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

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).

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

ultimately reported as inconclusive while half resulted in a definitive conclusion of identification or no value. Regarding the ability of the verification process to detect errors, results indicated that although verifying examiners prevented all false positive (erroneous identifications) from being reported, they failed to catch false negatives (erroneous exclusions). Moreover, the number of erroneous exclusions increased when the examiner was aware that verification would occur.

Given the importance of verification in broader error management and quality assurance processes, laboratories and policy-makers have begun to focus more attention in recent years on this stage. The PCAST report noted that verification procedures vary widely. Thus, policy recommendations included suggestions that additional categories of prints be subject to verification, and the suggestion that verification be enhanced to provide for blind verification, in which the verifier does not know what conclusions the first examiner had reached. Those policy and procedure questions can be informed by a pressing set of unexamined questions about how verification is used in practice. They include: In what types of cases is verification conducted in the field? How often does verification lead to changed results or new information? What transpires during any disagreement that results from verification?

4. Current study

Prior research has been valuable in providing examples of examiner reliability in a controlled environment, typically where the ground truth is known. But the field has even less data regarding "field reliability," or examiner agreement in routine practice. Indeed, we know of no published research detailing results from a functioning latent print section and the outcomes of routine casework. Therefore, this study describes case processing data following the ACE-V process, with a particular emphasis on describing verification, consultation, and conflict outcomes as they are performed in vivo at a large metropolitan lab: the Houston Forensic Science Center (HFSC).

The HFSC is a local government corporation that provides the City of Houston and other local agencies with forensic services. HFSC assumed control of the latent print section, which was previously controlled by the Houston Police Department and an independent contractor, on July 1, 2014. The section was accredited by ANSI-ASQ National Accreditation Board (ANAB) in November of 2015 by implementing and following a set of standard operating procedures.

Standard operating procedures dictate operations of the section and prescribe the ACE-V methodology. However, the requirements for the verification stage of the process have changed throughout the two years of the section's existence. The section has implemented a quantitative standard requiring eight minutiae to be present for a latent print impression to be deemed of value; however, the examiner has the discretion to determine that the latent impression is of no value if the quality of the impression is low and diminishes the examiner's confidence in observed minutia. Following the guidelines of the Scientific Working Group on Friction Ridge Analysis, Study and Technology (2013), the section requires all identifications to be verified by a second qualified examiner; however, the decision to verify identifications only was largely depended on the type of offense, with verification more often occurring on more serious offenses. The section settled on a policy of full verification of comparison conclusions, yet no value determinations remain unverified. Finally, following accreditation guidelines, the section follows a standard operating procedure for conflict resolution.

As a subjective decision making process, latent print examination inherently lends itself to consultation and conflict. *Consultation* is defined by HFSC as any significant interaction between examiners such that the consulting examiner conducts his or her own analysis or comparison; *conflict* occurs when each examiner arrives at a different conclusion that could not be resolved by consultation between the two. For each of these instances, the notes must reflect the latent impression conclusion that is being contested by the analyst and the outcome of the discussion between the two examiners.

If the two cannot agree, the latent impression conclusion is elevated to *conflict resolution,* which brings the supervisor into the examination. Each examiner involved must provide charted copies of the latent image that details what information was used to reach each examiner's respective conclusion. The supervisor may then perform his/her own analysis or send unmarked copies of the latent and exemplar to the section for a consensus agreement. Cases are reassigned when the supervisor, or consensus, is either identification when the case examiner concluded exclusion or when the case examiner concluded there was an identification and the ultimate conclusion was deemed to be exclusion.

5. Method

5.1. Participants

For this research, casework from twelve International Association for Identification certified latent print examiners, with experience ranging from six to twenty-eight years, was collected and coded. Seven examiners were men and five were women.

5.2. Procedures

Data from 2535 completed latent print cases was collected from HFSC laboratory information management system (LIMS). The survey focused on latent print section personnel responsible for reporting comparison results over a two-year period from July 1, 2014 to June 30, 2016.

