-257.

Six fingerprint analysts participated in a series of tests to measure the accuracy, precision, reproducibility, repeatability, and biasability during 60 ACE and 60 ACE-V trials. The results of the ACE testing, where each analyst received the same set of 60 fingerprint comparisons, showed 100% accuracy for all trials where an opinion of identification was reported (N=268) and an 86% accuracy for all trials where an opinion of exclusion was reported (N=14). The precision tests for the four categorical opinions reported (i.e., identification, exclusion, inconclusive, and no value) all passed threshold criteria that were determined before the test was administered. Reproducibility (the ability of the experts to all reach the same result independently) and repeatability (the ability of the test to provide the same answer upon re-analysis of the material) were both assessed in these experiments. The results varied depending on the amount of information present in the friction ridge impressions and generally how the images were presented to the participants.

XII. What scientific literature examines the range of innate human pattern matching capabilities? (General pattern recognition, latent print matching-specific pattern recognition)

93. Busey T.A. and Parada F.J. (2010). The Nature of Expertise in Fingerprint Examiners. *Psychonomic Bulletin & Review*, Vol. 17, No. 2, pp. 155-160.

Latent print examinations involve a complex set of psychological and cognitive processes. This article summarizes existing work that has addressed how training and experience creates changes in latent print examiners. Experience appears to improve overall accuracy, increase visual working memory, and lead to configural processing of upright fingerprints. Experts also demonstrate a narrower visual filter and, as a group, tend to show greater consistency when viewing ink prints. These findings address recent criticisms of latent print evidence, but many open questions still exist. Cognitive scientists are well positioned to conduct studies that will improve the training and practices of latent print examiners, and suggestions for becoming involved in fingerprint research are provided.

***94. Expert Working Group on Human Factors in Latent Print Analysis. (2012). *Latent Print Examination and Human Factors: Improving the Practice Through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.**

Fingerprints have provided a valuable method of personal identification in forensic science and criminal investigations for more than 100 years. Fingerprints left at crime scenes generally are latent prints—unintentional

reproductions of the arrangement of ridges on the skin made by the transfer of materials (such as amino acids, proteins, polypeptides, and salts) to a surface. Palms and the soles of feet also have friction ridge skin that can leave latent prints. The examination of a latent print consists of a series of steps involving a comparison of the latent print to a known (or exemplar) print. Courts have accepted latent print evidence for the past century. However, several high-profile cases in the United States and abroad have highlighted the fact that human errors can occur, and litigation and expressions of concern over the evidentiary reliability of latent print examinations and other forensic identification procedures have increased in the past decade. "Human factors" issues can arise in any experience- and judgment-based analytical process such as latent print examination. Inadequate training, extraneous knowledge about the suspects in the case or other matters, poor judgment, health problems, limitations of vision, complex technology, and stress are but a few factors that can contribute to errors. A lack of standards or quality control, poor management, insufficient resources, and substandard working conditions constitute other potentially contributing factors. In addition to reaching correct conclusions in the matching process, latent print examiners are expected to produce records of the examination and, in some cases, to present their conclusions and the reasoning behind them in the courtroom. Human factors issues related to the documentation and communication of an examiner's work and findings therefore merit attention as well. The study of human factors focuses on the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered at work and in daily living. Human factors analysis can advance our understanding of the nature of errors in complex work settings. Most preventable, adverse events are not just the result of isolated or idiosyncratic behavior but are in part caused by systemic factors. The forensic science community can benefit from the application of human factors research to enhance quality and productivity in friction ridge examinations and to reduce the likelihood and consequences of human error at various stages in the interpretation of evidence. To further this effort, the National Institute of Justice (NIJ) Office of Investigative and Forensic Sciences (OFIS) within the U.S. Department of Justice and the National Institute of Standards and Technology's (NIST's) Law Enforcement Standards Office (OLEs) sponsored the work of this expert panel to examine human factors in latent print analysis and to develop recommendations to reduce the risk of error and improve the practice of latent print analysis.

