

The potential (negative) influence of observational biases at the analysis stage of fingerprint individualisation

Beatrice Schiffer^{*}, Christophe Champod

*Université de Lausanne, Ecole des Sciences Criminelles, Institut de Police Scientifique,
Quartier UNIL-Sorge, Bâtiment Batochime, Lausanne CH-1015, Switzerland*

Received 9 June 2006; accepted 14 June 2006

Available online 28 July 2006

Abstract

Recent cases of erroneous identification have strengthened critical comments on the reliability of fingerprint identification. This goes hand in hand with recent publications regarding the lack of scientific foundation of the discipline. Combined with “legislative” needs, such as for instance the admissibility criteria under Daubert, or experimental studies revealing potential bias, the call for research on the identification process has become more urgent.

That background set the basis of this research project financed by the Swiss National Science Foundation (SNSF) which includes, among other parts, experimental tests to study potential observational biases in the analysis stage of fingerprint individualisation. These tests have been submitted to several groups of forensic science students at the University of Lausanne.

The aim is to study factors potentially influencing the analysis of fingerprints, more specifically the influence of training/education (test I) as well as the potential impact of case contextual information or known print availability (test II). For all tests students were given 11 or 12 fingerprints of a medium to difficult quality, with a range of 8–15 minutiae. For all tests the task was always the same for the participants but carried out in different contexts: to analyse the marks, to annotate the minutiae observed, to designate them and to decide on the status of the mark in two categories, exploitable and identifiable. The aim was to see how the fingerprints were annotated by different individuals so as to have an idea of the variation in annotation and counting in the analysis stage only.

For test I, students were submitted the same 12 fingerprints before and after having followed specific training in fingerprint individualisation. The aim was to see how training/education impacts the analysis of fingerprints. For test II, were participants given eleven fingerprints so as to study whether the presence of a comparison print changes the amount of minutiae found and whether low/high-profile background information influences the analysis stage.

Results show that for test I the effect of training can be observed, among other, in an increase of minutiae annotated and a higher consensus between participants. For test II no effect of the stimuli used to induce observational biases has been observed by all of the factors studied.

© 2006 Elsevier Ireland Ltd. All rights reserved.

Keywords: Fingermarks; Observational biases; Errors; Misidentification; Experimental study

1. Introduction

Traditional forensic identification evidence – especially fingerprint individualisation – though accepted in court for over 100 years, has been challenged lately. Reasons are, among others, the highly publicised erroneous identification by the FBI in the Mayfield affair [1] or legislative needs such as renewed attention on admissibility criteria [2]. The misleading influence

of observational biases in forensic science [3] is mentioned as a possible explanation for errors in the Mayfield misidentification [1]. It has been theorised that the lower the quality of the fingerprint, the more demanding and subjective the analysis process will be, and accordingly the more vulnerable to stimuli potentially inducing observational biases [4]. Those stimuli can be “circular reasoning” (looking for features found in the comparison print on the fingerprint), disregarding of the “one discrepancy rule” (explaining away discrepancies) and incorrect verification of the results when questioned [1]. All these problems do have their origin in the inaccurate following of the fingerprint individualisation process Analysis, Comparison,

^{*} Corresponding author. Tel.: +41 21 692 46 57; fax: +41 21 692 46 05.

E-mail address: Beatrice.Schiffer@unil.ch (B. Schiffer).

Evaluation and Verification (ACE-V) [5], especially jeopardising the independence of the four stages. Within this process basically three elements exist. They are:

1. the fingerprint: the object to be observed;
2. tools and methods used: the way the observation is made;
3. the fingerprint examiner: the observer.

Observational biases can have their influence on all three elements and might happen all along the ACE-V process. We will focus on the analysis stage, as Langenburg [6] showed that significant differences exist already in the beginning of the process. Indeed, the amount of minutiae found by fingerprint examiners in contrast to lay persons increased. Enlargement of the fingerprint was also observed to modify, that is to increase the amount of minutiae observed. Experimental tests on the influence of bottom-up and bottom-down factors on fingerprint individualisation have shown that the quality of the fingerprint (contrast, potential distortion and apposition) is an important factor in creating a difficult decisional situation (for lay-persons) and that this might be a critical factor in the individualisation process [4]. Furthermore, a study with five experienced fingerprint examiners showed that, when submitted twice the same fingerprint in a highly different (and emotionally charged) context three of them changed their initial opinion [7]. Kerstholt et al. studied observational biases in shoe print comparison [8]. They evaluated how expectations (background of the case) and complexity of task (difficulty of the comparison) as well as experience influence the evaluation of simulated case work of 12 shoe print examiners. Contrary to their expectations they did not find any of these influences having a bearing on the examiners results. Only experience changed the way decisions were justified, but not the results as such.