6. Results

The dataset represented 2535 cases and a total of 12,363 latent prints submitted for review. The cases were completed by 12 examiners, though 10 examiners were responsible for nearly all

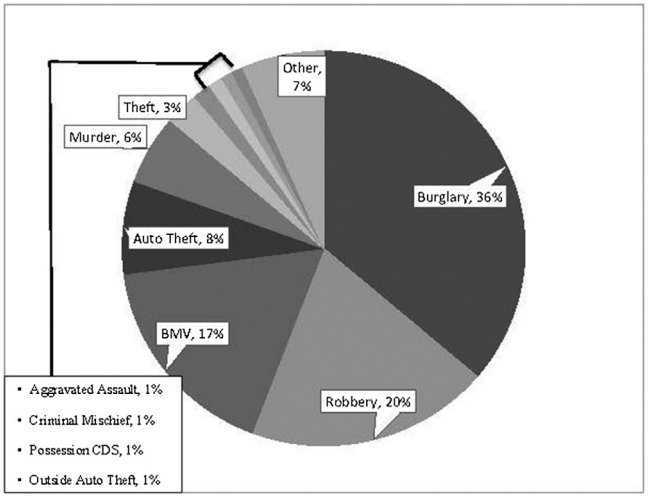

Fig. 1. HFSC caseload by offense type. The pie chart reflects the percentages of different offenses reflected in HFSC's latent print cases from 2014 to 2016. "BMV" refers to burglary of a motor vehicle, and "CDS" refers to possession of a controlled substance.

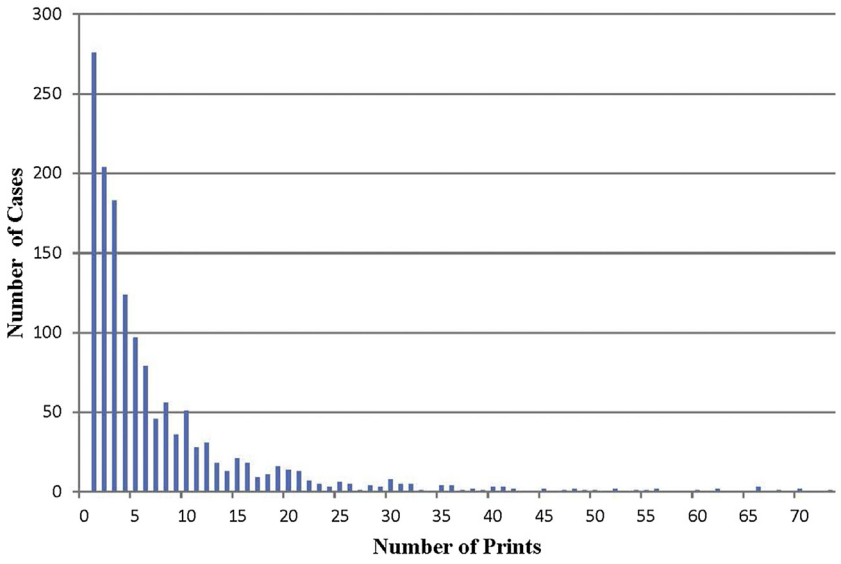

Fig. 2. Distribution of latent prints across cases. The graph above depicts how prints were distributed across cases from 2014 to 2016. Although the number of latent prints in a given case ranged from 1 to 153, for the sake of space we have only shown up to 73 prints, which accounts for 99% of cases. There were 15 cases with more than 73 prints.

