UNIVERSITY OF LAUSANNE
FACULTY OF LAW AND CRIMINAL JUSTICE
SCHOOL OF CRIMINAL JUSTICE
FORENSIC SCIENCE INSTITUTE

The Relationship between Forensic Science and Judicial Error: A Study Covering Error Sources, Bias, and Remedies

PhD thesis submitted to obtain the doctoral degree in forensic science

Beatrice SCHIFFER

Lausanne, 2009

UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de Mme Beatrice Schiffer, candidate au doctorat en sciences forensiques, intitulée

« The Relationship between Forensic Science and Judicial Error : A Study Covering Error Sources, Bias, and Remedies »

Le Président du Jury

Professeur Pierre Margot

Lausanne, le 17.10.2008

To my family, especially my parents

Summary

Forensic science - both as a source of and as a remedy for error potentially leading to judicial error - has been studied empirically in this research. A comprehensive literature review, experimental tests on the influence of observational biases in fingermark comparison, and semi-structured interviews with heads of forensic science laboratories/units in Switzerland and abroad were the tools used. For the *literature review*, some of the areas studied are: the quality of forensic science work in general, the complex interaction between science and law, and specific propositions as to error sources not directly related to the interaction between law and science. A list of potential error sources all the way from the crime scene to the writing of the report has been established as well. For the *empirical tests*, the ACE-V (Analysis, Comparison, Evaluation, and Verification) process of fingermark comparison was selected as an area of special interest for the study of observational biases, due to its heavy reliance on visual observation and recent cases of misidentifications. Results of the tests performed with forensic science students tend to show that decision-making stages are the most vulnerable to stimuli inducing observational biases. For the *semi-structured interviews,* eleven senior forensic scientists answered questions on several subjects, for example on potential and existing error sources in their work, of the limitations of what can be done with forensic science, and of the possibilities and tools to minimise errors. Training and education to augment the quality of forensic science have been discussed together with possible solutions to minimise the risk of errors in forensic science. In addition, the time that samples of physical evidence are kept has been determined as well. Results tend to show considerable agreement on most subjects among the international participants. Their opinions on possible explanations for the occurrence of such problems and the relative weight of such errors in the three stages of crime scene, laboratory, and report writing, disagree, however, with opinions widely represented in existing literature. Through the present research it was therefore possible to obtain a better view of the interaction of forensic science and judicial error to propose practical recommendations to minimise their occurrence.

Résumé

Les sciences forensiques – considérés aussi bien comme source de que comme remède à l’erreur judiciaire -ont été étudiées empiriquement dans cette recherche. Une revue complète de littérature, des tests expérimentaux sur l'influence du biais de l’observation dans l'individualisation de traces digitales et des entretiens semi-directifs avec des responsables de laboratoires et unités de sciences forensiques en Suisse et à l'étranger étaient les outils utilisés. Pour la revue de littérature, quelques éléments étudies comprennent: la qualité du travail en sciences forensiques en général, l'interaction complexe entre la science et le droit, et des propositions spécifiques quant aux sources d'erreur pas directement liées à l’interaction entre droit et science. Une liste des sources potentielles d'erreur tout le long du processus de la scène de crime à la rédaction du rapport a également été établie. Pour les tests empiriques, le processus d'ACE-V (analyse, comparaison, évaluation et vérification) de l'individualisation de

traces digitales a été choisi comme un sujet d'intérêt spécial pour l'étude des effets d'observation, due à son fort recours à l'observation visuelle et dû à des cas récents d'identification erronée. Les résultats des tests avec des étudiants tendent à prouver que les étapes de prise de décision sont les plus vulnérables aux stimuli induisant des biais d'observation. Pour les entretiens semi-structurés, onze forensiciens ont répondu à des questions sur des sujets variés, par exemple sur des sources potentielles et existantes d'erreur dans leur travail, des limitations de ce qui peut être fait en sciences forensiques, et des possibilités et des outils pour réduire au minimum ses erreurs. La formation et l'éducation pour augmenter la qualité des sciences forensiques ont été discutées ainsi que les solutions possibles pour réduire au minimum le risque d'erreurs en sciences forensiques. Le temps que des échantillons sont gardés a été également déterminé. En général, les résultats tendent à montrer un grand accord sur la plupart des sujets abordés pour les divers participants internationaux. Leur avis sur des explications possibles pour l'occurrence de tels problèmes et sur le poids relatif de telles erreurs dans les trois étapes scène de crime, laboratoire et rédaction de rapports est cependant en désaccord avec les avis largement représentés dans la littérature existante. Par cette recherche il était donc possible d'obtenir une meilleure vue de l'interaction des sciences forensiques et de l'erreur judiciaire afin de proposer des recommandations pratiques pour réduire au minimum leur occurrence.

Zusammenfassung

Forensische Wissenschaften - als Ursache und als Hilfsmittel gegen Fehler, die möglicherweise zu Justizirrtümern führen könnten – sind hier empirisch erforscht worden. Die eingestzten Methoden waren eine Literaturübersicht, experimentelle Tests über den Einfluss von Beobachtungseffekten (observer bias) in der Individualisierung von Fingerabdrücken und halbstandardisierte Interviews mit Verantwortlichen von kriminalistischen Labors/Diensten in der Schweiz und im Ausland. Der *Literaturüberblick* umfasst unter anderem: die Qualität der kriminalistischen Arbeit im Allgemeinen, die komplizierte Interaktion zwischen Wissenschaft und Recht und spezifische Fehlerquellen, welche nicht direkt auf der Interaktion von Recht und Wissenschaft beruhen. Eine Liste möglicher Fehlerquellen vom Tatort zum Rapportschreiben ist zudem erstellt worden. Für die e*mpirischen Tests* wurde der ACE-V (Analyse, Vergleich, Auswertung und Überprüfung) Prozess in der Fingerabdruck-Individualisierung als speziell interessantes Fachgebiet für die Studie von Beobachtungseffekten gewählt. Gründe sind die Wichtigkeit von visuellen Beobachtungen und kürzliche Fälle von Fehlidentifizierungen. Resultate der Tests, die mit Studenten durchgeführt wurden, neigen dazu Entscheidungsphasen als die anfälligsten für Stimuli aufzuzeigen, die Beobachtungseffekte anregen könnten. Für die halbstandardisierten Interviews beantworteten elf Forensiker Fragen über Themen wie zum Beispiel mögliche und vorhandene Fehlerquellen in ihrer Arbeit, Grenzen der forensischen Wissenschaften und Möglichkeiten und Mittel um Fehler zu verringern. Wie Training und Ausbildung die Qualität der forensischen Wissenschaften verbessern können ist

zusammen mit möglichen Lösungen zur Fehlervermeidung im selben Bereich diskutiert worden. Wie lange Beweismitten aufbewahrt werden wurde auch festgehalten. Resultate neigen dazu, für die meisten Themen eine grosse Übereinstimmung zwischen den verschiedenen internationalen Teilnehmern zu zeigen. Ihre Meinungen über mögliche Erklärungen für das Auftreten solcher Probleme und des relativen Gewichts solcher Fehler in den drei Phasen Tatort, Labor und Rapportschreiben gehen jedoch mit den Meinungen, welche in der Literatur vertreten werden auseinander. Durch diese Forschungsarbeit war es folglich möglich, ein besseres Verständnis der Interaktion von forensischen Wissenschaften und Justizirrtümer zu erhalten, um somit praktische Empfehlungen vorzuschlagen, welche diese verringern.