The examiner will be conditioned by all these previous factors, namely the fingerprint and the protocol to be followed for observation. However, in addition to them various other external and internal elements will influence him. Internal factors will be understood as those characteristics which are part of the individual. This comprises for instance his visual faculties, his training and his experience. They will not vary considerably from one fingerprint comparison to another, though they might evolve. External factors will be understood as elements that will influence one particular case because of exceptional circumstances. These might comprise special media coverage due to the profile of the case or working under special pressures due to time or other constraints.

All these factors might be leading to irreproducible analysis, especially if added to two peculiarities of forensic science work. First, there is the environment and the aim of forensic work that can be tense due to time and heavy context associated with the case. Second, known comparison material may be available upfront, and originate mostly from a “relevant” source from the investigation perspective. Expectation would be that the chance to find a match between the mark and the known comparison material might be a high a priori for the examiner. This phenomenon might even be reinforced by experiencing this happening repeatedly.

To summarise, it can be stated that the lower the quality of the fingerprint, the more demanding and subjective the analysis process will be. According to Tversky and Kahneman [9] the tendency to rely on additional, though not necessarily relevant information increases when the data present does not offer enough information for a clear decision. It is then that case relevant and not mark relevant information will be used to reach an opinion. In short, fingerprint individualisation is more vulnerable to expectations and biases if the difficulty of the task is increased.

The aim of the research was thus to study the potential (negative) influence of observational biases at the analysis stage of fingerprint individualisation by an experimental approach using low quality marks. Such research should help assess whether or not observational biases can develop into errors compromising the reliability of fingerprint individualisation as advanced by some authors [3,10].

Several factors were studied. The first was the effect of training on the analysis stage of fingerprints (test I). Forensic science students were tested before having acquired thorough knowledge in fingerprint individualisation and after having followed a course on forensic identification and having carried out practical examination. It was predicted that the number of minutiae annotated (per fingerprint and per individual) would increase with training and that the overall variation between individuals for the minutiae counted for a given fingerprint would diminish. It was also predicted that the amount of fingerprints considered exploitable (useful for comparison purposes) and identifiable would increase as well with training. Langenburg [6] showed already the higher efficiency of trained professionals compared to lay person. We felt important to attempt to show that such a trend (if any) can be related to specialised training.

Secondly, the potential impact of observational effects has been explored (test II). Stimuli in form of the presence of a matching and a non-matching comparison fingerprint was expected to vary the amount of minutiae found by the group submitted to condition A (group A) versus condition B (group B). The same phenomenon was expected for the disclosure of a high-profile (terrorism) versus low-profile (attempted petty burglary) background scenario.

2. Method

2.1. Participants

Participants were forensic science students of the School of Criminal Sciences of the University of Lausanne. Depending on the test, different classes were sampled. However, all forensic science students have followed the same theoretical and practical lessons in fingerprint individualisation.

2.2. Materials and design

Fingerprints from practical student case work were given for analysis to three experienced fingerprint examiners in order to determine among others the number of minutiae found. Their findings were used as criterion for selecting fingerprints for the tests. A total of 16 fingerprints relatively close to the Swiss threshold of 12 minutiae and of varying nature were included in both tests (12 in test I and 11 in test II).

Table 1
Illustration of the design for the test II on stimuli for potential observational biases

Group	Reference	Availability of known print			Context		
		None	Matching print	None	Non-matching print	High-profile	Low-profile
A	Same finger	None	Matching print	None	Non-matching print	High-profile	Low-profile
B	mark	Matching print	None	Non-matching print	None	Low-profile	High-profile
Finger mark	1	2	3	4	5	6–8	9–11

For test I (experience) a within group design over a period of time was used. Experience was manipulated by testing the same population (39 students) before and after (29 students) having acquired specific theoretical and practical knowledge, that is having followed a full course in forensic identification. For test II (observational biases) a between-group (between subjects) design was chosen. Observational biases were manipulated by subjecting group A to one condition and group B to the second condition for the same fingermark (Table 1). The sample included 20 master and 28 bachelor students, all having followed the same forensic identification course. To study the influence of the presence of a known comparison print at the analysis stage a matching and a non-matching case were given to each group. The same design was used for the high/low-profile manipulation (attempted petty burglary versus terrorism case). One reference fingermark was presented to both groups without specific stimuli, so as to have a direct between test condition comparison. The test sheets for groups A and B were randomly distributed within the class.