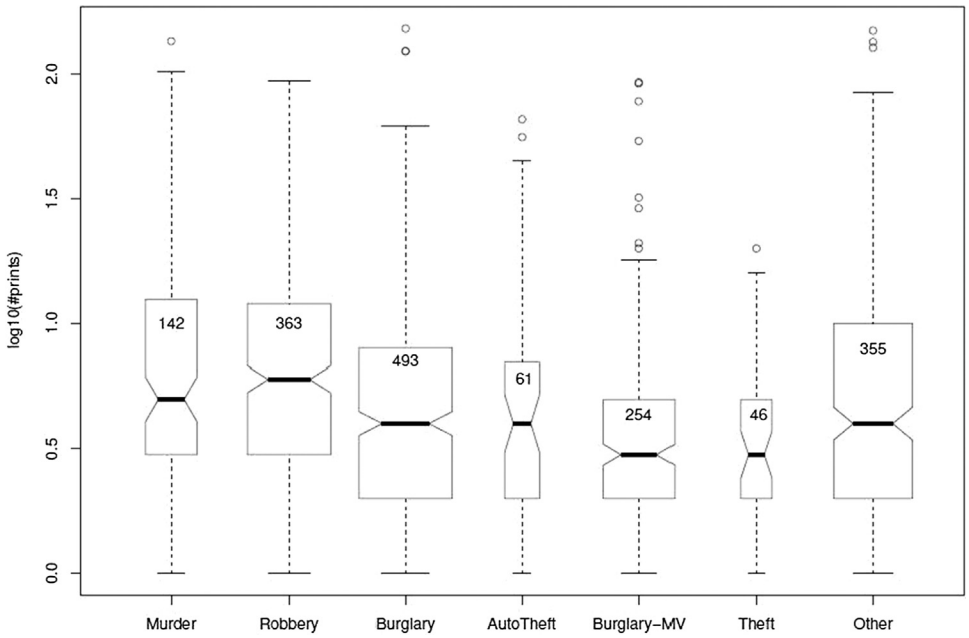

Fig. 3. Variability in prints across different types of offenses. The figure above is a box plot representing the variability in number of prints by offense type. The number in each box represents the number of cases in each offense category.

(98%) of the work (one left shortly into the data collection period, and a second was a section manager who completed little direct casework). At the start of the data collection period, examiners had between 6 and 28 years of experience ($M = 15.2$, $SD = 8.5$).

Offenses ranged from low level crimes such as trespassing to serious offenses such as rape and capital murder. The majority of HFSC cases, though, were burglary and robbery, as shown in Fig. 1.

Latent prints were unevenly distributed across cases (see Fig. 2); the number of prints per case ranged from 1 to 153 ($M = 8.5$, $Mdn = 4$). The mode, however, was one latent print per case. The number of latent prints submitted in a given case also varied by offense type (see Fig. 3). For example, more latent prints were submitted in murder cases than most other offense types (e.g., burglary, theft). Based on these data alone, however, we do not know whether this reflects the nature of the offense (i.e., more relevant latent print are available in murder cases), crime scene practices (i.e., more latent prints are collected in murder cases), or some combination of the two.

6.1. Analysis, comparison, and evaluation

Of the 12,363 latent prints submitted for review, slightly less than half (44%, n = 5430) were determined to be of value. These prints were distributed over 1047 cases (49% serious crimes like rape, murder and robbery; 51% volume crimes like burglary and theft). Of the latent prints of value, nearly three quarters (75%) were fingerprints, and the remaining prints were palm prints (26%), joint impressions (1%), and unknown anatomical source, orientation or other impressions (2%).

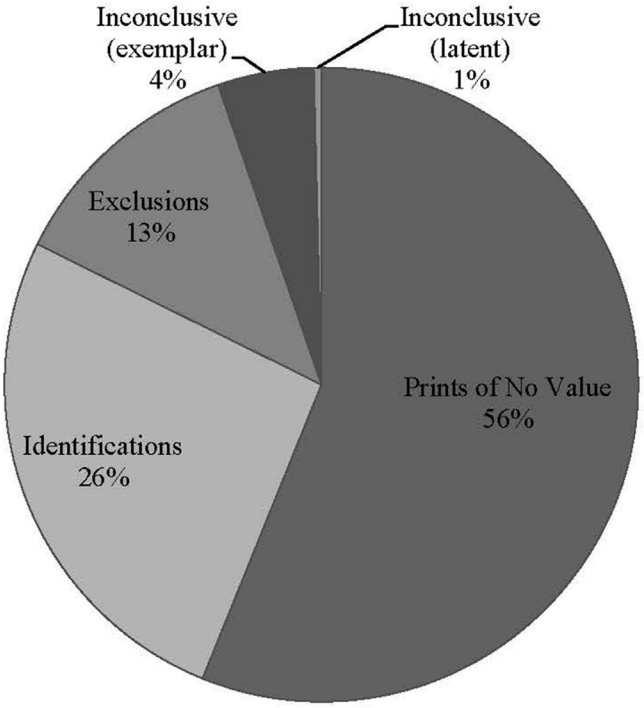

Fig. 4. Outcomes in latent print cases. The pie chart reflects the percentages of print comparisons resulting in each type of conclusion.

Subsequent analyses identified the relative proportion of each type of evaluative decision (i.e., identification, exclusion, and inconclusive) across latent prints of value. Sixty percent of latent print comparisons led to identifications (n = 3234), followed by exclusions (28%, n = 1532), inconclusive due to the exemplar (11%, n = 606) and inconclusive due to the latent (1%, n = 48).[1] Fig. 4 reflects these percentages in the context of all latent prints submitted for review.

6.2. Verification

Of the 2535 cases, 56% (n = 1418) received no verification, 36% (n = 913) received "full" verification (i.e., all conclusions in the case were verified), and 8% (n = 204) had only identifications verified. In all, 100% (n = 3232) of identification conclusions, 53% (n = 825) of exclusion conclusions, 70% (n = 426) inconclusive due to the exemplar and 79% (n = 38) inconclusive due to the latent print conclusions were verified.

