Science and Justice 55 (2015) 239–247

Contents lists available at ScienceDirect

Science and Justice

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

Fingermark submission decision-making within a UK fingerprint laboratory: Do experts get the marks that they need?

Helen Earwaker [a,*], Ruth M. Morgan [b], Adam J.L. Harris [c], Lisa J. Hall [d]

[a] Department of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, United Kingdom
[b] Centre for the Forensic Sciences, University College London, 35 Tavistock Square, London WC1H 9EZ, United Kingdom
[c] Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom
[d] Directorate of Forensic Sciences, Metropolitan Police, New Scotland Yard, London SW1H 0BG, United Kingdom

ARTICLE INFO

Article history:
Received 21 August 2014
Received in revised form 15 January 2015
Accepted 29 January 2015

Keywords:
Fingerprint development
Fingermark sufficiency
Decision-making
Contextual information

ABSTRACT

Within UK policing it is routinely the responsibility of fingerprint laboratory practitioners to chemically develop areas of latent fingerprint ridge detail on evidential items and to determine which areas of ridge detail are of sufficient quality to be submitted to fingerprint experts for search or comparison against persons of interest. This study assessed the effectiveness of the fingermark submission process within the Evidence Recovery Unit Fingerprint Laboratory of the Metropolitan Police Service. Laboratory practitioners were presented with known source fingermark images previously deemed identifiable or insufficient by fingerprint experts, and were asked to state which of the marks they would forward to the Fingerprint Bureau. The results indicated that practitioners forwarded a higher percentage of insufficient fingermarks than is acceptable according to current laboratory guidelines, and discarded a number of marks that were of sufficient quality for analysis. Practitioners forwarded more insufficient fingermarks when considering fingermarks thought to be related to a murder and discarded more sufficient fingermarks when considering those thought to be related to a crime of 'theft from vehicle'. The results highlight the need for fingerprint laboratories to work alongside fingerprint experts to ensure that a consistent approach to decision-making is, as far as possible, achieved, and that appropriate thresholds are adopted so as to prevent the loss of valuable evidence and improve the efficiency of the fingerprint filtering process.

1. Introduction

Items of forensic evidence that are recovered from scenes of crime are often submitted to a police force in-house fingerprint recovery laboratory where physical and chemical techniques are utilised to visualise latent fingermarks [1]. It is then commonly the responsibility of a fingerprint laboratory practitioner to determine which of the areas of developed friction ridge detail are of sufficient quality to be submitted to fingerprint experts for search or comparison against persons of interest. This interpretative process undertaken by laboratory practitioners is often overlooked, as it is the subsequent comparison and identification carried out by fingerprint experts that provides the end product of the fingerprint process in the form of a result which is fed back to the investigative team and the court. The importance of the decision-making that occurs in the laboratory prior to involvement of an expert is seldom recognised. The potential consequences of an incorrect practitioner fingermark submission decision in live casework, however, are severe; forwarding a poor quality fingermark will waste police force resources in the form of materials and the time of fingerprint experts, whilst discarding a good quality fingermark will lead to the potential loss of identifying evidence which, at the extreme, could lead to a guilty person not being detained or convicted. It is, therefore, crucial that the fingerprint filtering process conducted by fingerprint laboratory practitioners is closely aligned to the quality assessment of fingerprint experts and that this decision of fingerprint quality is made accurately and consistently by both parties.

In the United Kingdom, fingerprint laboratory practitioners are either trained in-house by police forces or externally by the College of Policing, with the primary focus of this training on selecting and carrying out chemical fingermark development techniques, rather than the recognition, quality assessment, and analysis of friction ridge detail. In contrast fingerprint experts carry out a five year training programme that focuses on the ACE-V process: the Analysis, Comparison, Evaluation and Verification of fingermarks. It is therefore crucial to establish the extent to which laboratory practitioners are able to distinguish fingermarks sufficient for comparison from insufficient ones.

In recent years forensic science practice has been scrutinised for not having a sufficient science grounding [2–4]. In the UK the appointment

* Corresponding author. Tel.: +44 7793225177.
E-mail addresses: h.earwaker.12@ucl.ac.uk (H. Earwaker), ruth.morgan@ucl.ac.uk (R.M. Morgan), adam.harris@ucl.ac.uk (A.J.L. Harris), Lisa.Hall@met.police.uk (L.J. Hall).

http://dx.doi.org/10.1016/j.scijus.2015.01.007
1355-0306/

of the Forensic Science Regulator, tasked with ensuring that the provision of forensic science services is subject to appropriate scientific quality standards [5], has paved the way for change. Forensic laboratories must now achieve accreditation to the international standard ISO17025 [6]. The regulator describes the four areas of assessment that accreditation targets as: organisational competence, individual competence, the validity of methodology, and objectivity and impartiality [5, p. 33].

Fingerprint recovery laboratories must now validate the chemical development techniques that they employ and this internal validation must be assessed by UKAS in order to fulfil the requirements of ISO17025 accreditation. However, the decision to submit an area of friction ridge detail to a fingerprint examiner or to discard the detail is not rigorously assessed for the purposes of ISO17025 accreditation. With or without an accredited procedure this decision is difficult to fully control, even if the procedure followed and documented by the organisation has been validated, as it relies upon on a subjective personal decision.

Laboratory practitioners thus receive little training on how to make these subjective decisions. Schiffer and Champod highlighted the importance of ACE-V training in the 'analysis' stage of fingerprint examination [7], which involves establishing the suitability of a developed latent fingermark for further comparison through identifying the characteristics present in the ridge detail [8]. Their study involved two groups of student participants: one group of students who had attended a full forensic identification course (including training in the ACE-V process) and another who had not received this training. Participants were asked to determine the number and type of minutiae present in a series of fingermarks and to determine whether each fingermark was sufficient for comparison. It was found that the group of participants who had been trained in the ACE-V process were able to identify a higher number of fingermark characteristics, with less variability between participants, than the group that had not received training [7]. Equally, the number of fingermarks deemed exploitable by the trained group was found to be almost double that of the non-trained group. This may suggest that fingerprint development practitioners, not primarily trained in the 'Analysis' stage of ACE-V may be likely to identify fewer characteristics within the ridge detail of a fingermark when making a decision as to the suitability of the mark to be submitted to an expert and may be less likely to decide that a fingermark is sufficient for comparison, than the expert themselves during the analysis process. This being the case it would seem that the sufficiency threshold of a fingerprint development practitioner may indeed be higher than that of a fingerprint expert due to their lesser capability to recognise ridge pattern and identify characteristics, potentially leading to the loss of evidentially useful fingermarks through incorrectly discarding them. This study by Schiffer and Champod also suggests that there may be more variability between the sufficiency thresholds of fingerprint laboratory practitioners than between trained fingerprint experts [7], suggesting a lack of consistency in the quality of marks that are submitted to the Bureau.