Resumen

Esta investigación ha analizado de manera empírica el rol de las ciencias forenses como fuente y como remedio de potenciales errores judiciales. La metodología empleada consistió en una revisión integral de la literatura, en una serie de experimentos sobre la influencia de los sesgos de observación en la individualización de huellas dactilares y en una serie de entrevistas semi-estructuradas con jefes de laboratorios o unidades de ciencias forenses en Suiza y en el extranjero. En la revisión de la literatura, algunas de las áreas estudiadas fueron: la calidad del trabajo en ciencias forenses en general, la interacción compleja entre la ciencia y el derecho, así como otras fuentes de error no relacionadas directamente con la interacción entre derecho y ciencia. También se ha establecido una lista exhaustiva de las fuentes potenciales de error desde la llegada a la escena del crimen a la redacción del informe. En el marco de los *tests empíricos*, al analizar los sesgos de observación dedicamos especial interés al proceso de ACE-V (análisis, comparación, evaluación y verificación) para la individualización de huellas dactilares puesto que este reposa sobre la observación visual y ha originado varios casos recientes de identificaciones erróneas. Los resultados de las experimentaciones realizadas con estudiantes sugieren que las etapas en las que deben tomarse decisiones son las más vulnerables a los factores que pueden generar sesgos de observación. En el contexto de las *entrevistas semi-estructuradas*, once científicos forenses de diversos países contestaron preguntas sobre varios temas, incluyendo las fuentes potenciales y existentes de error en su trabajo, las limitaciones propias a las ciencias forenses, las posibilidades de reducir al mínimo los errores y las herramientas que podrían ser utilizadas para ello. Se han sugerido diversas soluciones para alcanzar este objetivo, incluyendo el entrenamiento y la educación para aumentar la calidad de las ciencias forenses. Además, se ha establecido el periodo de conservación de las muestras judiciales. Los resultados apuntan a un elevado grado de consenso entre los entrevistados en la mayoría de los temas. Sin embargo, sus opiniones sobre las posibles causas de estos errores y su importancia relativa en las tres etapas de la investigación –la escena del crimen, el laboratorio y la redacción de informe– discrepan con las que predominan ampliamente en la literatura actual. De este modo, esta investigación nos ha permitido obtener una mejor imagen de la interacción entre ciencias forenses y errores

judiciales, y comenzar a formular una serie de recomendaciones prácticas para reducirlos al mínimo.

Table of content

Summary *1*

Résumé *1*

Zusammenfassung *2*

Resumen *3*

Table of content *5*

Preliminary remarks on the structure *9*

Part I – Introducing the context and the aim of the research *11*

1. Introduction *13*

1.1 Framework of the research **13**

1.2 Criminal justice system, criminal law, judicial error, forensic science, and criminology - some definitions **16**

1.3 Aims of the research **25**

1.4 What has been achieved – Summary of results **25**

Part II – Presentation of the three areas covered by the research: literature review, empirical tests, and semi-directive interviews *27*

2. Literature review *29*

2.1 Research questions asked concerning the literature **29**

2.2 Theoretical outline **30**

2.3 Methodology and results **32**

2.3.1 The status quo of publications concerning the interaction between forensic science and judicial errors and the lack of recent publications by forensic scientists on the risks of errors in forensic science 32

2.3.2 An inventory of various types of errors (error sources) in forensic science leading to judicial errors 53

2.3.2.1 Errors in the collection of traces/specimen or errors (e.g. confusions) in the manipulation of such traces 56

2.3.2.2 Errors in the choice and performance of technical tests 58

2.3.2.3 Errors in the establishment of factual findings and interpretation of these findings in a context 61

2.3.2.4 Errors in the presentation of evidence to (non) scientists 64

2.3.2.5 Errors due to various psychological reasons, such as observer bias (Risinger et al., 2002) 68
2.3.3 The lack of literature on the decision making process in forensic science (laboratory work and evaluation of results) 69
2.3.4 The lack of empirical studies of the observer (bias) effect (Risinger et al., 2002) in forensic science 72
2.3.5 The fact that people tend to be less critical of exculpatory than of inculpatory evidence, especially when a wrongful conviction is supposed. Therefore there is a need for symmetry (Edmond, 2002) 75
2.4 Comments 75
3. Empirical tests 79
3.1 Research questions asked 80
3.2 Theoretical outline 80
3.2.1 Fingermarks 80
3.2.2 Observer effects 83
3.2.3 Observer effects in the fingermark individualisation process 85
3.3 Methodology and results 89
3.4 Comments on methodological and theoretical issues 91
4. Semi-structured interviews 93
4.1 Research questions asked 93
4.2 Theoretical outline on organisations and research methods 94
4.3 Methodology and results 96
4.3.1 Methodology 96
4.3.2 Results 97
4.3.2.1 The status quo of awareness of forensic science as possible source leading to wrongful convictions in Switzerland and compare it to other countries 97
4.3.2.2 An inventory of “narrow escapes”, where something went wrong but was “corrected” before having a negative impact or, because its unimportance was established before it could be used 98
4.3.2.3 Reasons that led to the discovery of the near-mistake 99
4.3.2.4 Mechanisms in place in order to avoid such problems 99
4.3.2.5 Possible explanations for the occurrence of such problems (error sources) and their relative weight at different stages 100
4.3.2.6 The role of education, training, and experience to anticipate/avoid, recognise, and correct such problems 102
4.3.2.7 What would be needed to solve these problems, if they occur 102

4.3.2.8 Whether forensic specimen are kept beyond the moment a judicial decision becomes final 103
4.3.2.9 Additional points of interest revealed in the interviews 103
4.5 Comments 105
Part III – Connecting and discussing previous results 109
5. Discussion 111
5.1 Is forensic science a source of judicial error? 111
5.2 How do the main results fit in with existing research? 117
5.2.1 Introduction 117
5.2.2 Existing models 119
5.2.3 A possible approach to construct models for the relationship between forensic science and judicial error 122
5.3 Lessons learned and outlook 126
6. Conclusions 131
7. Acknowledgments 133
8. References 135
9. Appendices 153
9.1 Papers published or written 153
9.1.1 Judicial error and forensic science: pondering the contribution of DNA evidence 153
9.1.2 The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation 171
9.1.3 Is ACE vulnerable to contextual biases? 178
9.1.4 Forensic science and judicial errors: comparing the point of view resulting from a literature review with the opinion of interviewed practitioners as to error sources and remedies 193
9.2 Covering letter used for the experimental tests 206
9.3 Experimental grid for the interviews 207

Index of figures and tables:

Figure 1 Overview of the contextual framework of the research 14
Figure 2 Basic steps in the life-cycle from crime scene to discovery of a judicial error 15
Figure 3 The coming paradigm shift in forensic identification science. Saks and Koehler (2005, p. 892-895) 20
Figure 4 The (possible) role of forensic science in the "creation" of a judicial error 55
Figure 5 Illustration of top-down and bottom-up factors influencing fingermark individualisation 87
Figure 6 Illustration of the design for Test II on stimuli for potential observational biases 174
Figure 7 Summary of the test conditions for Tests I and II 183
Figure 8 Fingermark A – Values for the minutiae found under condition 1 and 2 184
Figure 9 Fingermark B – Values for the minutiae found under condition 1 and 2 185
Figure 10 Correctness of the decision made by the participants according to their condition 188
Figure 11 Correctness of the decision made by the participants according to their education 188
Figure 12 Forensic science and judicial error – a simplified model 196

Table 1 Selection of publications on misidentification and admissibility questions in fingermark individualisation 82
Table 2 Research publications on fingermark annotation and influence of observational and context effects 85

Preliminary remarks on the structure

The research presented here is at the intersection of forensic science and criminology, as well as other areas including social sciences and psychology. Starting from the general topic judicial errors, which by itself are already a small part of several more general subjects different approaches from different areas have been chosen for a better understanding of the relationship of judicial errors and forensic science. Because of the widespread and varied approach, the work has been divided into three parts.