Examiners were assigned to verify one another's work by section supervisors according to factors such as schedule and workload. As depicted in Fig. 5, some examiners had others verify their work at relatively equivalent rates (e.g., Examiner A). Others, however, had some examiners verify their work at much higher rates than other examiners (e.g., Examiner B).

6.3. Consultation and conflict resolution

Only 3% of cases (n = 82) had a documented consultation. Several of the cases in consultation involved multiple prints; a total of 132 latent prints were represented in consultation cases. The results of an independent samples t-test indicated that the cases that proceeded to consultation had, on average, a significantly greater number of latent prints than the cases that did not ($M = 21.87$ versus $M = 12.16$ respectively; $t(719, 1) = -4.04$, $p < .01$). Although consultations are not limited to verified cases and do not necessarily follow verification (they can happen at any point in the ACE process), in this sample all of the cases in consultation first proceeded through verification. The 82 cases in consultation represent 7% of verified cases—the cases in which examiners and verifiers disagreed. Thus, in this sample, examiners and verifiers agreed on the decisions in 93% of verified cases.

The results of consultation can be seen in Table 1. The top of the table contains decisions about the analysis stage, i.e., decisions about whether a latent print is of value for comparison. Results indicate that the modal (most common) outcome was that latent prints determined to be of no value at the analysis stage were determined to be of value during the consultation process (n = 18 prints). However, prints of value were determined to be of no value nearly the same number of times (n = 14 prints).

Consultation outcomes regarding evaluative decisions are also reported in Table 1. The modal outcome was an exclusion changed to an identification (n = 22 prints), followed by an exclusion changed to an inconclusive (n = 16 prints). The next three most frequent consultation decisions concerned the threshold between identification and inconclusive, followed by inconclusives changed to exclusions. There were no instances in which an identification was changed to an exclusion. Notably, for the majority of latent prints that proceeded to consultation in this sample, the consultation process produced a change in the analytic or evaluative conclusion.

Conflict resolution—the phase following verification or consultation if the examiners did not reach the same conclusion—was rare: only 8 cases (less than 1% of the overall sample and less than 1% of all verified cases), comprising a total of 10 latent prints, reached this stage. All cases in conflict resolution had been through the consultation process; thus, just fewer than 10% of cases seen in consultation proceeded to conflict resolution. The results of conflict resolution can be seen in Table 2. As shown in this table, each type of outcome occurred, at most, two times. Although conflict resolution sometimes led to a change from the original examiner's conclusion, it never produced a result that differed from the two examiners' conclusions in the consultation stage.

In addition to decisions about the latent prints themselves, we also evaluated the effects of examiners' roles (i.e., case examiner or verifier) and seniority. A chi square test for independence was used to determine whether there was a relationship between the seniority of the verifier relative to the case examiner and whether a case proceeded to consultation. Results indicated that there was not a significant difference between the proportion of cases that proceeded to consultation when the verifier was senior to the case examiner and the proportion of cases that proceeded to consultation when the verifier was junior to the case examiner, $\chi(1) = 1.26$, $p = .26$. In other words, seniority of the verifier did not influence whether a case proceeded to consultation. In consultation, the final decisions were those of the verifier—as opposed to the original case examiner—nearly three quarters (72%) of the time. A chi square test for independence was used to determine whether there was a relationship between the seniority of the verifier and the final decisions in consultation. Results indicated no significant differences between the proportion of final decisions belonging to a senior verifier and those belonging to a junior verifier, $\chi(1) = .09$, $p = .77$. Thus, the pattern of final decisions in consultation being those of the verifier, depicted in Fig. 6, held regardless of seniority.

We also looked at consultation data using the designations employed by the HFSC unit of latent print examiners (LPE), those with fewer than 10 years of experience, and senior latent print examiners (SLPE), those with 10 or more years of experience. Based on these dichotomous groups, slightly less than half of the case

[1] HFSC practice is to attribute inconclusive determinations to either the exemplar (e.g., record prints are sometimes of poor quality) or the latent.