XIII. What scientific literature establishes the key sources of bias and characterizes the effectiveness of measures to mitigate the bias? (How much does analyst knowledge of other case evidence influence accuracy? How much is accuracy affected by "blind" or "non-blind" status?)

****Arora S., et al. (2015). Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups. *Proceedings of the 8th International Conference on Biometrics (ICB)*, Phuket, Thailand, May 2015.**

Automatic matching of poor quality latent fingerprints to rolled/slap fingerprints using an Automated Fingerprint Identification System (AFIS) is still far from satisfactory. Therefore, it is a common practice to have a latent examiner mark features on a latent for improving the hit rate of the AFIS. We propose a synergistic crowd powered latent identification framework where multiple latent examiners and the AFIS work in conjunction with each other to boost the identification accuracy of the AFIS. Given a latent, the candidate list output by the AFIS is used to determine the likelihood that a hit at rank-1 was found. A latent for which this likelihood is low is crowdsourced to a pool of latent examiners for feature markup. The manual markups are then input to the AFIS to increase the likelihood of making a hit in the reference database. Experimental results show that the fusion of an AFIS with examiner markups improves the rank-1 identification accuracy of the AFIS by 7.75% (using six markups) on the 500 ppi NIST SD27, 11.37% (using two markups) on the 1000 ppi ELFT-EFS public challenge database, and by 2.5% (using a single markup) on the 1000 ppi RS&A database against 250,000 rolled prints in the reference database.

95. Dror I.E. and Cole S.A. (2010). The Vision in “Blind” Justice: Expert Perception, Judgment and Visual Cognition in Forensic Pattern Recognition. *Psychonomic Bulletin & Review*, Vol. 17, No. 2, pp. 161-167.

Many forensic disciplines require experts to judge whether two complex patterns are sufficiently similar to conclude that both originate from the same source. Studies in this area have revealed that there are a number of factors that affect perception and judgment and that decisions are subjective and susceptible to extraneous influences (such as emotional context, expectation, and motivation). Some studies have shown that the same expert examiner, examining the same prints but within different contexts, may reach different and contradictory decisions. However, such effects are not always present; some examiners seem more susceptible to such influences than do others- especially when the pattern matching is “hard to call” and when the forensic experts are not aware that they are being observed in an experimental study. Studying forensic examiners can contribute to our understanding of expertise and decision making, as well as have implications for forensic science and other areas of expertise.

96. Dror I.E., Champod C., Langenburg G., et al. (2011). Cognitive Issues in Fingerprint Analysis: Inter- and Intra-expert Consistency and the Effect of a ‘Target’ Comparison. *Forensic Science International*, Vol. 208, No. 1-3, pp. 10-17.

Deciding whether two fingerprint marks originate from the same source requires examination and comparison of their features. Many cognitive factors play a major role in such information processing. In this paper we examined the consistency (both between- and within-experts) in the analysis of latent marks, and whether the presence of a “target” comparison print affects this analysis. Our findings showed that the context of a comparison print affected analysis of the latent mark, possibly influencing allocation of attention, visual search, and threshold for determining a “signal”. We also found that even without the context of the comparison print, there was still a lack of consistency in analyzing latent marks. Not only was this reflected by inconsistency between different experts, but the same experts at different times were inconsistent with their own analysis. However, the characterization of these inconsistencies depends on the standard and definition of what constitutes inconsistent. Furthermore, these effects were not uniform; the lack of consistency varied across fingerprints and experts. We propose solutions to mediate variability in the analysis of friction ridge skin.

****Dror I.E., Wertheim K., Fraser-Mackenzie P. and Walajtys J. (2012). The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. *Journal of Forensic Sciences*, Vol. 57, No. 2, pp. 343-352.**

Experts play a critical role in forensic decision making, even when cognition is offloaded and distributed between human and machine. In this paper, we investigated the impact of using Automated Fingerprint Identification Systems (AFIS) on human decision makers. We provided 3680 AFIS lists (a total of 55,200 comparisons) to 23 latent fingerprint examiners as part of their normal casework. We manipulated the position of the matching print in the AFIS list. The data showed that latent fingerprint examiners were affected by the position of the matching print in terms of false exclusions and false inclusions. Furthermore, the data showed that false identification errors were more likely at the top of the list and that such errors occurred even when the correct match was present further down the list. These effects need to be studied and considered carefully, so as to optimize human decision making when using technologies such as AFIS.