In Part I, the framework of the research will be delimitated (1.1), the topic will be introduced by defining the main terms (1.2), the aim of the research will be presented (1.3), and the main achievements summarised (1.4). Once this more general context is explained and the scope of the research delimitated, **Part II** introduces a more specific approach, or in other words, a zoom into some of the areas of interest. It contains three different chapters, a literature review (Chapter 2), the empirical study of the observer bias in fingermark individualisation (Chapter 3), and semi-structured interviews with forensic scientists (Chapter 4). Each of these parts will have a short introduction, a presentation of the questions asked, a summary of relevant theory, a presentation of the results, and a discussion of methodological aspects called comments. Papers based on these results are summarised in the corresponding chapters and can be found in the appendices. They include substantial data necessary for the understanding of the research done and are preferably read along with the main body. The literature review is the first approach to get informed on the topics of interest; it also helps to obtain more information on more specific questions. The experimental tests allow for testing of a very specific and detailed aspect, the microscopic level. On the macroscopic level are the semi-directive interviews. They also help to link the ideas collected from the literature review and the interviews with professionals in the field.

In **Part III** the results presented as independent units in Part II will be examined and discussed together (Chapter 5), placing them again in the more general framework of judicial errors and forensic science. Some points will be examined in more detail: whether the initial hypothesis can be confirmed or not; how the results fit in with the existing research; and what the main lessons learned are. The thesis will end with conclusions (Chapter 6), acknowledgements (7), references (8), and appendices (9). Symmetrically to Part I, which from a very large topic defined the threads of research done in Part II, Part III gathers the threads together to put the data obtained back into a more general picture.

Part I – Introducing the context and the aim of the research

This part explains into which larger research context the subject of forensic science and judicial errors fits in. It also illustrates why criminological concepts and research methods can be used, though the subject explored is mainly forensic science. The main terms and concepts used throughout the thesis are defined, as well as the main research question that has been studied. The introductory part finishes with a short summary of results achieved, which are then presented in more detail in Part II.

“Go on with your theory – only do remember that to guess how a job might have been done isn’t the same thing as proving that it was done that way. If you will allow me to say so, that is a distinction which people of your profession are very liable to overlook. They will confuse moral certainty with legal proof.”

(Sayers, 1991, p. 228)

1. Introduction

1.1 Framework of the research

Formal context

This research is related to the multidisciplinary research project "Wrongful Convictions in Switzerland in a Comparative Perspective" financed by the Swiss National Science Foundation (SNSF 100012-105817). This study is in its turn integrated in the research agenda proposed by the International Network of Researchers on Wrongful Convictions (INRWC)[1]. The influence of the INRWC on this research is twofold. First, the same definition of wrongful conviction is used and second, the multidisciplinary nature of the project has been taken as an example to follow.

The objective of the SNSF project was to reach practical recommendations to minimise errors in the different domains, taking into account the interactions between scientific analysis and judicial decision-making. Emphasis was given to the comparative and interdisciplinary perspective, allowing for instance the combination of several research methods and approaches. Three different parts were considered: A legal part, which consisted in a review of comparative law and a survey among all Swiss cantonal courts; a legal medicine part, with a survey on quality management questions in legal medicine; and a forensic science part, which is an integral part of the research presented here. Additional research has been done for the present thesis which was not included in the SNSF project. Details as to the general SNSF project can be found in Killias et al. (2004).

Contextual framework

Besides this framework, the contextual situation of the research will be presented next. The terms used to define it will be explained later, as the aim here is first to position the research. The criminal justice system (CJS) forms the outer boundaries of the topic. Inside, one sector will be of interest: Criminal law. Within, the focus will be on a malfunctioning aspect of it – judicial error. The approach will be even more specific as one of several possible factors leading to it will be studied: forensic science[2].

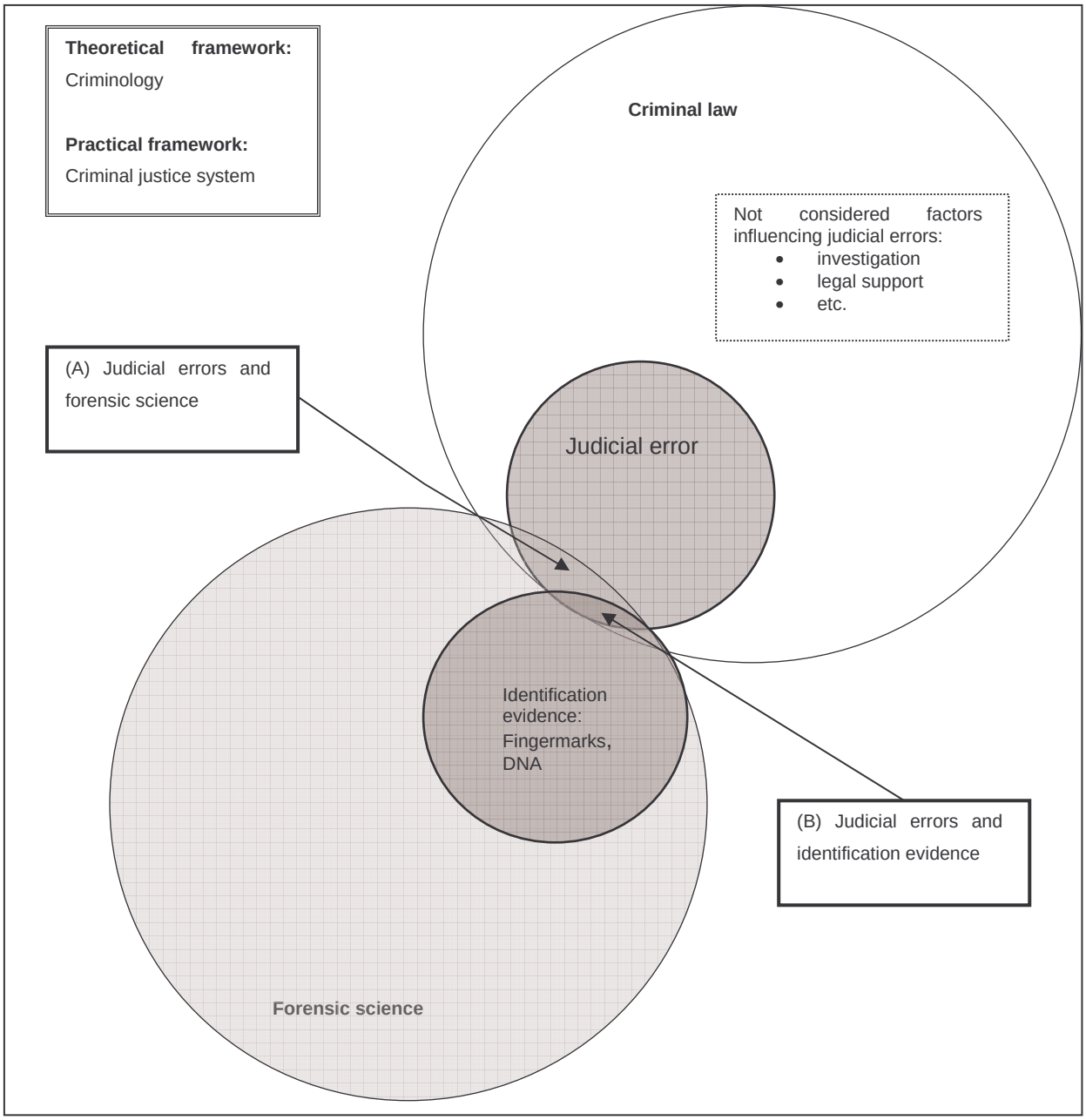
A general overview of the relationship of forensic science and judicial error will form one part of the present study (see Figure 1 (A); also Chapters 2 and 4). All areas of forensic science and of types of judicial error will be considered first, and then a restriction as to the chosen definitions for the rest of the work will be made. A particular focus will be on the area of identification evidence such as fingermarks and DNA profile analysis (see figure 1 (B); also Chapter 3).

[1] The project was the outcome of a workshop in Breil/Brigels sponsored by the Swiss National Science Foundation which allowed the creation of an International Network of Researchers on Wrongful Convictions.

[2] Other factors include eyewitness errors, police misconduct, etc., as detailed in figure 3. The focus here will be on situations where ("faulty") forensic science can lead to judicial error, though it does not necessarily do so.