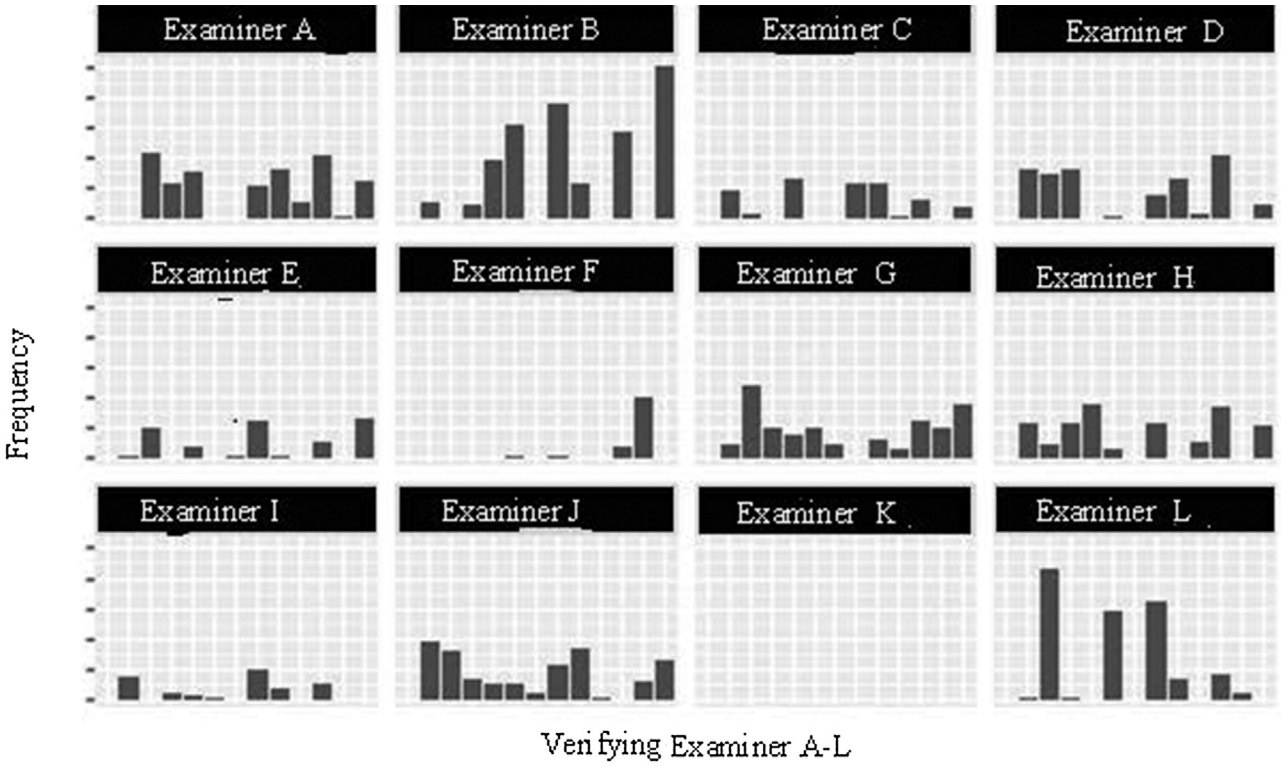

Fig. 5. Assignment of verifying examiners. The graphs show the frequency with which each examiner was assigned to verify one another's work. Each examiner's graph depicts how often his or her work was verified by the other examiners in the lab.

Table 1
Consultation results.

Case examiner's decision	Verifying examiner's decision	Conclusion changed during consultation (Y/N)	Frequency n (%)
Analysis decisions			
No value	**Value**	Y	18 (14%)
Value	**No value**	Y	14 (11%)
Value	No value	N	8 (6%)
No value	Value	N	3 (2%)
Evaluation decisions			
Exclusion	**Identification**	Y	22 (17%)
Exclusion	**Inconclusive**	Y	15 (11%)
Identification	**Inconclusive**	Y	14 (11%)
Identification	Inconclusive	N	13 (10%)
Inconclusive	**Identification**	Y	12 (9%)
Inconclusive	**Exclusion**	Y	10 (8%)
Inconclusive	Identification	N	2 (2%)
Exclusion	Inconclusive	N	1 (1%)

Note. The final conclusion is represented in **bold** type.

Table 2
Conflict resolution results.

