Forensic Science International 289 (2018) 215–222

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Resolving latent conflict: What happens when latent print examiners enter the cage?

Alicia Rairden[a,*], Brandon L. Garrett[b], Sharon Kelley[c], Daniel Murrie[c], Amy Castillo[a]

[a] Houston Forensic Science Center, 1301 Fannin St, Ste 170, Houston, TX 77002, United States
[b] University of Virginia School of Law, 580 Massie Road, Charlottesville, VA 22903-1738, United States
[c] Institute of Law, Psychiatry and Public Policy, University of Virginia, 1230 Cedars Ct, Charlottesville, VA 22903, United States

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form 16 April 2018
Accepted 19 April 2018
Available online xxx

Keywords:
Latent prints
Latents
Fingerprint
Verification
Consultation
Conflict

ABSTRACT

Latent print examination traditionally follows the ACE-V process, in which latent prints are first analyzed to determine whether they are suitable for comparison, and then compared to an exemplar and evaluated for similarities and differences. Despite standard operating procedures and quality controls designed, in part, to mitigate differences between examiners, latent print processing and review are inherently subjective. The ACE-V process addresses subjectivity, and the possibility of error, in the verification stage in which a second examiner repeats the analysis, comparison, and evaluation steps in a given case. Other procedures outside the ACE-V framework, such as consultation and conflict resolution, provide further opportunity to understand how differences between latent print examiners emerge. Despite the growing body of research on latent print examination, questions have emerged about how these procedures work in practice. This study reviews case processing data for two years of casework at the Houston Forensic Science Center (HFSC). We describe these data as cases proceed through each step of the ACE-V process, with a particular focus on verification, consultation, and conflict resolution. We discuss trends in these processes regarding modal types of disagreements, modal outcomes, and roles of the examiners involved. Results reveal implications for improving the practice of latent print examination.

The President's Council of Advisors on Science and Technology (PCAST) issued a high-profile report in September 2016, concluding that latent fingerprint comparison is a foundationally valid subjective methodology [1]. That report highlighted, however, that for any subjective method, the performance of individual examiners may vary and therefore monitoring examiner variability is crucial. One way of better understanding this variability—and possible sources of bias and error—is following routine case processing data through the traditional analysis, comparison, evaluation, and verification (ACE-V) process. The verification stage, as well as the procedures used to resolve disagreements between latent print examiners (i.e., consultation and conflict resolution), provide particularly useful frameworks for understanding the processes that result in differences among latent print examiners. To understand verification, consultation, and conflict resolution, it is important to highlight what transpires when they are used, specifically how often they produce new information or changed conclusions in latent print examinations. This study analyzes latent print examination outcomes from two years of case work, including the use of verification, consultation, and conflict resolution procedures, to determine trends in their occurrence at the Houston Forensic Science Center (HFSC). Although prior research on latent print examination has addressed the individual components of the ACE-V process separately, the goal of our review was to examine the potential for conflict at each stage in the latent print examination process.

1. Analysis

At the analysis stage, examiners gather and interpret data contained within the latent print impression according to three levels of detail: anatomical source, ridge flow and orientation, ridge path deviations, or minutia, and intrinsic morphological ridge characteristics [2,3]. As part of this process, examiners determine the relative weight for each observed feature and the tolerances for variabilities in appearance [4]. Then, examiners decide whether the latent print is of value for comparison and evaluation.

Existing research suggests that that value determinations are strongly influenced by minutiae count [5]. Ulery et al. [5] found

* Corresponding author.
E-mail address: arairden@houstonforensicscience.org (A. Rairden).

https://doi.org/10.1016/j.forsciint.2018.04.040
0379-0738/

that other metrics, such as image clarity or separately weighting "debatable" and "definitive" minutiae, did not improve prediction of examiners' value determinations. Research has not revealed a perfect line of demarcation, or a specific number of minutiae that clearly distinguished whether examiners consider a print to be of value. However, results suggest that counts greater than 7 were more strongly associated with value determinations and a threshold of 12 minutiae (the standard used in some countries) accurately predicted 84% of examiners' value for individualization determinations [5,6].

Individual differences among examiners at the analysis stage exist with respect to both process (e.g., number of minutiae annotated) and outcome (i.e., the ultimate value decision). A number of studies have documented that examiners vary widely in the number of minutiae they annotate or mark during the analysis stage [5–10]. For example, Langenburg [7] found that the number of minutiae documented for a single latent print ranged from 3 to 45 among experienced latent print examiners.