Neumann et al. [9] examined the evidential value of a sample of the fingermarks that were discarded by a US fingerprint laboratory practitioner and scene of crime officers during the process of mark recovery from evidential items to expert comparison. The crime type of casework included within the study was unknown, however typically the laboratory dealt with a majority of volume crime cases (66% of annual workload was reported as relating to property crimes). Progressing these discarded marks led to just a 2.3% increase in evidence (additional associations were gained in 38 out of 1619 additionally recovered marks across 17 of 178 cases), at an estimated cost of $138,000, suggesting that only a small number of evidentially useful fingermarks are discarded, and that to obtain additional evidentially useful fingermarks would require disproportionate investment. Results also showed that only a small proportion of insufficient marks were being submitted by the laboratory. This would seem to suggest that fingerprint laboratories may be reasonably successful in their role of submitting fingermarks to experts that are of sufficient quality for comparison and discarding marks of insufficient value, resulting in a system that is fairly efficient. However, Neumann et al. [9] did discover marks discarded by laboratory staff which contained very large numbers of minutiae, and would have been highly suitable for expert comparison, suggesting a difference between the quality judgements of the experts and the laboratory practitioners.

The decision made by laboratory practitioners as to whether to discard a fingermark or forward it for analysis merits further investigation as Neuman et al. [9] processed only a sample of discarded marks, so did not record the effectiveness of all decisions made, and did not distinguish between fingermarks that were recovered and submitted by scene of crime officers (who also have a role in filtering fingerprint evidence) and fingermarks recovered by the laboratory fingerprint practitioner. There could be seen to be benefit in making a distinction between these two roles as training, working environments, policies and procedures vary considerably between them. Equally the results reported are based on the decisions of one laboratory practitioner and four fingerprint experts in a small US laboratory. Therefore the findings are representative only of the decision-making ability and thresholds of one fingerprint practitioner. Given the amount of variability in the analysis ability of laboratory practitioners [7], similar research with a larger sample of laboratory staff is required in order to be able to generalise the results beyond this laboratory.

The context in which a fingermark is presented has been found to affect decision-making in relation to the mark. Researchers have investigated the effects of contextual information on the judgments of fingerprint experts, but such research has not yet been applied to fingerprint development practitioners. Fingerprint experts have been found to be affected by the context of a comparison print during the 'analysis' processes [10]. They were found to be more likely to make a fingerprint match in an emotional context [11], and inconsistent when making decisions about the same fingermark presented in a different context [12]. Charlton et al. found that the decision thresholds of fingerprint experts were vulnerable to distortion due to the effects of emotion and the need for closure [13] and Hall and Player also report that fingerprint experts perceive that emotional context affects their analysis process, even when no such effect is found [14]. Dror [15,16] states that expertise in a task leads to increased vulnerability to psychological effects such as selective attention, a reliance on top down information, and a vulnerability to confirmation bias through exposure to contextual information, but the extent to which access to contextual information affects the sufficiency decisions made by fingerprint practitioners in the laboratory has not been assessed. It is likely that contextual information will affect the outcome of the sufficiency decisions made by the practitioners. This may occur as a result of a number of psychological effects, including emotional effects which cause a broadening or narrowing of focus [17,18], or due to the effect that the contextual information has upon the utility values that practitioners associate with the outcomes of the decision to submit or discard a fingermark. For example, in the context of a murder case, the perceived negative consequences of discarding a useful fingerprint which could be attributed to the murderer are much greater than the perceived negative consequences of wasting resources by submitting an insufficient quality fingermark to a fingerprint expert, suggesting a reduction in the quality threshold for the submission of a fingermark, whereas the perceived negative consequences of discarding a fingerprint that could identify a suspect in the case of theft from a vehicle may be closer to the perceived negative consequences of wasting resources in this case, resulting in a higher quality threshold for fingermark submission in cases of volume crime. This may result in a difference in practitioner fingerprint submission threshold according to the category of crime.

1.1. Objectives of the present study

Access to the UK fingerprint community for the purpose of experimental research has been traditionally difficult to gain. The present

study is enabled by the operational experience of the first and fourth authors and builds on the work carried out in the US by Neumann et al. [9], to investigate the efficiency of sufficiency decision-making within fingerprint submission, focusing solely on submission decisions made by fingerprint laboratory practitioners in relation to fingermarks of borderline quality for submission to the fingerprint bureau. Fingermarks considered to be borderline for submission to the fingerprint bureau were investigated in this study as previous research has shown decision-making to be more challenging in ambiguous cases [11]. This present study investigated how appropriate fingermark submission decisions made by laboratory practitioners were in relation to the sufficiency threshold of in-force fingerprint experts, in these borderline cases. Through doing so, this study aimed to investigate the efficiency and variability of laboratory practitioner fingermark submission decision-making through the execution of an experimental study that was valid within the UK Metropolitan Police Service Evidence Recovery Unit, Fingerprint Laboratory. It also sought to establish how the category of crime affects the submission threshold of fingermarks by laboratory practitioners, and assessed whether this effect is in line with Metropolitan Police Service procedure.