Case examiner's decision	Verifying examiner's decision	Conclusion changed during consultation (Y/N)	Frequency n (%)
Analysis decisions			
Value	**No value**	Y	2 (20%)
No value	Value	N	2 (20%)
Evaluation decisions			
Inconclusive	**Identification**	Y	2 (20%)
Identification	**Inconclusive**	Y	1 (10%)
Identification	Inconclusive	N	1 (10%)
Inconclusive	Identification	N	1 (10%)
Exclusion	Inconclusive	N	1 (10%)

Note. The final conclusion is represented in **bold** type.

examiner-verifier dyads were peers (42%; 24% LPE dyads and 18% SLPE dyads). The remaining 58% of cases involved an LPE and SLPE. In over half of these cases (61%, or 36% of all cases in consultation), an SLPE verified the work of an LPE, and in 38% of these cases (or 22% of all cases in consultation), an LPE verified the work of an SLPE (see Fig. 7). Again, seniority did not influence the overall pattern of verifiers determining the final decision. Even among the 19 cases in which an LPE verified the work of an SLPE, the verifier determined the final decision 63% of the time.

A different pattern was observed in conflict resolution. Regarding final outcomes, the decisions were evenly distributed between those of the original case examiner (50%) and those of the verifier (50%) (see Table 2). Here, peer dyads were much less common (10%), and it was much more common for an LPE to verify the work of an SLPE (68%).

6.4. Examiner differences

Examiners differed in the percentage of cases that were verified from approximately one third (36%) to slightly less than two thirds (63%) of their caseload in this period ($M = 47\%$). Examiners had between 2% and 18% ($M = 8\%$) of their verified cases proceed to consultation. Once in the consultation process, examiners had 17–100% ($M = 69\%$) of their initial decisions changed. For some examiners at the higher end of the range, these percentages reflected relatively few contested decisions (e.g., three or four); one examiner, though, had 22 out of 23 decisions in consultation overturned. Finally, examiners tended not to have any cases proceed to conflict resolution (mode = 0, range 0–2).

7. Discussion

This study, the first to present field data concerning latent fingerprint processing, including verification, consultation, and

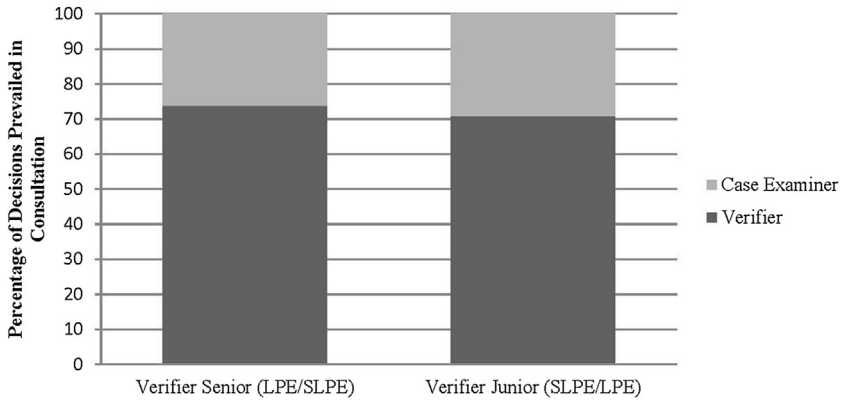

Fig. 6. Consultation decisions and seniority. The graph shows the percentage of post-consultation decisions belonging to either the case examiner or verifier. The bar on the left depicts cases in which a senior examiner verified the work of a more junior examiner. The bar on the right depicts cases in which a senior examiner's work was verified by a more junior examiner.

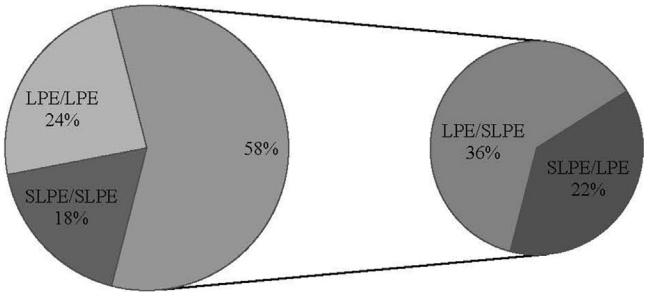

Fig. 7. Consultation cases and seniority. The charts show the percentage of cases in consultation according to the examiner-verifier dyad.

conflict resolution, describes first the flow of cases in a large crime laboratories latent print examination unit. Cases vary greatly in the number of latent prints per case and therefore the work required to analyze the evidence in each case. The caseload also varied by crime, with burglary and robbery comprising the largest portion of the laboratory caseload.