Materials comprised the four times enhanced fingermarks all printed with a photo-quality Fuji Pictrostrat printer for both tests I and II.

The results to be observed were the same for both tests I and II. For each fingermark, the response consisted in the total number of minutiae annotated, the respective type of minutiae – ridge ending, bifurcation, point or unknown – as well as the classification of the fingermark into exploitable or identifiable.

2.3. Procedure

For both tests I and II, participants carried out the task during a normal class. They first read a written instruction sheet explaining the task to them and the way to proceed. Explanations as to the aim of the research were minimal and kept very general. Then, the task was described, namely, to analyse each given fingermark by annotating each minutiae found, to determine the type of minutiae found – ridge ending, bifurcation, point or unknown – (see illustration below) and to assess the usefulness of the fingermark using two categories (exploitable or identifiable).



Participants were encouraged to differentiate the minutiae marked either by letter (A, B, P and U) and/or by colours. Participants were expected to use the standard procedure of analysing a fingerprint learned during their theoretical and practical lessons based on Ashbaugh [5]. The illustrations of the types of minutiae were given in order to avoid misgivings about the definition of a minutiae.

3. Results

3.1. Test I: training

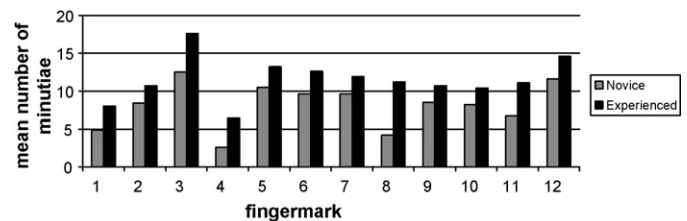
The aim was to observe possible differences in the annotation of minutiae before and after training of forensic science students. In a first time the mean number of minutiae found by participants over all fingermarks was compared. If

they did find an average of 8.1 minutiae when novice, they found 11.5 when having been trained, that is an average of 3.4 minutiae more when more experienced. Increase was highest for fingermarks with background noise and/or bad contrast. Then, for each fingermark the means and the difference of minutiae observed were compared. The minimal and maximal number of minutiae found per fingermark and the range of minutiae found (difference between minimum and maximum) were noted as well.

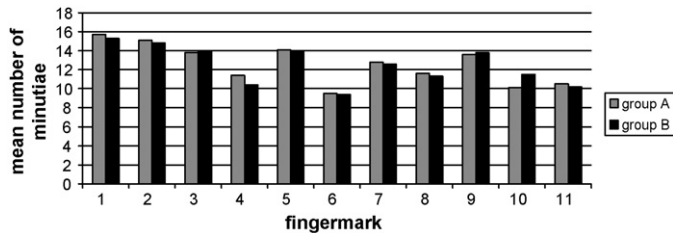
For each of the 12 fingermarks the mean of minutiae found does increase considerably from the novice to the experienced condition, with a minimum of two minutiae more observed, up to a maximum of nearly seven minutiae (Graph 1). All of these values are highly significant by the statistical *t*-test. The minimum amount of minutiae found per fingermark does increase as well for all fingermarks. The same is true for the maximum, excluding however 3 out of 12 fingermarks. In general, the range of minutiae found by all participants for the same fingermark does decrease after having been trained in fingerprint examination (two out of three fingermarks). As to the types of minutiae found, it has been observed that over all minutiae marked the total number of ridge endings passes from 34.2% to 44.1%, a very significant increase. Inversely, the amount of bifurcations decreases from 54.7% to 47.6%. The rest of the change due to points and unknown minutiae is negligible in comparison. Concerning the classification of the fingermarks in terms of exploitable and identifiable in can be observed that the mean number of exploitable nearly doubles from novice to experienced, while it does more than double for the identifiable condition. In summary, results illustrate that participants do see more minutiae for the different fingermarks while showing a greater consensus in their observations.

3.2. Test II: observational biases

The aim was to observe possible differences in the annotation of minutiae between the two groups due to the manipulation of the factors: (1) availability of a known



Graph 1. Variation of the mean number of minutiae annotated by novice and more experienced participants on 12 fingermarks.



Graph 2. Variation of the mean number of minutiae annotated by groups A and B on the 11 fingerprints studied using different stimuli.

comparison print and (2) differing background case information. For the mean number of minutiae over all fingerprints (group A: 138, 01; group B: 137, 28) no difference between the groups A and B was noted. The same observation applies for all other parameters studied (see Graph 2 for an illustration of the difference in mean number of minutiae found by the two groups).