2. Materials and method

2.1. Development of experimental fingermarks

A series of latent fingermarks of known source were deposited by a single 'poor quality' donor on clean sheets of white A4 paper to provide only known fingermarks for use in the study. These were deposited at a range of pressures and with a range of movement to provide marks of varied appearance, quality and type. Marks were developed at the UK Metropolitan Police Service Evidence Recovery Unit Fingerprint Laboratory, London, with the use of Ninhydrin (a reagent proven by the Home Office Centre for Applied Science and Technology to be effective at the development of fingermarks on smooth porous items [1]), to visualise the amino acid constituents in the deposited latent fingermarks. The same development technique was used to enhance all fingermarks deposited to allow consistency in the general appearance of each fingermark, and Ninhydrin was selected as it is a reagent commonly used as part of a sequence of chemical techniques in both serious and volume casework [1]. The Metropolitan Police Service ISO17025 accredited procedure for carrying out this technique was followed. Each paper was submerged in pre-prepared Ninhydrin Working Solution (Samuel Banner & co. Ltd) and left to dry within a fume cabinet. The papers were then placed in a Ninhydrin Oven (Weiss Galenkamp, calibrated 7/3/13) set at 80 °C, with 65% relative humidity. The door was closed and the oven was allowed to regain its regulated temperature. After two minutes the papers were removed from the oven and placed in a plastic folder to prevent contamination. Papers were left for 6 days to allow any further mark development to occur. Forty areas of developed ridge detail considered by the first author to be borderline in quality for submission to an expert were selected so as to ensure that the study focused on decision-making in relation to more challenging, ambiguous fingermarks. These areas of ridge detail were captured in colour using a Nikon D4 camera at 500 ppi. Adobe Photoshop was used to apply very minor adjustments to contrast and brightness so as to optimise the screen image for maximum quality at print as per Metropolitan Police Service standard mark photography procedures. Images were printed 1:1 on photographic paper at a resolution of 400 dpi using an AGFA D Lab printer.

2.2. Expert comparison of fingermarks

A set of inked fingerprints (tenprints) were taken from the donor of the developed fingermarks (described in Section 2.1) by two Metropolitan Police Service Fingerprint Experts. Three sets of prints were taken from each hand, including tips and sides, to ensure that all friction ridge detail was captured to a good quality. Capturing all ridge detail in this way ensured that the tenprints represented the highest possible quality suspect fingerprints achievable. Prints were recorded under an assumed name to ensure the anonymity of the donor.

Three fingerprint experts assisted with the setup of the study. Two experts were employed by the Metropolitan Police Service Fingerprint Bureau, and one was a registered independent expert. Each expert was given the set of tenprints described and one set of printed mark photographs (described in Section 2.1). They were asked to compare each of the 40 mark photographs against the tenprints and, whilst doing so, to make a general observation of the quality and clarity of the mark, to state whether or not the mark was searchable and comparable, and whether they were able to identify the mark (if so to which finger), or whether it was inconclusive, excluded or insufficient. Each expert worked independently of the others, resulting in three independent expert opinions of the quality of each of the 40 fingermarks.

2.3. Selection of experimental fingermarks

From the 40 expert assessed fingermarks, 20 were selected to form the 'experimental image set'. The opinions of the three experts were consistent in the comparability of the selected marks and whether or not they could be identified to the tenprints provided, so as to provide an experimental set upon which expert opinion was consistent. This resulted in a benchmark standard by which to judge the effectiveness of laboratory practitioner submission decisions. The fingermarks selected were an equal mixture of insufficient and identifiable marks as stated by the experts and were considered to be borderline for submission by the first author, who has three years' experience as a fingerprint laboratory practitioner.[1] An example of an insufficient and an identifiable fingermark from the experimental image set is provided in Fig. 1.

The remaining fingermark images not selected to form the experimental set were divided into two sets of ten 'decoy prints' ('set 1' and 'set 2') to be included among the experimental images in order to make two sets of the same images appear to be different. Agreement by all three experts was not necessary for these images as decisions made in relation to them were not analysed. The experimental image set and decoy set 1 were mixed to form 'serious crime set 1' and the experimental image set was mixed with decoy set 2 to form 'volume crime set 2'. Order and content information for both sets can be found in Table 1.

Each set of fingermark photographs was bound in a hardback A6 notebook, with one image per page, ensuring that images were considered in isolation in the order stated. The reference and the crime type to which the mark related were written above each mark ('theft from vehicle' for the volume context and 'murder' for the serious context). The inclusion of the crime type on each page in this way was intended to ensure that the crime context was considered in each case.

2.4. Participants

Participants were 11 fingerprint practitioners working within the Metropolitan Police Service Evidence Recovery Unit Fingerprint Laboratory. They had an average of 12 years' experience, a minimum experience of approximately two years and a maximum of 38 years. All participants had received Metropolitan Police Service in-house fingerprint development training and were subject to existing quality assurance procedures. Two of the participants had previously been

[1] One possible definition of a 'borderline' fingermark would be 'a fingermark about which there is a lack of consistency in the determination of its usability between fingerprint experts'. Given that the assessment of fingerprint experts is the only relevant standard against which to compare practitioners' decisions, the implementation of this definition would prevent the use of an appropriate analysis standard and so would be inappropriate in the present study. Instead the definition of 'borderline' for the purpose of the present study is 'a fingermark considered to be ambiguous for submission to a fingerprint expert'.

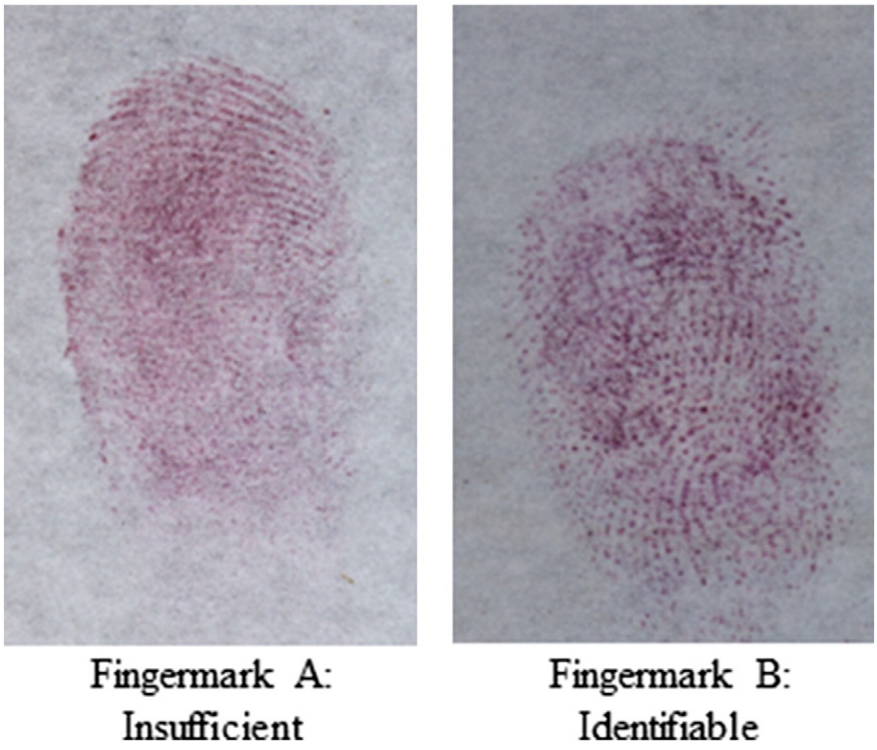

Fig. 1. An example of insufficient and identifiable fingermarks from the experimental image set.