In the analysis stage, over half of latent prints were considered of no value for comparison and evaluation. In the other half of the caseload, the prints were compared, with more evaluations resulting in identification than exclusion or inconclusive. Approximately half of the cases during this period were verified. Among the cases that received verification, case examiners and verifiers agreed 93% of the time; the 7% of cases in which case examiners and verifiers disagreed proceeded to consultation. The most frequent types of disagreements between examiners are consistent with extant research documenting inter-examiner variability at the analysis stage (e.g., [5,10–12]). Although the use of real case data—in which ground truth is not known—precludes a substantive analysis of error rates, the consultation data showing disagreements about evaluative conclusions allows for some informed speculation about error. The relatively high frequency of changes from exclusion to identification, and the absence of *any* changes from identification to exclusion, are consistent with literature demonstrating a much higher rate of false negative than false positive errors among latent print examiners [15].

Although consultation was a rare occurrence in this sample (82 cases and 132 prints, or 3% of the total caseload), the data reveal how verification and consultation can lead to meaningful changes in analytic and evaluative conclusions. For instance, the two most frequent outcomes of consultation were an exclusion being changed to an identification and a print of no value being changed to a print of value. Given that half of this unit's caseload was determinations that a print is not suitable for comparison, future research and training might focus on better understanding and standardizing the initial value determination. Further, expanding the use of verification would potentially yield more prints considered useful for investigations. Overall, these results support investment in verification procedures, and suggest that consultation is an important stage in the fingerprint examination process.

Despite the fact that many initial decisions were changed as a result of the consultation process, these changes were not associated with years of experience or seniority. Thus, despite the possibility that examiners might be inclined to defer to their more senior colleagues when a difference of opinion emerges in verification, the data suggest otherwise.

Given that consultation cases were divided across several types of disagreements (e.g., value/no value, exclusion/identification, identification/inconclusive) and consultation outcomes were not predicted by experience or seniority, future research should explore other variables that might contribute to better understanding of examiner disagreement and how disagreements are resolved, particularly as they relate to individual differences among examiners. It is possible, for example, that personality variables—particularly those related to confrontation and conflict—might predict how examiners interact and reach decisions during consultation and conflict resolution processes. Psychological research suggests that certain personality traits (e.g., agreeableness, extraversion) predict general conflict resolution styles (e.g., collaborating versus competing) [18]. This work could be extended to the specific context of latent print examination and conflict resolution to determine if and when personality traits matter. Additional areas for future inquiry include other examiner variables more particular to latent print examination, such as the effects of personal decision thresholds along a conservative-liberal spectrum (e.g., do examiners differ in terms of how much they tolerate the risk of false positives versus the risk of false negatives?).

Important differences among examiners in this sample emerged, particularly with respect to the proportion of each examiner's caseload that goes to consultation, and the proportion of examiners' decisions that are overturned during this process. For example, at least one examiner appeared highly likely to have conclusions changed upon review. Future research should explore the processes that result in these individual differences, potentially by incorporating metrics of print quality and difficulty of the comparison (e.g., based on distortion, rotation, etc.).

We hope that these results show the value of examining case processing data at crime laboratories. We emphasize, however, that the primary value of these results is probably not the actual

figures they present—for example, the 7% rate of consultation—but rather the process they present: i.e., increased transparency and dissemination of routine case processing information. Indeed, the primary numerical findings will become more meaningful only as other labs begin to share their findings or as HFSC shares updated data. For example, it is possible that labs using different procedures may find meaningfully different agreement rates among examiners. Findings like this would provide better insight into the sensitivity of different procedures to detect examiner differences (which certainly exist) and provide fodder for future research. Basic case processing information can also provide a baseline for comparison if labs change procedures (e.g., from selective verification to 100% verification, or from non-blind to blind verification). Case processing information offers objective data about how such changes affect case processing efficiency, examiner agreement, and the overall reliability of the discipline.

In addition to addressing these foundational questions, future in-house case processing research can also monitor and evaluate exposure to potentially biasing or task-irrelevant information such as the nature of the offense, victim or suspect information, case details (e.g., suspect's confession). Gathering these data allow for subsequent analyses of whether exposure to these types of details influence case processing procedures (e.g., the likelihood of analyst disagreement) or outcomes (e.g., the likelihood of an exclusion or identification). In sum, we consider these results a first step in case processing research that will continue and expand, leading to more transparency, better efficiency, and greater consistency across disciplines.

Acknowledgments

Garrett, Kelley, and Murrie are with the University of Virginia site of the Center for Statistics and Applications in Forensic Evidence (CSAFE) sponsored by the National Institute of Standards and Technology (NIST).