****Dror IE, et al. (2015). Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making. *Journal of Forensic Sciences*, Vol. 60, No. 4, pp. 1111-1112.**

Letter to the Editor.

***97. Fraser-Mackenzie P.A.F., Dror I.E. and Wertheim K. (2013). Cognitive and Contextual Influences in Determination of Latent Fingerprint Suitability for Identification Judgments. *Science and Justice*, Vol. 53, No. 1, pp. 144-153.**

We examined forensic fingerprint examiners' suitability determinations of latent fingerprints comparing situations in which the latent is assessed solo (in isolation) versus situations in which it is presented alongside a comparison (matching or non-matching) exemplar print. The presence of a non-matching comparison exemplar led examiners to be more inclined to draw the conclusion that the latent was suitable for comparison compared to when the latent was presented solo. This effect persisted even when the latent presented was highly unsuitable for comparison. The presence of a matching comparison exemplar led examiners to be less likely to decide that the latent was suitable and more likely to decide the latent was questionable compared to solo analysis. This effect persisted even when the latent presented was highly suitable, suggesting a strong main effect. Knowledge of another examiner's previous determination that the latent was unsuitable was found to increase the likelihood that the examiner would conclude that the latent was unsuitable. However, knowledge of a previous "suitable" determination by another examiner did not increase the likelihood of a "suitable" conclusion by examiners. The finding that effects were weaker, although not entirely removed, in those with IAI certification suggests that training may be an appropriate route for reducing the effect of contextual influence and bias in suitability determinations. It was also shown that latent prints that were previously classed as "unsuitable" in a non-biasing context continued to be judged to be "unsuitable" in a strongly biasing context (a major case in which a previous examiner was purported to have made an Individualization).

98. Langenburg G. Champod C. and Wertheim P. (2009). Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology When Conducting Fingerprint Comparisons. *Journal of Forensic Sciences*, Vol. 54, No. 3, pp. 571-582.

This study was conducted to assess if fingerprint specialists could be influenced by extraneous contextual information during a verification process. Participants were separated into three groups: a control group (no contextual information was given), a low bias group (minimal contextual information was given in the form of a report prompting conclusions), and a high bias group (an internationally recognized fingerprint expert provided conclusions and case information to deceive this group into believing that it was his case and conclusions). A similar experiment was later conducted with laypersons. The results showed that fingerprint experts were influenced by contextual information during fingerprint comparisons, but not towards making errors. Instead, fingerprint experts under the biasing conditions provided significantly fewer definitive and erroneous conclusions than the control group. In contrast, the novice participants were more influenced by the bias conditions and did tend to make incorrect judgments, especially when prompted towards an incorrect response by the bias prompt.

99. Thompson W.C. (2009). Interpretation: Observer Effects, in *Wiley Encyclopedia of Forensic Sciences*, A Moenssens and A Jamieson, Eds. Chichester, UK: John Wiley & Sons, pp. 1575-1579.

The tendency of human observers to interpret data in a manner consistent with their expectations and desires is well established. Scientists in many fields use blind or double-blind procedures to minimize this tendency. Forensic scientists have recently given greater attention to this issue as a result of high profile errors in forensic testing and empirical research confirming that observer effects influence forensic interpretations. To minimize such problems, forensic laboratories can separate case management from evidence interpretation or adopt sequential unmasking procedures that allow key interpretations to occur while the analyst is blind to extraneous facts.

XIV. What scientific literature establishes the basis for appropriate false match vs. missed match probabilities? (How do the probabilities differ for automatic and manual methods?)

XV. What scientific literature establishes the overall accuracy of current fingerprint analysis methods? (What is the error performance of individual steps in each method? Does the number of comparisons performed in single evaluation influence accuracy?)