employed as fingerprint experts within the Metropolitan Police Service but neither had practised as a fingerprint expert for a significant number of years or carried out any additional duties as part of their current job role, so were deemed suitable for inclusion within the sample.

2.5. Design

A within-participants design was employed. Each participant was given both the serious and the volume set of mark photographs, one set at time 1 and the second set at time 2 (three weeks later). The order of presentation of the image sets was counterbalanced (participants were assigned to either group A ($N = 5$), who received the 'serious crime' images on time one, or group B ($N = 6$), who received the volume crime images first). This was intended to balance the data for carryover effects of having been exposed to a particular context first. The three week gap between analyses was designed to ensure that the participants were unlikely to recognise similarities within the two sets of fingermarks in order to make the different origins of the fingermarks believable.

2.6. Procedure

Each participant was given the set of fingermark images relating to their assigned first time context along with an instruction sheet, which stated that they should look at each fingermark in turn, in the order presented, and to decide, given the context provided, whether or not they would 'mark-up' (submit to a fingerprint expert) each of the fingermarks. They were also instructed to state their reasons for making each decision. Participants then returned their responses along with the book of fingermark photographs, and a time gap of three weeks was left before the participants were given the photo set relating to the alternative context according to the group to which they were assigned. This meant that each practitioner made a

Table 1
Order and content of experimental image sets.

Contents of experimental image sets			
'Serious crime' image set		'Volume crime' image set	
1: Decoy mark	16: Decoy mark	1: Decoy mark	16: Image J — identifiable
2: Decoy mark	17: Image I — insufficient	2: Decoy mark	17: Image K — identifiable
3: Image A — insufficient	18: Image J — identifiable	3: Image A — insufficient	18: Decoy mark
4: Decoy mark	19: Image K — identifiable	4: Image B — identifiable	19: Image L — identifiable
5: Image B — identifiable	20: Image L — identifiable	5: Decoy mark	20: Decoy mark
6: Image C — insufficient	21: Image M — insufficient	6: Decoy mark	21: Image M — insufficient
7: Image D — insufficient	22: Image N — insufficient	7: Image C — insufficient	22: Decoy mark
8: Image E — identifiable	23: Image O — insufficient	8: Image D — insufficient	23: Image N — insufficient
9: Image F — insufficient	24: Image P — insufficient	9: Image E — identifiable	24: Image O — insufficient
10: Image G — identifiable	25: Decoy mark	10: Image F — insufficient	25: Image P — insufficient
11: Decoy mark	26: Decoy mark	11: Decoy mark	26: Image Q — identifiable
12: Decoy mark	27: Image Q — identifiable	12: Image G — identifiable	27: Image R — identifiable
13: Decoy mark	28: Image R — identifiable	13: Image H — identifiable	28: Image S — identifiable
14: Image H — identifiable	29: Image S — identifiable	14: Image I — insufficient	29: Image T — insufficient
15: Decoy mark	30: Image T — insufficient	15: Decoy mark	30: Decoy mark

submission decision in relation to each of the 20 experimental images twice; once when the fingermark was presented in a volume context and once in a serious crime context. Participants were asked to carry out this task at their normal place of work, under normal working conditions, using any standard practices and equipment, so as to maximise the ecological validity of the study.

2.7. Laboratory policy interview

The Fingerprint Laboratory Manager (who was also one of the 11 practitioner participants) was interviewed in a semi-structured manner after completing the fingermark assessment task in order to gain additional qualitative data concerning business structure, recruitment and training, fingerprint recovery and submission policy, and quality assurance practices. This additional information was collected in order to provide a bench mark for the analysis of decision-making performance data in relation to the current policies and procedures of the Metropolitan Police Evidence Recovery Unit Fingerprint Laboratory. The interview was tape recorded and later transcribed.

3. Results

3.1. Effectiveness of fingerprint laboratory sufficiency decision-making

Overall 11 participating practitioners each made 2 submission decisions on each of the 20 experimental fingermark photographs (one decision upon each mark in relation to both of the crime categories), thus a total of 440 submission decisions were made over the course of the study. Each of these decision outcomes were classified as either an 'agreement' (the decision of the practitioner matching the decision desired by the expert), or an 'erroneous' result, with erroneous results further divided into 'false positives' (the decision to submit a mark that was of insufficient quality for comparison) and 'false negatives' (the decision to discard a mark that was of sufficient quality for comparison).

Overall an error was made in 34% of the decisions made by the laboratory practitioners, with 17% of the decisions being false positives and 17% false negatives. The practitioners forwarded a total of 76 marks that were of insufficient quality for comparison and 73 marks that could have been compared and identified were discarded. 34.1% of the marks that were selected by the practitioners to be submitted to the fingerprint experts were of insufficient quality to be compared, and 33.6% of the marks that were discarded by the practitioners were of good enough quality to have been compared.