More specifically, the first fingerprint which was used as reference does not show any significant difference between the groups, be it the amount of minutiae found, the mean or any other factor compared. The same results apply to all other 10 fingerprints. Considering the presence/absence of a known fingerprint, be it the matching or the non-matching condition, no significant difference can be observed neither. The same result does apply to the low/high-profile condition for all data observed. The type of minutiae observed does not vary between the two groups, nor does the qualification of the fingerprints as exploitable or identifiable.

4. Discussion

4.1. Test I

In agreement with our expectations for test I a pronounced effect (an increase of the total number of designated minutiae) of training on the analysis stage of fingerprint individualisation was found for test I. It might be inferred that participants have gained knowledge on the observation of minutiae. A certain move towards consensus of the way a fingerprint is perceived can be observed as well. However, this effect is limited, as still quite important variations do subsist between examiners. This is in accordance with the observations of Evett and Williams [11]. Furthermore, training does affect the type of minutiae found, as bifurcations are less often attributed. Though these variations are smaller for the more experienced participants in comparison with their novice performance, the abovementioned tests show that differences will still persist. Indeed, observation is linked to individuals, which might differ slightly in what they observe. Therefore, a clear subjective element persists.

4.2. Test II

Contrary to our initial expectations for test II on the potential effects of stimuli inducing observational biases, no effect of availability of known print nor context information has been

observed. This was true for all fingerprints used in the test. These results do, to a certain degree, contradict previous findings or hypotheses, for instance Risinger et al. [3] and their overview of studies on the detrimental effects of expectation on reasoning and perception. Also the study by Dror et al. [7], showing with a within-subject design that context information might influence the conclusion drawn from a fingerprint as to its identification or not. However, these studies focus on the *outcome*, the moment when observations have to be evaluated in terms of evidential value. It is argued here, that not all stages of the process of ACE-V are similarly vulnerable to observational biases. Indeed, the less decisional tasks are involved in a stage, the less the risk of “oversimplifying” information in order to reach a decision might be [9]. Thus, individuals might quite correctly observe minutiae if the task is only to designate them, but they might not be able to undertake this task correctly, if forced to compare with reference material and draw a conclusion from that all encompassing process.

The study by Kerstholt et al. [8], although based on shoe print examinations by experienced examiners and a slightly different design, does come to a similar conclusion as far as the effects of expectation and complexity go. They did not observe any difference due to the manipulation of the background, which should modify expectation. They explain this result partly by the presence of a formal guideline employed by the examiners, but also by the potential positive influence of experience on vulnerability to these potential sources of biases. As the experience of our participants is limited, this second thesis does not explain our results. The first aspect, the presence of a guideline, or a “structured” approach to fingerprint examination might be a better way into explaining the results. Indeed, as long as the guidelines are followed properly, there is a small risk of drift happening.

Although further research is needed for a better understanding of how the ACE-V protocol could be influenced by potential observational biases, especially in the phases following the analysis—the results presented here tend to show the robustness of the analysis phase.

Acknowledgment

This research was financed by the Swiss National Science Foundation (SNSF) 100012-105817/1.

References

- [1] O.o.t.I.G. U.S. Department of Justice (Ed.), A Review of the FBI's Handling of the Brandon Mayfield Case, 2006, pp. 1–330.
- [2] P.C. Giannelli, Scientific Evidence and Miscarriages of Justice, in: Proceedings of the 16th International Conference of the International Society for the Reform of Criminal Law, Charleston, SC, USA, (2002), pp. 1–14.
- [3] M.D. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion, *California Law Rev* 90 (2002) 1–56.
- [4] I.E. Dror, A. Péron, S.-L. Hind, D. Charlton, When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints, *Appl. Cognit. Psychol.* 19 (6) (2005) 799–809.

- [5] D.R. Ashbaugh, *Quantitative–Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*, CRC Press LLC, Boca Raton, NY, 1999.
- [6] G.M. Langenburg, Pilot study: a statistical analysis of the ACE-V methodology—analysis Stage, *J. Forensic Identif.* 54 (2004) 64–79.
- [7] I.E. Dror, D. Charlton, A. Péron, Contextual information renders experts vulnerable to making erroneous identifications, *Forensic Sci. Int.* 156 (2006) 74–78.
- [8] J.H. Kerstholt, R. Paashuis, M. Sjerps, Shoe print examinations: effects of expectation, complexity and experience, *Forensic Sci. Int.* (2006).
- [9] A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, *Science* 185 (1974) 1124–1131.
- [10] S.A. Cole, Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponents' discourse, *Law Policy* 28 (2006) 109–135.
- [11] I.W. Evett, R.L. Williams, A review of the sixteen point fingerprint standard in England and Wales, *Fingerprint World* 21 (1995).