***100. Expert Working Group on Human Factors in Latent Print Analysis. (2012). *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*. U.S. Department of Commerce, National Institute of Standards and Technology.**

Fingerprints have provided a valuable method of personal identification in forensic science and criminal investigations for more than 100 years. Fingerprints left at crime scenes generally are latent prints—unintentional reproductions of the arrangement of ridges on the skin made by the transfer of materials (such as amino acids, proteins, polypeptides, and salts) to a surface. Palms and the soles of feet also have friction ridge skin that can leave latent prints. The examination of a latent print consists of a series of steps involving a comparison of the latent print to a known (or exemplar) print. Courts have accepted latent print evidence for the past century. However, several high-profile cases in the United States and abroad have highlighted the fact that human errors can occur, and litigation and expressions of concern over the evidentiary reliability of latent print examinations and other forensic identification procedures has increased in the past decade. “Human factors” issues can arise in any experience- and judgment-based analytical process such as latent print examination. Inadequate training, extraneous knowledge about the suspects in the case or other matters, poor judgment, health problems, limitations of vision, complex technology, and stress are but a few factors that can contribute to errors. A lack of standards or quality control, poor management, insufficient resources, and substandard working conditions constitute other potentially contributing factors. In addition to reaching correct conclusions in the matching process, latent print examiners are expected to produce records of the examination and, in some cases, to present their conclusions and the reasoning behind them in the courtroom. Human factors issues related to the documentation and communication of an examiner’s work and findings therefore merit attention as well. The study of human factors focuses on the interaction between humans and products, decisions, procedures, workspaces, and the overall environment encountered at work and in daily living. Human factors analysis can advance our understanding of the nature of errors in complex work settings. Most preventable, adverse events are not just the result of isolated or idiosyncratic behavior but are in part caused by systemic factors. The forensic science community can benefit from the application of human factors research to enhance quality and productivity in friction ridge examinations and to reduce the likelihood and consequences of human error at various stages in the interpretation of evidence. To further this effort, the National Institute of Justice (NIJ) Office of Investigative and Forensic Sciences (OFIS) within the U.S. Department of Justice and the National Institute of Standards and Technology’s (NIST’s) Law Enforcement Standards Office (OLES) sponsored the work of this expert panel to examine human factors in latent print analysis and to develop recommendations to reduce the risk of error and improve the practice of latent print analysis.

***101. Langenburg G., Champod C. and Genessay T. (2012). *Informing the Judgments of Fingerprint Analysts Using Quality Metric and Statistical Assessment Tools*. *Forensic Science International*, Vol. 219, No. 1, pp. 183-198.**

The aim of this research was to evaluate how fingerprint analysts would incorporate information from newly developed tools into their decision-making processes. Specifically, we assessed effects using the following: (1) a quality tool to aid in the assessment of the clarity of the friction ridge details, (2) a statistical tool to provide likelihood ratios representing the strength of the corresponding features between compared fingerprints, and (3) consensus information from a group of trained fingerprint experts. The measured variables for the effect on examiner performance were the accuracy and reproducibility of the conclusions against the ground truth (including the impact on error rates) and the analyst accuracy and variation for feature selection and comparison. The results showed that participants using the consensus information from other fingerprint experts demonstrated more consistency and accuracy in minutiae selection. They also demonstrated higher accuracy, sensitivity, and

specificity in the decisions reported. The quality tool also affected minutiae selection (which, in turn, had limited influence on the reported decisions); the statistical tool did not appear to influence the reported decisions.

102. Neumann C. Mateos-Garcia I., Langenburg G., et al. (2011). Operational Benefits and Challenges of the Use of Fingerprint Statistical Models: A Field Study. *Forensic Science International*, Vol. 212, No. 1-3, pp. 32-46.

Research projects aimed at proposing fingerprint statistical models based on the likelihood ratio framework have shown that low quality finger impressions left on crime scenes may have significant evidential value. These impressions are currently either not recovered, considered to be of no value when first analyzed by fingerprint examiners or lead to inconclusive results when compared to control prints. There are growing concerns within the fingerprint community that recovering and examining these low quality impressions will result in a significant increase of the workload of fingerprint units and ultimately of the number of backlogged cases. This study was designed to measure the number of impressions not currently recovered or not considered for examination and to assess the usefulness of these impressions in terms of the number of additional detections that would result from their examination.