Identification evidence has then been selected to be studied in more detail by empirical tests. The methodology chosen pertains to criminology.

Figure 1 Overview of the contextual framework of the research

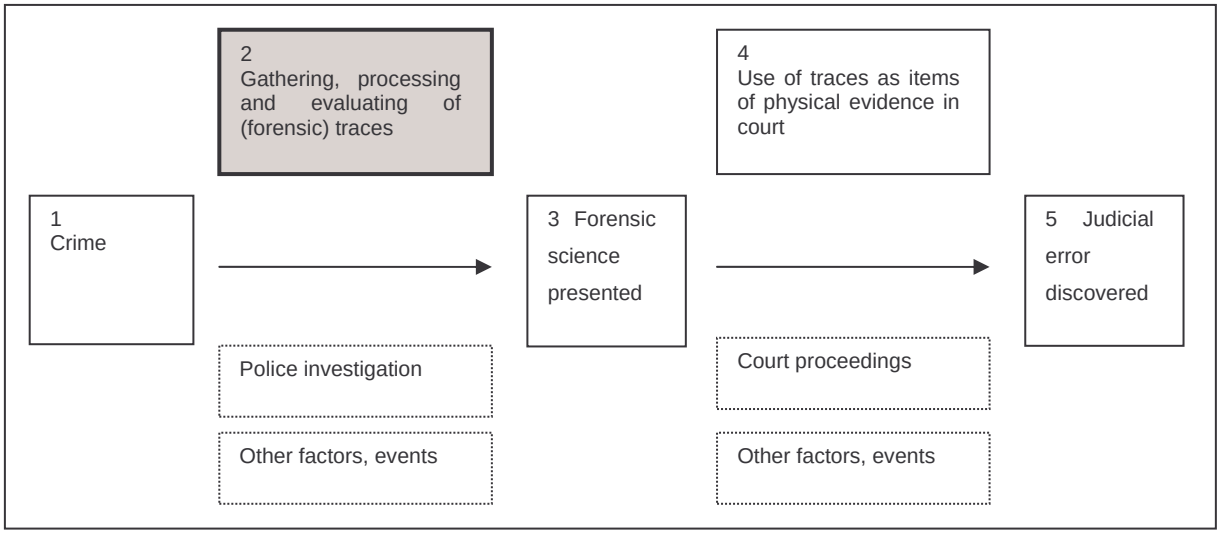


Within the framework given by the criminal justice system (CJS) one area of interest has been selected: judicial error. One of the causes (or the remedy), among others - such as the police investigation or inadequate legal support - is forensic science. This overlap is shown under (A). One specific area of forensic science is identification evidence, whose potential contribution to judicial errors will be studied in more detail (B). The size of the circles does not correspond to any proportional issues of the problem; they have been chosen to help visually illustrate the problem.

Viewing the subject in a more chronological way, the following basic steps of the phenomenon of interest can be defined:

1. Discovery of crime
2. Gathering, processing and evaluating of (forensic) traces
3. Presentation of traces and acceptance as items of physical evidence in court
4. Use of forensic science in the judicial process
5. Discovery of judicial error (due to forensic science)

Figure 2 Basic steps in the life-cycle from crime scene to discovery of a judicial error



Only the most important elements are shown here, namely the discovery of the crime (1), the presentation of items of physical elements in court (3), and the discovery of the judicial error (5). Forensic science is present both at the gathering, processing, and evaluating of traces (2) and as physical evidence used in court (4)[3]. Other factors and elements have their influence all along this process, especially the police investigation. However, in order to keep things simple they have not been detailed, nor will they along the research presented here. They include media pressure, the nature of the case, the quality of defence available, the structure and nature of the justice system, and still other elements (see MacFarlane (2003) for a discussion of such factors).

In this research special focus has been given to the gathering, processing and evaluating of traces. This corresponds to three steps of forensic science work, namely: crime scene, laboratory, and evaluation (and presentation) of results. The literature review, the empirical tests, and the semi-directive interviews with forensic science practitioners illustrate this.

Other elements that happen at the same time as the forensic science work, such as the police investigation for instance, will not be detailed in this research, though a direct collaboration between investigation and forensic science exists in reality (though not everywhere to the same

[3] As will be distinguished later, forensic science "as such" (gathering and processing traces) and the "use made" of forensic science (evidentiary proof in court).

degree). It is called investigative[4] use of forensic science, as it offers intelligence and support regarding which leads are to be followed in an investigation. In contrast, when forensic science data is used in court, it is called evaluative[5], as it helps determine the case and is used as evidence (Jackson et al., 2004). One of the main differences is the "quality" of forensic science used. What might not pass the test as evidence in court can still give useful inputs into the investigation. However, if the distinction between both is not well understood, this can be a source of confusion and error.

1.2 Criminal justice system, criminal law, judicial error, forensic science, and criminology - some definitions

Criminal justice system and criminal law

Judicial errors – a malfunctioning of the criminal justice system, which will be defined more precisely later - are of interest both to society in general and to researchers. One of the reasons is the crucial role law has in a society forming a state[6]. Deviations from the proper functioning of the justice system – errors – do question not only the reliability of justice, but also the working of the institutions of the state. Therefore, to have a well-functioning criminal justice system seems, to a certain degree, to mirror the functioning of the institution/state proposing it[7].

On a more individual level, a functioning justice system guarantees each person that conflicts between him and other individual/s, taking the form of criminal offences for example, will be solved according to a predetermined and impartial rule or law for everybody in the given social and political context. If there is a flaw in the system, the solving of conflicts between individuals is not forcibly given any more, nor the equity of treatment. The compulsion to follow its rules tends to diminish, which in its turn has a deteriorating effect on general compliance to the rules of law. In other words, errors in the justice system might compromise its functioning by undermining the acceptance of individuals to follow its rules.

---

[4] "an expression of hypotheses that have been generated to explain observations" (Jackson et al., 2006; p. 37)

[5] "an expression of the magnitude of the LR [Likelihood Ratio]" (Jackson et al., 2006; p. 38)

[6] Law: "Body of rules, enacted by public officials in a legitimate manner and backed by the force of the state." (Neubauer (1999, p. 519).

[7] The universal declaration of the human rights articles 6-8 and 10-12, especially 10 and 11 insist on the importance on the right of "fair trial". (http://www.un.org/Overview/rights.html) The various international tribunals are a further element to demonstrate the importance of well-functioning judicial system. Thus, Cassese, president of ICTY (1995, on the conclusion of the Dayton peace agreement held): *"Justice is an indispensable ingredient of the process of national reconciliation. It is essential to the restoration of peaceful and normal relations between people who have had to live under a reign of terror. It breaks the cycle of violence, hatred and extra-judicial retribution. Thus Peace and Justice go hand-in-hand."* (http://www.un.org/icty/glance-e/index.htm; last visited 2008/09/01)
That is perhaps one of the reason why the UN is given more and more weight to help countries to have a well-functional justice system where the rights of the individuals are respected and old wrongdoings are corrected as illustrated by the International Criminal Tribunal for the former Yugoslavia (ICTY), the International Criminal Tribunal for Rwanda (ICTR), the United Nations International Independent Investigation Commission (UNIIIC), the Extraordinary Chambers in the Courts of Cambodia (ECCC), and the Comisión Internacional contra la Impunidad en Guatemala (CICIG). The political dimension of a well-functioning justice system can not be denied.

Criminal Law, as known today, depends on the concept of modern state (boundaries and one source of power) and the disappearance of private justice (Killias et al., 2008, chapter 1). Two main "types" of justice systems developed. They are known as Common law and Continental law.