3.2. Laboratory manager interview

During the interview with the laboratory manager, it was disclosed that the optimum position would be no false positive and no false negative fingerprint submissions, but acknowledged that this was not feasible at an operational level. Metropolitan Police Evidence Recovery Unit Fingerprint Laboratory procedure was that if false positive fingermarks constituted over 20% of a submission to the Bureau this was picked up through a fingerprint expert feedback mechanism and led to further training. Three additional quality assurance mechanisms were described during the interview: dip sampling, expert feedback, undisclosed testing, and UKAS inspection. A quarterly dip sampling process is carried out through which an exhibit is removed from workflow and re-examined by an internal practitioner to quality assess the fingermark selection carried out. This process assesses fingermark selection but from the view point of another fingerprint practitioner not that of a fingerprint expert. Undisclosed tests are carried out through the placement of mock casework into workflow, but these tests are designed to assess the overall continuity of the evidence recovery process and turnaround times, not fingermark selection. Annual inspection of the Metropolitan Police Evidence Recovery Unit Fingerprint Laboratory is carried out on behalf of UKAS in fulfilment of the requirements of accreditation. This involves UKAS assessors observing practitioners carrying out treatment processes and observing a sampled selection of training records and controlled documents, but does not focus upon the decision to submit or discard a friction ridge detail. There was no routine mechanism by which experts were able to determine that good quality marks had been discarded.

All but two of the participating practitioners submitted a higher than acceptable level of insufficient fingermarks to experts, according to the stated laboratory procedure of allowing 20% of a fingermark submission to be of insufficient quality for comparison. The percentage of discarded marks that could have been identified and forwarded marks that could not have been used to identify are given for each participant in Table 2.

3.3. Effect of contextual information on fingerprint laboratory sufficiency decision-making

The 11 practitioners each made one decision about each of the 20 experimental marks according to each crime context, resulting in a total of 220 decisions relating to each crime context of serious and volume crime. Each of these decision outcomes were classified as either 'agreement', 'false positive' or 'false negative', as previously described. The results from a chi-squared test showed that the relation between crime context and decision outcome was significant ($(x^2,2) = 9.817$, $p < 0.01$), although the Cramer V test statistic was 0.149 indicating a weak relationship between the two variables. In the serious crime category false negatives were underrepresented and false positives were overrepresented, whereas in the volume crime category false negatives were overrepresented and false positives were underrepresented. The frequencies of each decision type made according to the context presented are presented in Table 3.

The submission of more fingermarks in the serious crime context (Table 3) shows that the threshold for submitting a mark was lower in cases of serious crime (recall that the set of marks for each context was identical). It was disclosed during interview, however, that a different threshold for the submission of fingermarks according to crime type was not part of Metropolitan Police Evidence Recovery Unit Fingerprint

Table 2
Percentage of 'false negative' and 'false positive' decision outcomes for each participant.

Participating practitioner	Percentage of marks discarded that were of sufficient quality (False negatives)	Percentage of marks submitted that were of insufficient quality (False positives)	Did false positive submission fall above or below Metropolitan Police Service maximum false positive submission threshold of 20%?
A	45.45 (N = 33)	28.57 (N = 7)	Above maximum threshold
B	30.00 (N = 20)	30.00 (N = 20)	Above maximum threshold
C	25.00 (N = 20)	25.00 (N = 20)	Above maximum threshold
D	20.00 (N = 25)	0.00 (N = 15)	Below maximum threshold
E	35.71 (N = 28)	16.67 (N = 12)	Below maximum threshold
F	37.04 (N = 27)	23.08 (N = 13)	Above maximum threshold
G	38.46 (N = 13)	44.44 (N = 27)	Above maximum threshold
H	45.16 (N = 31)	33.33 (N = 9)	Above maximum threshold
I	20.00 (N = 10)	40.00 (N = 30)	Above maximum threshold
J	11.11 (N = 9)	38.71 (N = 31)	Above maximum threshold
K	0.00 (N = 1)	48.72 (N = 39)	Above maximum threshold

Table 3
Frequencies of each decision type made according to the context presented.

Context provided	Decision frequency (percentage of decisions within crime context)		
	Decisions in agreement	False positive decisions	False negative decisions
Serious crime	139 (63.18%)	50 (22.73%)	31 (14.09%)
Volume crime	152 (69.09%)	26 (11.82%)	42 (19.09%)

Laboratory policy. The policy was that any fingermarks that are of good enough quality for comparison should be submitted in all cases.

4. Discussion

4.1. Effectiveness of fingerprint laboratory sufficiency decision-making

Thirty four percent of fingermark submission decisions made by fingerprint laboratory practitioners during the course of this study were erroneous in relation to the outcomes of the sufficiency decisions made by the fingerprint experts. These erroneous decisions were approximately evenly distributed between fingermarks of too poor quality for comparison that were submitted and fingermarks of good enough quality for comparison that were discarded, resulting in an equal proportion of wasted resources (in terms of subsequent analysis being carried out on marks of insufficient quality) and lost potential evidence (in acceptable marks not being submitted for analysis). It is acknowledged by the fingerprint laboratory that there will be a degree of error when submitting fingermarks and that submitting at a lower than optimum threshold, which creates some false positive submissions, should ensure that good quality marks are not discarded. Within the Metropolitan Police Service Evidence Recovery Unit Fingerprint Laboratory it is acceptable for fingermarks of too poor quality for comparison to make up to 20% of each submission to fingerprint experts, with levels above this flagged by an expert and raised as a training requirement. The results of the present study showed that, in the case of the borderline sample of fingermarks used in this study, the percentage of false positive marks exceeded the acceptable level overall, and that all but two practitioners surpassed this level when deciding to submit marks from the experimental set, suggesting that this level is not often achieved in practice when dealing with borderline fingermarks and that existing mechanisms for detecting and providing training to prevent these errors are not currently addressing this particular issue.

However, the high level of false negatives (33.6% of marks that were discarded could have been identified) is, arguably, the most concerning and significant finding of this study. It is potentially concerning as there is a possibility that this could equate to a loss of evidence and possible detections, and, ultimately, the potential for perpetrators of serious crimes to go unidentified. However, it is also arguably the most concerning type of error as there is not currently a sufficient process in place which is able to identify or prevent it. An existing quality assurance feedback mechanism within the Metropolitan Police Evidence Recovery Unit Fingerprint Laboratory (as stated during the laboratory manager interview) is able to identify higher than desirable levels of false positive mark submission, according to the opinion of the fingerprint expert who has received the marks. The existing dip sampling quality assurance procedure in place within the laboratory is able to identify potential false negative results, but only according to the judgement of another laboratory practitioner and not that of a fingerprint expert, meaning that true detection of false negative decisions is not currently achieved. Given that regular and accurate feedback is often assumed necessary for the development of expertise [19], it seems unlikely that laboratory practitioners would be able to develop expertise in this sufficiency determination task through experience.