***103. Swofford H.J. (2012). Individualization Using Friction Skin Impressions: Scientifically Reliable, Legally Valid. *Journal of Forensic Identification*, Vol. 62, No. 1, pp. 62-79.**

The adversarial structure of the American judicial system encourages critical reviews and challenges of forensic evidence. As a result, the discriminatory power of friction ridge skin impression evidence has been a prime target of debate among critics of the latent print discipline for years, the primary argument being friction ridge skin examination is neither scientifically reliable nor legally valid. Therefore, these critics advocate the exclusion of expert testimony to identifications from the legal system. This article reviews some long-held challenges to the science of friction ridge examination, including challenges to the premise of friction ridge skin uniqueness, testimonial claims of individualization, reliability of comparative interpretations, errors and error rate data, and the legal admissibility according to Daubert standards. The flawed logic on which these challenges are based is presented along with evidence in response to the challenges regarding the scientific reliability and legal validity of the science of the examination of friction ridge skin examination.

***104. Swofford H., Steffan S., Warner G., et al. (2013). Inter- and Intra- Examiner Variation in the Detection of Friction Ridge Skin Minutiae. *Journal of Forensic Identification*, Vol. 63, No. 5, pp. 553-570.**

Friction ridge skin minutiae (bifurcations, ridge endings, dots) and their unique arrangements are the primary information detected and evaluated by ten-print and latent print examiners when comparing unknown friction ridge skin impressions to known (record) impressions. During the analysis of friction ridge skin impressions, examiners usually detect and interpret the minutiae available for comparison to the known impression. Because this is a subjective process, the detection and interpretation of minutiae is prone to variation. Whereas earlier studies have demonstrated inter-examiner variation using impressions having a wide range of quality, this study focuses on high-quality impressions to evaluate a base-line level of variation that can be expected when detecting and interpreting friction ridge skin minutiae under optimal conditions. The standard deviation (SD) of total minutiae detected fluctuated depending on the image, whereas it was much higher for those impressions bearing breaks in the ridges as a result of creases. When comparing various examiner demographics, many of the observed inter- and intra-examiner variations in the detection of minutiae were to a statistically significant degree (95% confidence level). Although the analysis of friction ridge skin minutiae is inherently subjective, variation in the detection of minutiae may not necessarily translate into variation of examiners' conclusions nor should be necessarily considered a limitation of the discipline. Nevertheless, efforts should be made by the discipline to reduce as much variation as possible. Accordingly, these findings suggest that attention should be given towards the creation of standards and guidelines related to defining and selecting minutiae and further emphasize the

importance of documenting the specific minutiae and related information detected by examiners during the analysis of friction ridge skin impressions to facilitate greater transparency of the information relied upon to reach a suitability determination or conclusions (identification, exclusion, or inconclusive).

***105. Swofford H., Steffan S., Warner G., et al. (2013). Impact of Minutiae Quantity on the Behavior and Performance of Latent Print Examiners. *Journal of Forensic Identification*, Vol. 63, No. 5, pp. 571-591.**

Although friction ridge skin is widely accepted to be unique, impressions of the friction ridge skin are not perfect reproductions of the skin and, therefore, will vary in their discriminating strength, depending on the quantity and quality of the minutiae and other features reproduced. Forensic examiners routinely analyze impressions and make determinations, based on their training and experience, of whether the discriminating strength of the features in an impression is such that a decision of identification or exclusion is warranted (e.g., whether the print is of value). Although minutiae quantity is not the sole factor for basing value determinations, it has been found through previous studies to play a major role. Because examiners' training and experience will vary, this study seeks to understand, in general, how examiners' decision-making behavior changes when faced with comparisons of friction ridge skin when minutiae quantity varies, but quality remains very high. The results indicate the decision-making behavior is affected in a predictable manner between inconclusive and identification decisions (for mated sources) based on the number of minutiae present. Eighty percent (80%) or more of examiners' decisions were identification for mated sources when seven or more minutiae were present. No further increase in the relationship between examiner decision and minutiae quantity was observed for impressions with more than seven minutiae. These findings correspond well to the sufficiency chart published by the Scientific Working Group for Friction Ridge Analysis, Study and Technology (SWGFAST) in the area pertaining to high-quality impressions. Additionally, there appears to be no relationship between minutiae quantity and erroneous exclusion decision. When presented with the same comparison twice, nine examiners (17%) changed their decision between inconclusive and the correct decision or vice versa. This study provides greater understanding of how minutiae quantity may affect examiners' decision-making behavior when faced with high-quality impressions. Although further research is needed with lower quality impressions, the results from this study suggest minutiae quantity may be a factor that forensic laboratories may consider when triaging which impressions should undergo enhanced measures of quality assurance.