Common law can be defined as: "Law developed in England by judges who made legal decisions in absence of written law. Such decisions served as precedents and became "common" to all of England. Common law is judge-made, it uses precedent, and it is uncodified." (Neubauer, 1999, p. 516)

Continental law, in contrast, has developed based on Roman Law and has a written code, e.g. criminal law which has been established by the legislative. Besides the codified, non-codified distinction, there are others. One is that common law is linked to an adversarial or adversary system, whereas continental law tends to be more inquisitorial. A further difference exists in the role of the judge, who tends to ensure fair play in the first case and helps determine the facts in the second. Moreover, in the adversarial system the prosecutor is seen as one of two equal parties represented in court, where the one with the best arguments will win. In contrast, in the inquisitorial system the prosecutor is seen as a neutral entity presenting the facts to the court. With the introduction of written statutes in Common Law countries and for instance "plea bargaining" in continental law countries, distinctions are becoming less clear-cut (Dongois, 2006).

Swiss criminal law is mainly based on continental law (Killias et al., 2008; chapter 1). There is a penal code dating from 1937, and accessory laws and regulations, implemented in 1942 and a thoroughly revised form has been in force since January 2007 (www.admin.ch/f/rs/3/311.0.fr.pdf). For the moment each Canton still has its own code of criminal procedure. However, a unified code will be introduced in 2011. Some of its changes, for instance having access to a lawyer from the first hours a person is questioned by the police, are expected to provoke considerable modifications of the functioning of the whole criminal procedure.

The interest of the differences of the common law and continental law system in the context of judicial errors lies in the fact that specificities of both systems have been enumerated both for contributing to and preventing of judicial errors, but also for facilitating or rendering difficult their discovery. In a simplified way, the adversary system has been accused of contributing more than the inquisitorial system to judicial errors, by provoking more errors by its (mal)functioning but also by offering an increased ease to discover them (Killias, 2006). Another factor has been the different ways the forensic science evidence can find its way into court, the concept of «admissibility»[8]. A more detailed discussion of these points can be found in Giannelli (2002) and Schwartz (2005). In the context of this research, the difference between legal systems will only be considered and discussed if of direct relevance to the topics studied.

8 Some jurisdictions, especially the USA system evaluate whether traces can be used as physical evidence in court or not, based on criteria such as Daubert (and related decisions), Frye, Federal rules of evidence, etc. (Kaufman, 2001).

Judicial error, wrongful conviction, and miscarriage of justice

These are three expressions/terms that describe malfunctioning in the criminal justice system. For the present thesis, **judicial error** will be used as a general term. Wrongful conviction and miscarriage of justice will be subcategories defined after the presentation of the concept of judicial error, which is seen as a general expression to describe a malfunction in the justice system. The term includes two aspects: the judicial context and the notion of error. The necessity of the former has already been described earlier on; the focus will now be on the second. In law, error can happen mainly for two reasons: errors on the facts, or the application of the law/procedure (due process in the USA). In the first case, a legal decision is taken, although later on it is established that it was based on facts that have to be revised. One example would be a new exculpatory DNA profile analysis on a biological specimen having been examined for blood groups showing consistency. In the second case, the law has been applied inappropriately because one or another rule/law has not been considered. One example would be the refusal of a justified (scientific) counter-expertise.

Accepted grounds for an appeal vary from one legal system to another. So do the definitions of judicial error, wrongful conviction, miscarriage of justice, and other related terms (Roberts, 2003b; Gross, 1998; Naughton, 2005; Huff and Killias, 2008). The possibilities to review errors and change the judgment can also vary considerably between criminal justice systems (for an overview see Huff and Killias, 2008).

The question of what is to be understood by the terms wrongfully convicted person, miscarriage of justice, or judicial error will not be discussed in detail[9], as some difficulties arising have already been considered by Aebi (2003), Jehle (2003), Killias (2003), Naughton (2005), Rattner (2003), and Walker (2003). Generally, the definition adopted by the International Network of Researchers on Wrongful Convictions (INRWC, 2003) will be used here, unless mentioned otherwise:

> "to deal only with factual errors, those cases in which people have been found guilty, after having exhausted all available appeals, and have been convicted although they were indeed innocent"[10] (INRWC, 2003).

Wrongful conviction will be used when referring to specific cases, which do fulfil these criteria. Otherwise, either the more general term of judicial error or the term miscarriage of justice as introduced below will be used. Some of the advantages of this terminology are to be sure that only recognised cases of wrongful convictions are included, thus diminishing potential criticism on objectivity and inclusion criteria. The selection of cases, so as to know whether a judicial error has happened objectively or not, is also easier.

---

[9] It does not directly influence the question of erroneous forensic science, as in a broad view one could be interested in all possible errors, regardless of whether they do or do not have a consequence. This point was raised by Itiel Dror in a discussion with the author.

[10] As well as exonerated formally by a legal procedure; this excludes cases as described by Aebi (2003) wherein somebody agrees to be convicted for a crime he/she did not commit (in the context of ETA terrorism in Spain).

Miscarriage of justice will be used for cases where something went wrong, but where the criteria defining the wrongful conviction do not apply. It is a residual category that includes (admitted) errors in the pre-trial stage (e.g. wrongful "preventive detention") or an acquittal though the person is supposed to be in fact guilty.

Examples of wrongful convictions are the Birmingham six[11] and Patrick Dils[12]. Examples of miscarriages of justice are Brandon Mayfield and Seznec. The first case does not fit the former definition, as no conviction took place; though wrongful handling of the case (including a fingermark misidentification by the FBI) has been admitted after the release of Brandon Mayfield (OIG, 2006). The Seznec case is another case not fitting the definition of the INRWC, despite decades of trials to recognise it as wrongful conviction. According to the latest decision of the "cour de revision" in 2006, no wrongful conviction has been admitted (http://www.france-justice.org/html/affaire_seznec/chronologie.htm). The definition also excludes cases where a guilty person is not convicted, be it because of the lack of evidence or formal errors, etc.

The discovery that judicial errors do and can happen is not new. Indeed, first documented cases seem to go back to ancient Greece (Otto, 2004). Since then, cases of judicial errors have been reported from time to time (Otto, 2004). Case compilations started to be more current in the 20th century, as illustrated by the historical overview of research on judicial errors in the USA by Leo (2005). The subject will be developed more in Chapter 2. The amount of cases of judicial errors happening has not yet been determined. Estimates range from 0.5 % up to 15% of all cases (Huff, 2008, chapter 4; Naughton, 2003; Poveda, 2001; Risinger, 2007). In a survey by Ramsey (2003) the error percentage varies from 0.5% up to 5%, depending whether law enforcement, prosecutors or defence lawyers were questioned. The results will also vary in function of the definition given for judicial errors.

Without entering into the details of different definitions of judicial error from one jurisdiction to another, it can nevertheless be understood that it is not straightforward to answer the question whether the amount of judicial error depends on their actual happening, on the openness with which errors are accepted or on the structural means (such as revisions, reviews etc), that actually facilitate the discovery of errors[13].

As to the reasons leading to judicial errors as mentioned by Huff, Rattner, and Sagarin (1996), MacFarlane (2003) and The Innocence Project (Scheck et al., 2003) they correspond to those given by Saks and Koehler (2005) in the figure below (Figure 3)[14].

---

[11] Six men were accused and tried in 1975 for the Birmingham pub bombings, which were attributed to IRA terrorists. Their conviction was declared unsafe and overturned in 1991.