There are a number of potential explanations for a difference in decision threshold between the fingerprint experts and the laboratory practitioners. Experts' focused training and experience with fingermark analysis might have led to enhanced pattern recognition, enabling them to identify more minutiae than laboratory practitioners [7,20], or experts may work to a 'winner takes all' decision making threshold [13], meaning the quantity of information required to make a decision about a fingerprint may vary between an expert and practitioners. In the present study it may have been the case that differences arose because fingerprint experts carried out their analysis with reference to a control print in order to reach an implied sufficiency determination through an attempted identification (contrary to Metropolitan Police Service standard fingerprint expert policy of blind analysis of a fingermark prior to comparison) which meant that they judged a fingermark differently from a practitioner viewing the print in isolation [21].

However the results of the present study suggest that there is not a straightforward difference in perceived fingermark sufficiency threshold between the experts and the laboratory practitioners when dealing with borderline fingermarks. Practitioners, in the context of this study, were not simply selecting marks at too high a threshold (missing good marks) or at too low a threshold (sending poor quality marks); they were selecting and discarding the wrong marks in these challenging cases. The lack of a clear cut threshold difference suggests the need to carry out further research to ascertain the differences in laboratory practitioner and fingerprint expert decision-making.

The practitioner decision to submit or discard a fingermark, especially when dealing with those of borderline quality is, by definition, a difficult decision. This decision is all the more difficult since the appropriate evaluation standard is the subjective judgement of the fingerprint expert who receives (or does not receive) the fingermark. The lack of agreement between the fingerprint experts on the sufficiency of some of the fingermarks not included within the final experimental set during the set-up of this study highlights further the subjective nature of fingermark sufficiency decision-making and the individual differences inherent in the decision. Even though a zero error rate would be idealistic, in particular in relation to false negative decisions, the inherent subjectivity of the decision means that an evaluation standard of zero error is unrealistic and, arguably, unattainable. However, it is important to identify the degree of error occurring, particularly in challenging marks, and to attempt to identify procedures to minimise this error as far as is possible.

It should be highlighted that the fingermarks presented in the current study were intended to be ambiguous and borderline for submission to an expert. It could, therefore, have been the case that the poor quality of the fingermarks was, itself, affecting practitioner decision-making causing some practitioners to forward fewer of the fingermarks during the experimental scenario as they felt that they had already submitted a series of poor quality marks and were concerned not to submit too many poor quality fingermarks, whereas others may have felt that it was better to submit rather than discard these borderline quality marks so that the decision was passed to a fingerprint examiner. The intentional use of fingermarks that were borderline for submission in this study does therefore, limit the extent to which the results can be extrapolated to the population of fingermarks typically encountered in casework. Casework typically encompasses both borderline and clear-cut marks, for which one would expect greater agreement between laboratory practitioners and fingerprint experts, increasing the overall percentage of agreement with regard to the sufficiency of the marks. However, whilst the fingermarks used were considered of borderline quality for practitioner submission, all marks had been agreed to be either identifiable or of insufficient quality for identification by two Metropolitan

Police Service fingerprint experts and one external expert. As a result these 'borderline' fingermarks could not be considered to also be borderline from an identification perspective, and as such were only borderline from a laboratory practitioner perspective. This demonstrates, that, whilst the fingermarks utilised were not necessarily representative of the population of marks found in casework as a whole, they were examples of fingermarks that could easily have been determined sufficient or insufficient by a fingerprint expert and so were not the most challenging marks that may be come across during case work which may cause disagreement in their sufficiency according to fingerprint experts. This would suggest the importance of further research into practitioner decision-making outcomes in relation to the whole range of fingermarks encountered during casework including those considered straight forward for sufficiency determination, those considered borderline for submission from the perspective of a practitioner, and those considered borderline in sufficiency for comparison by fingerprint experts.

It should be born in mind that the experimental task used to generate these findings was not real life case work or, indeed, a direct imitation of real life case work, but was instead a decision-making task in which practitioners were given a set number of marks in relation to a mock case and were asked to make sufficiency decisions in relation to these marks. This approach was advantageous over examining decision outcomes in live casework as it enabled control over a number of experimental variables such as the quality, origin and development method of fingermarks presented to practitioners, and allowed for a large sample of fingermarks and participating practitioners to be used in the study. Whilst this methodology was a good representation of sufficiency decisions made in relation to fingermarks viewed in isolation, it was not necessarily representative of decisions that would have been made on exhibits from live case work in which the process used and information available may differ.

Equally it is important to bear in mind that this study required laboratory practitioners to determine the sufficiency of fingermarks from photographic images of the original marks as opposed to from the original marks in situ, as would be the case operationally. This was intended to ensure consistency in the appearance of the fingermarks viewed between the first and second experimental sessions in case of any further enhancement or degradation, and also enabled copies of the marks to be viewed by multiple practitioners simultaneously. Whilst these were colour images designed to imitate working from an original chemically treated exhibit as far as possible, there is likely to have been a difference in the clarity and quality of the fingermark that was visible on the original exhibit and the clarity and quality of the fingermark displayed in the photograph. In fact, during experimental image production it became apparent that there had been a potential increase in the quality of the fingermarks from the original exhibit to the photograph of the fingermark. This being the case it may be that an even higher proportion of comparable fingermarks are being discarded by laboratory practitioners when making sufficiency decisions based upon unenhanced developed fingermarks on original exhibits, as operationally fingerprint experts will make their sufficiency decisions based upon a black and white photographic image of the original fingermark which may be of enhanced quality, exaggerating the difference in decision-making between practitioners and fingerprint experts. It is important to further investigate the possible enhancement properties of fingermark image capture so as to allow for this difference when determining desirable decision thresholds for laboratory practitioners.

No clear fingermark submission procedure was stated by the laboratory manager during this study. It may be the case that the generation of such a procedure, in line with the requirements of the fingerprint experts (and taking into account the enhancement properties of image capture), would be successful in increasing the accuracy of decision-making according to the needs of the experts, particularly in the case of borderline marks.