WORKING GROUP QUESTIONS – LATENT FINGERPRINT EXAMINATION

I. Is there an adequate scientific foundation for understanding the degree of variability of fingerprints: (a) among unrelated individuals; and (b) among relatives?

II. Is there an adequate scientific foundation for understanding the degree of variability among prints made by the same finger: (a) on different surfaces, under different environmental conditions; and (b) over time as a person ages or is subject to injury?

III. Is there an adequate scientific foundation for understanding the accuracy of automated fingerprint identification systems (AFIS)?

IV. Is there an adequate scientific foundation for understanding the potential for contextual bias in latent print analysis and how might it be addressed?

V. Is there an adequate scientific foundation for understanding the accuracy of human fingerprint examiners and how their accuracy is affected by: (a) level of training and experience; (b) individual differences in perceptual ability; (c) analytic procedures and standards of practice; (d) quality control and quality assurance procedures; and (e) the quality of prints? If not, what kinds of research are needed to improve understanding of these issues?

VI. In light of the existing scientific literature, what kind of statements might fingerprint examiners reasonably make in reports and testimony in order to appropriately convey both the strength and uncertainty associated with fingerprint evidence?

D. WORKING GROUP

John Black – Owner, Black & White Forensics, LLC
(Forensic Scientist) CLPE, CFWE, CSCSA

Anil Jain, Ph.D. – Michigan State University
(Biometric Engineer)

Jay Kadane, Ph.D. – Carnegie Mellon University
(Statistician)

William Thompson, Ph.D., J.D. (CHAIR) – University of California, Irvine
(Human Factors)

E. WORKING GROUP BIOS

JOHN P. BLACK

John P. Black is the owner of Black & White Forensics, LLC in South Carolina. He holds a Bachelor of Science degree in Forensic Chemistry from Ohio University. John has 26 years' experience as a forensic scientist and has worked for Ron Smith and Associates, Inc., the South Carolina Law Enforcement Division and the Drug Enforcement Administration. He has spent the last 22 years as a latent print and footwear examiner, with 12 of those years also being devoted to crime scene investigation. John is certified by the International Association for Identification (IAI) as a Latent Print Examiner, Footwear Examiner and Senior Crime Scene Analyst. He is a former member of the IAI's Crime Scene Certification Board, Ethics Committee, Editorial Review Board and Footwear/Tire Track Subcommittee. John also served as a member of SWGFAST. He has provided expert witness testimony in Federal and State courts in over 100 trials. John's research interests include the examination of simultaneous impressions, exclusion decisions and the verification of all conclusions. His research has been published in the Journal of Forensic Identification and the Wiley Encyclopedia of Forensic Science. John has conducted nearly 250 training classes throughout the United States, Africa, Asia, Canada, Central America and Europe. John was awarded Distinguished Member status with the IAI in 2007. In October 2014, he was selected through a competitive application process to the Friction Ridge Subcommittee within the Organization of Scientific Area Committees (OSAC).