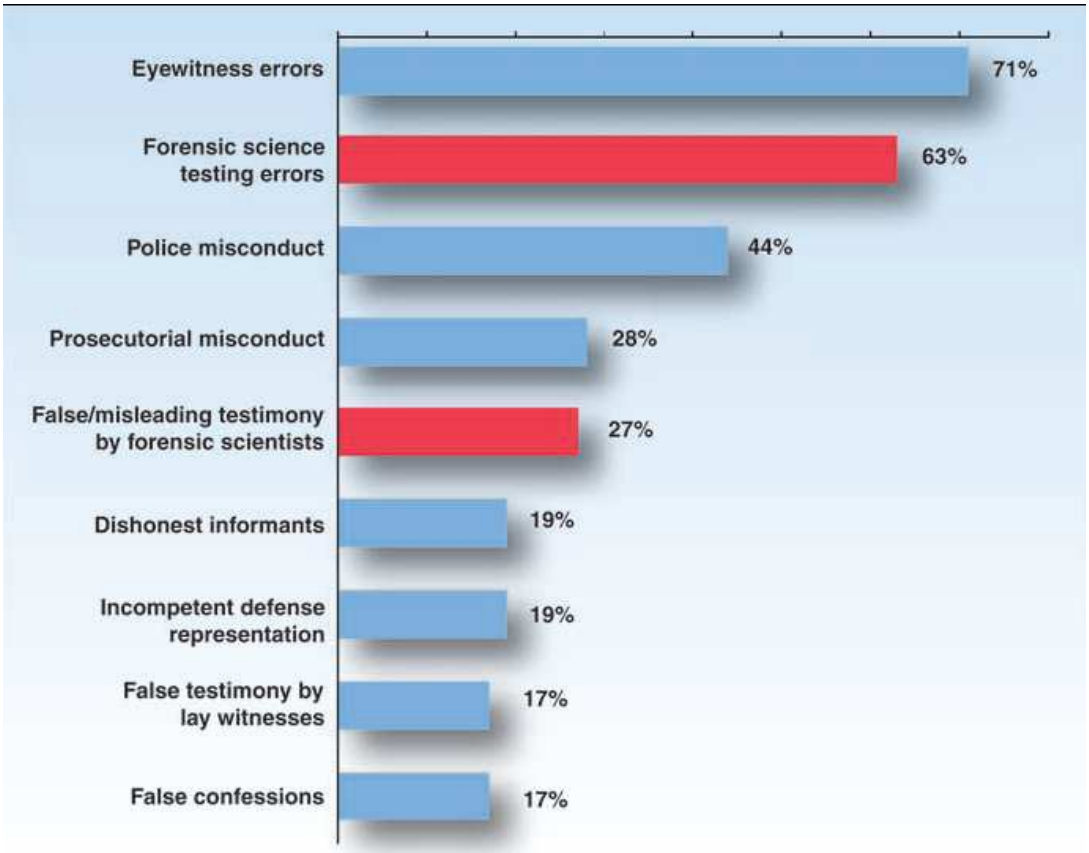
[12] Accused and tried of having killed at the age of 16 two young boys in 1989. In 2002 a judicial error was admitted and Patrick Dils freed after 15 years in prison.

[13] For more about this see Huff and Killias (2008) especially chapters 5 (Zalman, 2008), 8 (Killias, 2008), 9 (Brants, 2008), 11 (Kessler, 2008) and 12 (Dongois, 2008).

[14] For more details see Chapter 2.

Summarising these results, it can be stated that at least one third of the factors seem to be linked to forensic science[15]. This will be discussed in more detail next.

Figure 3 The coming paradigm shift in forensic identification science. Saks and Koehler (2005, p. 892-895)[16]



This table illustrates a list of factors often cited as sources of judicial errors, highlighting the factors associated with forensic science. Two kinds of forensic science errors seem to be important: testing errors (by inference work done in the laboratory) and testimonies given by forensic scientists (by inference the evaluation and presentation of forensic science as evidence on court). As will be discussed later, the first stage of forensic science work, the gathering of forensic science traces at the crime scene, is not mentioned here.

Forensic Science

First, a short overview on forensic science will be given[17]. Second, the special interest of identification evidence for this research will be explained. Third, its contribution to judicial errors will be presented.

---

[15] However, a recent publication seems to contradict the findings by stating that the incidence and impact of forensic science has considerably been overstated in the previous reports based on data from the Innocence Project Collins and Jarvis (2008).

[16] See Collins and Jarvis (2008) for a critical review of this table.

[17] More detailed information can be found in Inman and Rudin (2000) and other introductory works to forensic science.

Forensic science is an instrument to help the justice system. Definitions tend to include this concept in some form or another, though there seems to be none which is generally agreed upon (Inman and Rudin, 2000). One inspired by Margot (personal communication) will be used here:

> "The body of scientific principles applied to traces used to help Justice determine whether a crime has been committed, identify its perpetrator(s) and victim(s), and determine his/their modus operandi."

Forensic science covers a wide range of areas such as handwriting, fingerprints, drug or DNA profile analysis. All of these areas are used to answer questions of the following kind: who was at the crime scene? What happened and by whom? Why/how did it happen? Where and when did it happen?

Fields of forensic science allowing an answer to the question of who imply that there is a means of linking a (crime scene or offence-related) specimen directly to an individual. This is possible for example with fingermarks and biological specimens containing DNA. By extension it is also feasible to associate an object as source of a trace, for instance ammunition shot by a given firearm or shoe wear mark impressions left by a given shoe sole. Other areas imply the analysis of substances, e.g. to find out whether the white powder found is an illicit substance or not. However, to connect a person directly to a (crime scene) specimen is among the strongest demands on forensic science. Errors in these areas could therefore have important consequences, which is why special interest will be given to them in this research. A second differentiation is sometimes made, in terms of "soft" and "hard" (forensic) science (McClure, 2007). Hard sciences are considered those that rely on methodologies and techniques similar to the ones used in fields such as chemistry. They include DNA profile analysis. Soft sciences are understood to be the ones relying on less objectively measurable "sciences", such as visual observation needed for fingermark comparison. The second are often perceived as being more vulnerable to errors than the first (Risinger et al., 2002).

On a pragmatic note it is not possible to discuss all forensic science areas in detail, thus a choice has to be made[18]. To select areas of forensic science with the strongest individualisation character, accepted court use, and representing both hard and soft sciences seems a logical option. Thus fingermarks, with its heavy reliance on human comparison abilities might be considered to be more vulnerable to errors than more automated domains such as drug analysis by GC-MS (gas chromatography-mass spectrometry) (Risinger et al., 2002). In addition, its longstanding use and acceptance mark it as a respected and reliable forensic science field. However, Cole (1998; 2005; 2006a; 2006b) gives a critical overview of the history of the acceptance of fingermark evidence and thus offers a more critical perspective. DNA

[18] This is one reason why handwriting analysis has not been considered in this work, though some research into the process of examination and issues of quality has been published of lately, for instance by Found (Found and Rogers, 2008).

profile analysis, on the other hand, has become part of forensic science in the late 1980s (Aronson, 2005)[19]. It represents the counterpart, a "hard" science of individualising character. Its very technical nature and its statistical and probabilistic approach have changed forensic science considerably. One of the important effects has been the (partial) replacement of existing techniques for biological specimen such as blood, semen, and hair. By reanalysing such traces in most cases much more detailed information could be gained which led sometimes to the re-evaluation of the case. This can happen due to two mechanisms. It can be that in one specific case the necessary interest exists due to pressure by the defence, the media, family or other supporters (e.g. Innocence Project; http://www.innocenceproject.org/). But it is also possible that in a regular review or investigation of unsolved cases – so-called cold case review - (new) DNA profile tests are made that produce new information and help elucidate the affair[20].
Indeed, DNA profile analysis has been an important tool for the discovery of judicial errors; it has led to the establishment of whole programs such as the Innocence Project (Connors et al., 1996; Scheck et al., 2001). Forensic science, most exemplified by DNA profile analysis, thus led to the discovery of judicial errors, both miscarriages of justice and wrongful convictions. Once errors have been accepted, the interest switched to study the reasons that led to them. Within the forensic science field Saks and Koehler (2005) have stipulated that problems arise due to testing errors and/or misleading forensic scientist testimony (see Figure 3). Bellemare and Finlayson (2004), Cooley (2004), and Giannelli (2007) mention forensic science as source of judicial error as well[21].