Research documenting poor reliability regarding minutiae and value determinations has prompted calls for more structure in the analysis stage. Langenburg and Champod [8] developed a system, termed GYRO (green, yellow, red, orange), that visualizes the relative weight, or confidence, an examiner places on any given minutiae. Their approach emphasized transparency in the analytical process by tasking examiners with marking survey latent prints with different colors that indicate the weight, tolerance, and expectation to note the same feature in a control exemplar of the same area. The study also explored differences between U.S. and Dutch examiners, as Dutch examiners receive standardized training before qualifying as experts in latent print examination and conform to a twelve-point standard before formulating a conclusion. Even with the GYRO system, substantial differences were observed both between and within examiners regarding number of minutiae, though the Dutch examiners produced less variability in their responses. Thus, there is some evidence that standardized training and increased structure can reduce examiner variability at the analysis stage. Even so, the task of determining whether a latent print is of value for comparison and evaluation remains vulnerable to individual differences between examiners.

Research examining the reliability of conclusions at the analysis stage suggests that disagreement about whether a print is of value for further comparison is relatively common. For instance, Neumann et al. [10] found that, of the 15 latent prints used in their study, 14 of them received all three determinations (no value, value for identification, value for exclusion only). Research by Ulery et al. [11,12] indicated that examiners differed on their value conclusions for 57% of latent prints used in their studies. Given the strong relationship between minutiae count and value determinations, it is perhaps unsurprising that inter-rater agreement of value for identification determinations with low minutiae counts and no value determinations with high minutiae counts is particularly low [5].

2. Comparison and evaluation

Just as number of minutiae was strongly associated with value determinations at the analysis stage, number of corresponding minutiae detected during comparison is strongly associated with examiners' ultimate conclusions at the evaluation stage [9]. This research suggested a "tipping point" of seven minutiae in evaluative conclusions: counts of greater than seven corresponding minutiae were associated with individualization, and the transition from inconclusive to individualization generally occurred between about six to nine corresponding minutiae [9,p. 5].

The corresponding minutiae examiners use to make evaluative conclusions may not have been annotated during the analysis stage, however. In what appears to be the only study that evaluated changes in markup of latent prints after examiners were exposed to exemplars, Ulery et al. [6] found that changes were common, particularly among examiners concluding individualization. In fact, examiners added or deleted minutiae in 90% of individualizations, such that individualizations were associated with more moved, deleted, and added minutiae than any other determination. Overall, in this study, the comparison stage resulted in a net increase in number of minutiae annotated on latent impressions.

Similar to findings from research on the analysis stage, variability in minutiae count is strongly associated with low examiner agreement at the evaluation stage [9,13]. As an example of the range of this disagreement, Evett and Williams [13] found that on the most extreme comparison, the range of corresponding minutiae ranged from 13 to 54. One group summarized such findings in the following way:

> The extensive variability means we must treat any individual examiner's minutia [sic] 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." [9,p. 7].

This variability can have important implications for examiners' ultimate decisions. For instance, disagreements about inconclusive and individualization determinations—which carry significant implications—are often associated with disagreements about corresponding minutiae [9].

Across studies, it is common for latent-exemplar pairs to receive different evaluative conclusions from different examiners, with many pairs receiving three different conclusions (of either three or four possible conclusions; [10,13,14]). For example, 8 of the 12 pairs used by Langenburg et al. [14] and 13 of the 15 pairs used by Neumann et al. [10] received all three available conclusions: identification, exclusion, and inconclusive. Ulery et al. [11] found that examiners disagreed about 39% of the same source pairs and 20% of the different source pairs used in their study.

Studies of differences in error rates among evaluative decisions (i.e., individualization, inconclusive, and exclusion) also provide possible avenues to understanding examiner differences. Across studies, the erroneous identification rate is consistently much lower (.1–3%) than the erroneous exclusion rate (1–13%) or missed identification rate (9–55%), reflecting the field's stated preference for false negative (i.e., erroneously concluding prints do not match) over false positive (i.e., erroneously concluding prints do match) errors [15]. This preference likely influences examiners' determinations in ambiguous cases, and may lead examiners to reach an evaluative decision of inconclusive in an effort to avoid making a more serious error [16].

3. Verification

During the verification phase, a second latent print examiner scrutinizes the latent print conclusion by performing another analysis, comparison, and evaluation. Thus, verification is an opportunity, for an ostensibly independent examiner, to either corroborate the first examiner's conclusions or detect errors.

In one of the only studies to investigate verification, Langenburg [17] found that verifiers agreed with the original analysts on 94% of trials. Disagreements (on the remaining 6% of trials) involved either identifications versus inconclusive (or vice versa) or no value versus inconclusive (or vice versa) decisions. The "consensus opinion" (reached after discussion between the original analyst and verifier) was split almost equally between decisions of the original analyst (45%) and verifier (55%). Half of the decisions were