References

[1] 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 fundamentally valid subjective methodology.
[2] M. Triplett, L. Cooney, The etiology of ACE-V and its proper use: an exploration of the relationships between ACE-V and the scientific method of hypothesis testing, J. Forensic Identif. 56 (3) (2006) 345–355. Retrieved from: http://www.fprints.nwlean.net/JFI.pdf.
[3] J. Vanderkolk, Examination process, The Fingerprint Sourcebook National Institute of Justice, (2011) . (Chapter 9), Retrieved from: https://www.ncjrs.gov/pdffiles1/nij/225320.pdf.
[4] G. Langenburg, A Critical Analysis and Study of the ACE-V Process Retrieved from, Université de Lausanne, 2012. http://www.unil.ch/esc/files/live/sites/esc/files/shared/Langenburg_Thesis_Critical_Analysis_of_ACE-V_2012. pdf.
[5] B. Ulery, R. Hicklin, G. Kiebuzinski, M. Roberts, J. Buscaglia, Understanding the sufficiency of information for fingerprint value determinations, Forensic Sci. Int. 26 (1) (2013) 106–117. Available file:///C:/Users/HFS0016/Downloads/ValueDecisions_FSI-Preprint_Final.pdf.
[6] B. Ulery, R. Hicklin, M. Roberts, J. Buscaglia, Changes in latent fingerprint examiners' markup between analysis and comparison, Forensic Sci. Int. 247 (2015) 54–61.
[7] G. Langenburg, Pilot study: a statistical analysis of the ACE-V methodology—analysis stage, J. Forensic Identif. 54 (1) (2004) 64–79.
[8] G. Langenburg, C. Champod, The GYRO system—a recommended approach to more transparent documentation, J. Forensic Identif. 61 (4) (2011) 373–384.
[9] B. Ulery, R. Hicklin, G. Kiebuzinski, M. Roberts, J. Buscaglia, Measuring what latent print examiners consider sufficient information for individualization determinations, PLoS One 9 (2014) 1–16.
[10] C. Neumann, C. Champod, M. Yoo, T. Genessay, G. Langenburg, Improving the understanding and the reliability of the concept of "Sufficiency" in friction ridge examination, national institute of justice, gNCJ (2013) 244231. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/244231.pdf.
[11] B. Ulery, R. Hicklin, J. Buscaglia, M. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, Proc. Natl. Acad. Sci. U. S. A. 108 (19) (2011) 7733–7738.
[12] B. Ulery, R. Hicklin, J. Buscaglia, M. Roberts, Repeatability and reproducibility of decisions by latent fingerprint experts, PLoS One 7 (3) (2012). Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0032800.
[13] W. Evett, R. Williams, Review of the Sixteen Points Fingerprint Standard in England and Wales, J. Forensic Identif. 46 (1) (2018) 49–73.
[14] G. Langenburg, C. Chapod, T. Genessay, Informing the judgments of fingerprint analysts using quality metric and statistical assessment tools, Forensic Sci. Int. 219 (2012) 183–198.
[15] R.N. Haber, L. Haber, Experimental results of fingerprint comparison validity and reliability: a review and critical analysis, Sci. Justice 54 (2014) 375–389.
[16] E. Ray, P. Dechant, Sufficiency and standards for exclusion decisions, J. Forensic Identif. 63 (6) (2013) 675–697.
[17] G. Langenburg, 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, J Forensic Identif. 59 (2) (2009) 219–256.
[18] V.F. Wood, P.A. Bell, Predicting interpersonal conflict resolution styles from personality characteristics, Pers. Individ. Dif. 45 (2008) 126–131.

Further reading

[19] G. Langenburg, C. Champod, P. Wertheim. Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons J. Forensic Sci. 2009; 54: (3) 571-582.
[20] H. Swofford, S. Steffan, G. Warner, C. Bridge, J. Salyards. Impact of minutiae quantity on the behavior and performance of latent print examiners J. Forensic Identif. 2013; 63: (5) 571-591.
[21] Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint), SWGFAST 2013; Retrieved from http://www.swgfast.org/documents/examinations-conclusions/130427_Examinations-Conclusions_2.1.pdf.
[22] J. Tangen, M. Thompson, D. McCarthy. Identifying fingerprint expertise Psychol. Sci. 2011; 22: (8) 995-997.
[23] M. Thompson, J. Tangen. The nature of expertise in fingerprint matching: experts can do a lot with a little PLoS One 2014; Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4269392/.
[24] M. Thompson, J. Tangen, D. McCarthy. Expertise in fingerprint identification J. Forensic Sci. 2013; 58: (6) 1519-1530.