One of the reasons why forensic science is of special interest within the topic of judicial errors is because it has also been advanced as being a tool to avoid them by replacing other types of evidence believed to be less reliable, such as eyewitness testimony. Thus forensic science has been attributed the double role of error source and also remedy, which has been discussed to some extent by Edmond (2002) and will be presented in more detail under 2.3.5. It is not only a double role, but shows also a direct growing reliance on forensic science in detriment of other sources of evidence. The reasons for this trend will not be discussed here. A few explanatory hypotheses could be: an increased reliance in technical fields as opposed to more social (science) topics in society; an increased critical view of other items of evidence such as eyewitness testimony; the potential influence of TV shows promulgating an idealised vision of forensic science (Cole and Dioso-Villa, 2007; Schweitzer and Saks, 2007; Starrs, 2005); and changing legal requirements. For instance, as mentioned before, the new Swiss federal criminal procedural law (in force in 2011) includes the lawyer from the first hour of arrest which is expected to reduce the amount of confessions and thus leads to an increased reliance on other

[19] First legal case is immigration case (question of "disputed paternity"); first criminal law case is mass screening in Pitchfork murder affair, though DNA was more a tool for investigation rather than needed proof in court (Aronson, 2005).
[20] http://www.ncjrs.gov/pdffiles1/nij/194197.pdf for a definition of cold cases and the use of DNA
[21] For more details see Chapter 2.

means of proof such as forensic science evidence. It seems therefore legitimate to want to know if it does this correctly or not.

Forensic science is named as one of three main contributing factors, as acknowledged in several studies (see chapter 2.3.1). More details on types and categories of errors will be given later in Chapter 2. In this research the focus will be on non-voluntary errors, as voluntary errors seem to be mainly a question of applying policies correctly to avoid (Risinger et al., 2002). This means that cases such as Fred Zain and Pamela Fish (Giannelli, 2007), examples of willing fraud, will not be discussed in detail in this work.

Criminology

Criminology[22] – a research field interested in (empirically) studying crime and the reaction of society towards it, deriving its qualitative and quantitative methodology mainly from social sciences – is prone to be interested in law(s) and the (criminal) justice system, which to some measure "create" (or at least define) crime. In doing so, it is inevitably also interested in occurrences where the criminal justice system errs, by creating a miscarriage of justice or even a wrongful conviction or execution. These individual cases often serve as a qualitative study to propose a hypothesis for an observed tendency or trend in society in general or in some of its sub-groups. As it happened for judicial error, the compilation of several of such cases can be seen as the trigger of further, more quantitative research on the topic (Leo, 2005). A feature which distinguishes criminology from other social sciences is not only the focus on "crime" related topics, but its strong vocation for multidisciplinary and comparative research, both internationally and interdisciplinary. Again, these properties do concur with the needs of any methodological tool interested in studying judicial errors. A further element that encourages the use of a criminological framework and tools for the present study is the research methodology developed to tackle sensitive issues. To discuss error sources in one's work as it is done by the semi-structured interviews presented in Chapter 4 is not without practical difficulties. By drawing by analogy on existing methodology and experiences in criminology, a better approach can be guaranteed.

In addition, a considerable part of literature published on the general topic of judicial errors has been done so by criminologists and/or published for a public interested in criminology. Some of them can be seen as being part of **critical criminology**, a movement starting in the late 1960s that criticises the integrity of the justice system and people studying it on several points, which will be discussed in more detail in Chapter 2. Areas of interest touching judicial errors discuss topics such as the effects of (long-term) wrongful imprisonment (see the special edition of the Canadian Journal of Criminology and Criminal Justice 2004 (Campbell and Denov, 2004; Denov

[22] A different definition of Criminology is given by Sutherland and Cressey, (1970), p. 3 "Criminology is the body of knowledge regarding delinquency and crime as social phenomena. It includes within its scope the processes of making laws, of breaking laws, and of reacting toward the breaking of laws. ...The objective of criminology is the development of a body of general and verified principles and of other types of knowledge regarding this process of law, crime, and reaction to crime. ...In addition, criminology is concerned with the immediate application of knowledge to programs of crime control."

and Campbell, 2004; Gervais, 2004; Hickman, 2004; Huff, 2004; Kennedy, 2004, Scullion, 2004)); the effect on society of judicial error (Walker, 2003); problems encountered by released wrongfully convicted persons and their families (e.g. access to compensation) (Hoel, 2008); the costs to society resulting from judicial error (Forst, 2003); and the possible judicial solution to minimise certain types of errors[23] (Brants, 2008). Those topics will not be developed in detail. The interaction between forensic science and judicial error - the focal point of this research - is, however, seldom mentioned by these abovementioned researchers.

In addition, criminology can also be perceived as a tool for empirically testing the underlying theories that inspire criminal law and related disciplines, which may then, depending on the results, influence "criminal policy" (Killias, 2001)[24]. Considering this view, to study judicial error – the errors of the justice system – does enter into the range of the definition. One advantage of this view is that it does directly include the possibility of change, as it implies an interaction between academia and the legal system in a given nation[25].
In short, criminology is interested, among other things, in crime and the criminal justice system and relies on social science methodology and theories to study it. This explains its interest in judicial errors. Several different schools of thought have existed and still exist in criminology. Critical criminology (Hil and Robertson, 2003; Jones, 2006, chapter 10), by pointing, among other things, at the inequalities of the criminal justice system has influenced, and perhaps to some degree triggered the initial discussions on judicial errors. It has a strong focus on the injustice lived through by the wrongfully "convicted". This point of view tends, however, to neglect another point of view: the search for causes or contributors to judicial errors, as presented next.

In summary, any (criminal) justice system, be it based on common law or continental European law, functioning according to the adversary or inquisitorial system, has a vital interest in having a small error rate, as every malfunctioning tends to erode its credibility and, therefore, its right for existence. However, the occurrence of errors is an accepted fact and several measures, such as the possibility of reviews, do offer a way to remedy them. Each nation has developed its own way to cope with this eventuality. Definitions of error do vary considerably and are limited somewhat by their legal context.

Putting all the elements defined in this chapter together – justice system, criminal law, judicial error, forensic science, and criminology – it results that the interaction judicial error and forensic science (identification evidence) can only be studied and discussed appropriately if the larger framework described here is considered.

---

[23] Which tend not to include measures to take in the field of forensic science.
[24] "[…] la criminologie est appelée à vérifier si les théories implicites qui inspirent le droit pénal et les branches juridiques adjacentes sont pertinentes ou non d'un point de vue empirique." (Killias, 2001; p. 2)
[25] The typical example here would be the political decisions on drug dependency policy taken by the Swiss government when agreeing to test heroine prescription to addicts and adapting their politics in reaction to the results of this pioneer study (Killias, 2001, p. 506-510).

1.3 Aims of the research

For this research the main hypothesis is:
Forensic science as a source (or remedy) for judicial errors has already been identified theoretically but has to be further established experimentally by

1. A literature review composed of case reviews and a review of publications to identify possible sources of judicial errors in the pre-trial forensic science work, as well as to identify the use of items of physical evidence in the exoneration phase.
2. An empirical survey protocol to carry out the testing of several working hypotheses on the nature of non-voluntary error as for instance the observer bias (Risinger et al., 2002), by using forensic science students as test population.
3. A survey among forensic science units in Switzerland and abroad in order to observe their awareness of the interaction between judicial error and forensic science.

The objective of this research is to obtain a better view of the interaction of forensic science and judicial error (incidence and impact; source and remedy), both by literature review and experimental proceedings, in order to suggest practical recommendations to minimise the occurrence of wrongful convictions.
A special interest in an experimental section is given to a specific source of possible error - the observer bias. The sub-arguments necessary to its testing as well as additional hypotheses are detailed in the chapters dealing with the three research stages.