A further challenge to the adoption of a fingermark submission procedure in this case may have been that in this study the fingerprint laboratory and bureau were not based at the same site. In this case, the two units worked in relative isolation with little integration in casework. Whilst this could be seen to be an increasingly common way of working within the UK due to mergers between scientific support units and an increased use of remote transmission technology, further research could potentially compare the results of this study with the differences in practitioner and expert sufficiency decisions made in other police forces with joined laboratories and bureaux to assess the effect of this isolated approach to workflow on the efficiency of the fingerprint laboratory evidence filtering process.

4.2. The effect of contextual information and policy on sufficiency decision-making

Although no distinct overall threshold difference was found between experts and practitioners, there did appear to be a difference in the practitioner decision threshold according to the context in which the mark was presented. A higher submission threshold was shown when fingermarks were presented in a volume context. In these cases, fewer marks were submitted overall, fewer insufficient fingermarks were submitted, and more sufficient prints were discarded. A lower threshold was observed in a serious context where more fingermarks were submitted overall, with more insufficient fingermarks submitted and fewer sufficient marks discarded.

This threshold difference may seem to be a desirable effect, as it channels resources towards serious crime cases. More fingermarks were submitted in these cases, which, whilst resulting in an increased waste of resources on processing insufficient marks, meant that less evidence of sufficient quality was lost (although it was observed that laboratory practitioners did not successfully prevent the loss of all evidentially useful fingermarks). In (arguably less critical) volume cases, fewer resources were wasted on poor quality marks, but, as observed in this study, at the price of the loss of some evidentially useful marks. However, there is no documented procedure within the Metropolitan Police Evidence Recovery Unit Fingerprint Laboratory that prescribes a different fingermark submission criterion between cases of serious and volume crime. The laboratory manager stated that staff should not treat fingermark selection differently according to crime type, regardless of resource constraints. It is important to bear in mind that the laboratory that participated in this study did not routinely deal with volume crime cases and so would not be expected to have a separate procedure for dealing with such cases. It would seem that either this protocol had not been well communicated to practitioners who believe they should in fact be employing differing thresholds, that they themselves judge that they should use different thresholds, or that they are not aware of their use of different thresholds.

If practitioners are aware of the laboratory procedure, there are a number of reasons why they might, nevertheless, employ a different threshold for different crime contexts. In cases of volume crime dealt with by the Metropolitan Police Service it is common for no persons of interest to be supplied for the case. Rather, any fingermarks found will be searched on Ident1 (the central national database for fingerprints within the UK), which requires a minimum of eight characteristics in a fingermark to perform a search. It may be that a number of practitioners taking part in this study were aware of this criterion and so raised their suitability threshold in cases of volume crime in order to fulfil it. From a decision-theoretic perspective see e.g. [22,23], if practitioners believe that it is a more severe error to fail to convict the correct individual in the case of a serious crime than a volume crime, it is rational to lower the threshold at which fingermarks are submitted for analysis so as to provide better protection against missing vital evidence. Indeed, laboratory practitioners might also consciously employ such a strategy as they feel that submitting an insufficient quality fingermark is a more easily justifiable error — ease of justification being one previously identified goal of human decision-making processes [24]. Previous research within the fingerprint domain has shown the emotional context of a case to

affect the decision-making thresholds of fingerprint experts [10–12]. The results of this study could subsequently be shown to be a further example of emotional influences on the process of fingerprint recovery. Emotional aspects may, indeed, have a particularly strong effect on decisions made in a fingerprint development laboratory where emotive items of evidence from crime scenes are dealt with directly during routine casework. It may be the case that the negative emotions associated with a murder case cause a narrowing of focus leading to a greater perception of local minutiae detail causing a reduction of the submission threshold in these instances [17,18], or it may be that practitioners are vulnerable to 'seizing and freezing' [25], where the amount of information required to make a decision is lowered when there is a high need for closure, such as in a serious case, resulting in a lower sufficiency threshold for serious case fingermarks.

Schiffer and Champod [7], however, found that there was no effect of the serious or volume category of a case on the number of minutiae identified or on the outcome of the mark when using student participants within a university environment, rather than practitioners working within the resource constraints of their department. This finding may provide support for the idea that the differences between the two contextual scenarios found in the present study occurred because the practitioners had been exposed to operational resource constraints, or believed that they should be working to differing thresholds (perhaps due to differing policies elsewhere in the organisation) and had consciously lowered their fingermark submission threshold in more serious crimes in order to target resources to prints from cases with more resources allocated. It could also be that practitioners had raised their submission threshold in cases of volume crime to meet Ident1 requirements, or, alternatively, that they had followed a decision theoretic perspective. If it were the case that more automatic emotional-based reactions (such as a narrowing of processing or 'seizing and freezing') altered thresholds irrespective of practical resource constraints it might have been expected that the students would also have been vulnerable to these cognitive effects. It is, finally, also possible that a combination of all these potential processes is affecting the decision threshold.

It is important to further investigate this area to first establish if the threshold differences are problematic or are actually desirable. It could be argued that falling back on contextual information in the absence of policy in the case of ambiguous marks is, in this case, a positive aspect of the decision process as it has enabled a reduction in the number of 'missed' marks in relation to serious crime in comparison to volume crime. However, this needs to be supported by a decision-making strategy or policy of some kind to ensure that these effects are actually desired by the organisation. Any unwanted effects need to be examined further with the ultimate aim of removing them in the decision-making process.

5. Conclusion

The results of the present study demonstrate a clear need to further investigate the process of fingerprint laboratory fingermark submission, an area in which there is currently a paucity of empirical published research. If the results of this study were translated directly to casework then it would appear that the quality assurance procedures currently adopted within the Metropolitan Police Service may not be sufficient in identifying the loss of evidentially valuable marks to an appropriate standard, and the quality assurance processes designed to raise and tackle or indeed rectify the submission of poor quality marks over a 20% threshold may not be sufficiently effective, in the case of borderline fingermarks. It may also be the case that other UK laboratories are not currently picking up upon the loss of evidentially valuable fingermarks as the process of initial quality assessment by a laboratory practitioner is common across the UK and this area of quality assurance is not covered under the requirements of UKAS accreditation. At a time when forensic science is receiving a high level of national scrutiny, and the ability of police force in-house laboratories to provide unbiased evidence recovery is in question [26], it would seem essential that these issues are openly investigated and that solutions are actively sought throughout UK fingerprint laboratories where this is found to be necessary.