ANIL K. JAIN

Anil Jain is a University Distinguished Professor in the Department of Computer Science & Engineering at Michigan State University. His research interests are computer vision, pattern recognition and biometric recognition. He has received Guggenheim Fellowship, Humboldt Senior Research Award, Fulbright Fellowship, IEEE Computer Society Technical Achievement Award, IEEE W. Wallace McDowell Award, IAPR King-Sun Fu Prize, and IEEE Data Mining Research Contributions Award. He served as the Editor-in-Chief of the IEEE Trans. Pattern Analysis and Machine Intelligence and is a Fellow of ACM, IEEE, AAAS, IAPR and SPIE. He holds a number of patents on fingerprint matching and is the author of several books on biometric recognition and data clustering. He was a member of the National Research Council studies on Whither Biometrics and Improvised Explosive Devices. He served as a member of the Defense Science Board and the Forensic Science Standards Board. He is a member of the National Academy of Engineering and a foreign fellow of the Indian National Academy of Engineering.

JOSEPH B. (JAY) KADANE

Joseph B. (Jay) Kadane is the Leonard J. Savage University Professor of Statistics and Social Sciences, Emeritus at Carnegie Mellon University. One of the early proponents of Bayesian Statistics, Kadane has used the theory in both its decision-theoretic foundations and in problems of elicitation and computation to solve political science, law, physics, medicine and computer science problems. He is also well known for his contributions to the field of econometrics and for his work in applying statistics to decision-making. After joining Carnegie Mellon in 1971, he served as head of the Department of Statistics from 1972-1981 and instilled a balance within the department between theoretical and applied work. Kadane has authored more than 250 peer-review articles. His book, "Principles of Uncertainty" won the 2011 International Society for Bayesian Analysis' DeGroot Prize. His most recent book is "Pragmatics of Uncertainty". Kadane has been elected as Fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, American Statistical Association and Institute of Mathematical Sciences. He has served the statistical community in many capacities over the years; he recently served as chair of the American Statistical Association's Committee on Scientific Freedom and Human Rights. Kadane received his bachelor's degree in mathematics from Harvard University in 1962 and his Ph.D. in statistics from Stanford University in 1966. Prior to joining Carnegie Mellon's faculty, Kadane was assistant professor of statistics at Yale University (1966-1968) and a member of the professional staff at the Center for Analyses from 1968-1971.

WILLIAM C. THOMPSON

William C. Thompson is a professor in the Department of Criminology, Law & Society at the University of California, Irvine (UCI); he has joint appointments in Psychology and in UCI's School of Law, where he has taught Evidence. He received a Ph.D. in psychology from Stanford University and a J.D. from the University of California, Berkeley. He studies human judgment and decision making with a particular focus on cognitive and contextual bias in scientific assessments, on the use (and occasional misuse) of scientific and statistical evidence in the courtroom, and on lay perceptions of scientific and statistical evidence. His research has been funded by the National Science Foundation, the National Institute of Justice, and the Center for Statistics and Applications in Forensic Evidence (CSAFE). Thompson was a member of the Task Force that drafted the ABA's Standards on DNA Evidence, he served on the California Crime Laboratory Review Task Force, and he has been a member of SWG-Speaker—the scientific working group on speaker identification. He was a member of the Human Factors Subcommittee of the National Commission on Forensic Science and is Chair of the Human Factors Committee of the Organization of Scientific Advisory Committees (OSAC), a federal standards-setting organization for forensic science that is jointly sponsored by the US Department of Justice and the National Institute of Standards and Technology (NIST).

F. PROJECT ADVISORY COMMITTEE

Martha Bashford, J.D.

Chief, Sex Crimes Unit
New York County District Attorney

Shari Seidman Diamond, J.D., Ph.D.

Professor of Law and Psychology
Northwestern University School of Law
Research Professor, American Bar
Foundation

Itiel Dror, Ph.D.

University College of London &
Cognitive Consultants International Ltd.

Jules Epstein, J.D.

Professor of Law and Director of Advocacy
Programs
Temple University Beasley School of Law

Barbara Hervey, J.D.

Judge, Texas Court of Criminal Appeals

Gilbert S. Omenn, M.D., Ph.D.

Director, Center for Computational Medicine and Bioinformatics
University of Michigan

Hal Stern, Ph.D.

Professor of Statistics
University of California, Irvine

AAAS STAFF

Mark S. Frankel

Former Project Director

Deborah Runkle

Project Manager

Michelle Barretta

Project Assistant