The present study found that a very similar number of borderline marks are being erroneously discarded and erroneously submitted. This highlights the potential for maximising evidential value using existing resources through correcting these decisions. It would seem that there is high value in working to better understand the decision processes involved and to develop improved communication, corroboration and training links between fingerprint experts and fingerprint development laboratories so as to ensure that procedures are in place which would provide experts with the marks that they need. Further investigation is then needed to determine if there are psychological factors impinging on practitioners and experts operating within these procedures. In times of fiscal uncertainty [27], it would seem that maximising the value of the resources that are already present is paramount to delivering a best value fingerprint recovery service.

Acknowledgements

This research was funded by the Engineering and Physical Sciences Research Council of the UK through the Security Science Doctoral Research Training Centre (UCL SECReT) based at University College London (EP/G037264/1).

The researchers would like to thank Professor Gary Pugh, Director of Forensic Services at the Metropolitan Police Service, for his support of this research, and the management and staff of the Metropolitan Police Service Fingerprint Bureau and Evidence Recovery Unit Fingerprint Laboratory for allowing access to facilities and for their participation in this study.

It must be recognised that this was a single experiment targeting one particular aspect within a wider process. It does not reflect the whole process or the final product presented in evidence to the courts. The Metropolitan Police Service has procedures in place to monitor and reduce the risk of an error being made by a single practitioner decision, however, the results of this research have highlighted that a review of current procedures is required, and changes will be made where necessary to further mitigate any risks.

References

[1] V. Bowman (Ed.), Manual of Fingerprint Development Techniques, second ed.Home Office, London, ISBN: 1 85893 9720, 1998 (revised 2002, 2004, 2009).
[2] Daubert v. Merrell Dow Pharmaceuticals, Inc. 509, US, 579 (1993).
[3] The Law Commission, Expert Evidence in Criminal Proceedings in England and Wales, The Stationery Office Limited, London, 2011.
[4] National Research Council Committee on Identifying the Needs of the Forensic Science Community, Strengthening Forensic Science in the United States: A Path Forward, National Academy Press, Washington DC, 2009.
[5] House of Commons Science and Technology Committee, The Forensic Science Service — Seventh Report of Session 2010–2011, The Stationery Office Limited, London, 2011.
[6] British Standards Institute, BS EN ISO/IEC 17025:2005 General Requirements for the Competence of Testing and Calibration Laboratories, 2005.
[7] B. Schiffer, C. Champod, The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation, Forensic Sci. Int. 167 (2007) 116–120.
[8] Scientific Working Group on Friction Ridge Analysis, Study and Technology, Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint), version 1.0, 2011.
[9] C. Neumann, I. Mateos Garcia, G. Langenburg, J. Kostroski, Operational benefits and challenges of the use of fingerprint statistical models: a field study, Forensic Sci. Int. 212 (2011) 32–46.
[10] I.E. Dror, C. Champod, G. Langenburg, D. Charlton, H. Hunt, R. Rosenthal, Cognitive issues in fingerprint analysis: inter and intra expert consistency and the effect of a 'target' comparison, Forensic Sci. Int. 208 (2011) 10–17.
[11] I.E. Dror, A.E. Peron, S.L. Hind, D. Charlton, When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints, Appl. Cogn. Psychol. 19 (6) (2005) 799–809.
[12] I.E. Dror, D. Charlton, A.E. Peron, Contextual information renders experts vulnerable to making erroneous identifications, Forensic Sci. Int. 156 (2006) 74–78.

[13] D. Charlton, P. Fraser-Mackenzie, I.E. Dror, Emotional experiences and motivating factors associated with fingerprint analysis, J. Forensic Sci. 55 (2) (2010) 385–393.
[14] L.J. Hall, E. Player, Will the introduction of an emotional context affect fingerprint analysis and decision-making? Forensic Sci. Int. 181 (2008) 36–39.
[15] I.E. Dror, Expectations, contextual information, and other cognitive influences in forensic laboratories, Sci. Justice 52 (2012) 132.
[16] I.E. Dror, The paradox of human expertise: why experts get it wrong, in: N. Kapur (Ed.), The Paradoxical Brain, Cambridge University Press, Cambridge, UK, 2011, p. 177.
[17] P.A. Gable, E. Harmon-Jones, Approach-motivated positive affect reduces breadth of attention, Psychol. Sci. 19 (5) (2008) 476–482.
[18] P.A. Gable, E. Harmon-Jones, The blues broaden, but the nasty narrows: attentional consequences of negative affects low and high in motivational intensity, Psychol. Sci. 21 (2) (2010) 211–215.
[19] K.A. Ericsson, A.C. Lehmann, Expert and exceptional performance: evidence of maximal adaption to task constraints, Annu. Rev. Psychol. 47 (1996) 273–305.
[20] G.M. Langenburg, Pilot study: a statistical analysis of the ACE-V methodology — analysis stage, J. Forensic Identif. 54 (2004) 64–79.
[21] P. Fraser-Mackenzie, I.E. Dror, K. Wertheim, Cognitive and contextual influences in determination of latent fingerprint suitability for identification judgments, Sci. Justice 53 (2013) 144–153.
[22] J. Pratt, H. Raiffa, R. Schlaifer, Introduction to Statistical Decision Theory, MIT Press, Massachusetts, 1995.
[23] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, Forensic Sci. Int. 177 (2008) 120–132.
[24] J.R. Bettman, M.F. Luce, J.W. Payne, Constructive consumer choice processes, J. Consum. Res. 25 (1998) 187–217.
[25] A.W. Kruglanski, D.M. Webster, Motivated closing of the mind: seizing and freezing, Physiol. Rev. 103 (2) (1996) 263–283.
[26] House of Commons Science and Technology Committee, Forensic Science, vol. 1, The Stationery Office Limited, London, 2013.
[27] D. Charlton, Standards to avoid bias in fingerprint examination? Are such standards doomed to be based on fiscal expediency? J. Appl. Res. Memory Cogn. 2 (2013) 71–72.