1.4 What has been achieved – Summary of results

Most of the objectives fixed in the PhD progress report for the “Wrongful Convictions in Switzerland in a Comparative Perspective” for the forensic science portion have been achieved. They included the study of various error sources throughout the process from the crime scene to court. This has been done by a literature review (Chapter 2), an empirical approach using semi-structured interviews of senior forensic scientists (Chapter 4), as well as the study of one specific potential error source, the observer bias (Chapter 3). The three surveys proposed – covering forensic science students, Swiss forensic science units, and other forensic science units abroad – have been done, though several changes have been made, which will be explained in more detail later.
Partial answers have been found for the questions asked in the SNSF project about the role and utility of forensic science (SNSF, 2.5 Relevance of the project) in the context of judicial errors. This includes, among others, studying the interest of the introduction of quality management measures in forensic science, determining the importance of forensic science to detect and avoid judicial errors, and introducing classes on the risk of miscarriages of justice. Intermediate results have been communicated through several publications and conferences.

- Schiffer and Champod (2004). Scientific Expertise: Pondering the Contribution of (DNA) Evidence, presented at the 4th European Society of Criminology conference, 28 -31 August, Amsterdam.
- Schiffer (2005). Scientific Expertise: Incidence and Impact on Judicial Errors (Work in Progress) presented at the 5th European Society of Criminology conference, 31 August - to 3 September, Cracow, Poland.
- Schiffer and Champod (2006). Experimental Research on the Possible (Negative) Impact of Observational Biases in the Analysis Stage of Fingerprint Individualisation, poster presented at the 4th EAFS conference, 14 – 16 June, Helsinki, Finland.
- Schiffer et al. (2006). Miscarriages of Justice: The Comparative Point of View of Senior Forensic Scientists or Managers, presented at the 6th European Society of Criminology conference, 26 - 29 August, Tübingen, Germany.
- Schiffer and Champod (2007). The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation. *Forensic Science International*, 167(2-3), 116-120.
- Schiffer, B. and Champod, C., (2008). "Judicial Error and Forensic Science: Pondering the Contribution of DNA Evidence", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 33-55.

Part II – Presentation of the three areas covered by the research: literature review, empirical tests, and semi-directive interviews

While in Part I it was possible to delimitate the research topic and to give some general ideas of what will be explored, Part II explains this in more detail. On one hand, the literature review gives an overview on different types of existing research, namely academic, official (requested by official organisms) or private (done by individuals such as journalists). This helps at the same time to see where the lacunae are and to illustrate that forensic science is one of the aspects of judicial errors that still has many under-researched facets, thus justifying the interest of this work. However, one of the main observations made is that existing research tends to be descriptive, and very often not very systematic, which makes it difficult to pinpoint specific problems in the field of forensic science. One of the consequences is that practical recommendations are rather difficult to make. The literature review continues the attempt to give an exhaustive inventory of possible errors in forensic science, highlighting some of them. Those often tend to be linked to areas relying heavily on human judgment. One of them is the possible presence of observational biases in forensic science work. In order to test this assumption more closely, empirical tests (Chapter 3) were designed to be implemented in one field of forensic science thought to be specially relevant for and vulnerable to such biases, identification evidence, and more precisely, fingermarks. Results tend to show that though under some of the tested conditions biases can be observed, they can also be absent on other areas where they have been suspected to exist. For the population tested, forensic science students at the University of Lausanne, the role of education seems to play an important role for minimising errors and is therefore an interesting factor to be retained. In general, the results tend to show a certain discrepancy between comments made in literature and those observed under the test conditions. Although the design of the test and many other factors might explain (part of) these differences, they do not clarify everything.

In a third approach (Chapter 4) the opinion of professionals in forensic science as to error risks in their field were collected. The semi-structured interviews with senior forensic scientists and managers of forensic science laboratories also allowed them to summarise their view on a variety of related topics. The comparison of the professionals' point of view with the results of the literature review for instance, allowed to gain a new perception of the subject, which will be discussed in more detail in Part III, which puts all the threads together.

2. Literature review

The literature review will give a summary of publications in relation to judicial errors and forensic science. In order to do this, a list of (research) questions has been prepared, and the chapter is subdivided into sections answering those questions. At the beginning, the concept of critical criminology will be introduced (2.2), as it is proposed here that it helps to explain the focus of the publications discovered on the subject. It seems also to a certain degree to influence future directions that could or should be taken, including the need for empirical research.
After a general overview, including a comparative and chronological view of forensic science and judicial errors, an inventory of possible errors in the forensic science process will be given. One of these, observer effects (or in other words subjectivity), will be then developed in more details, as it will also be the focus point of Chapter 3, the experimental research on fingermark individualisation. The chapter also includes a publication on the asymmetrical way forensic science used to inculpate or exonerate is considered (2.3.5); it will close with some comments on methodological issues (2.4).

2.1 Research questions asked concerning the literature

After a first exploratory overview on existing publications and bearing in mind the considerable variation of these[26], the literature review was structured according to some questions or topics thought to be of special interest to the research. The aim was to establish and study:

1. the status quo of publications concerning the interaction between forensic science and judicial error and the lack of recent publications by forensic scientists on the risks of errors in forensic science
2. an inventory of various types of errors (error sources) in forensic science leading to judicial error[27]:
   (1) errors in the collection of specimen or errors (e.g. confusions) in the manipulation of forensic traces;
   (2) errors in the choice and performance of technical tests;
   (3) errors in the establishment of factual findings and interpretation of these findings in a context;
   (4) errors in the presentation of evidence to non scientists (typically in the courtroom);
   (5) errors due to various psychological reasons.
3. the lack of literature on the decision-making process in forensic science
4. the lack of empirical studies of the observer (bias) effect (Risinger et al., 2002) in forensic science

[26] See the PhD progress report (Schiffer, 2005)
[27] Without considering the question of causality; based mainly on the various case reviews studied and the inventory of error sources given.

5. the fact that people tend to be less critical of exculpatory than of inculpatory evidence (especially when a wrongful conviction is supposed), therefore there is a need for increased symmetry (Edmond, 2002).

2.2 Theoretical outline

Several schools of thought exist in criminology and have been described in introductory works (Jones, 2006; Killias, 2001). One of them, called critical criminology, seems particularly interested in the malfunctioning of the criminal justice system, and by extension in judicial errors. Why this might be so and how it might have shaped the existing publications discussed in this research will be exposed in the following. However, the aim here is not to discuss "trends" in (critical) criminology, nor their potential benefits or disadvantages, but rather to offer some explanations as to why existing publications can be better understood if those concepts are kept in mind[28].

Critical criminology[29] starts as a movement in the late 1960s and early 1970s. It is characterised by a certain loss in faith in the state as unfailing institution and in consequence also in the justice system being just. According to Cullen and Agnew (2006, p. 294ff) and Jones (2006, p. 247ff), five main points were criticised. First, the state protects the inequality of power and means through its institutions, the criminal justice system being one of them. Second, the definition of crime is political; it reflects power structures and is therefore to some degree subjective. Third, the existing system defends the existing social order. "The rich get richer and the poor get prison." (Cullen and Agnew, 2006 p. 295). Fourth, "capitalism is the root cause of criminal behavior", (Cullen and Agnew, 2006 p. 295). Fifth, criminologists should actively help construct a more equitable society, the solution to crime. In this perspective, criminology has been understood differently. "In the process of redefining crime, criminologists will redefine themselves, no longer be the defenders of order but rather the guardians of human rights." (Cullen and Agnew, 2006 p. 295). This last concept is stressed by Hil and Robertson (2003) to be one of the future directions that should be, but has not yet been taken.
Judicial errors can be seen as an example of what is of interest for (critical) criminologists and therefore relevant for them. In other words, it is advanced here that it offers a possible theoretical framework for studying the subject. Out of the five critical points mentioned before, two main arguments will be considered in more detail.

---

[28] Readers interested in those questions are invited to refer to Cullen and Agnew (2006, p. 294ff), Hopkins Burke (2005, 173ff), Jones (2006, p. 247ff), and special journals (e.g. critical criminology). Here a very simplified representation is given, due to the criteria mentioned before.

[29] It is difficult to find an agreed upon definition of critical criminology. Hil and Robertson (2003) discuss this and comment on p. 98: "The business of being critical [criminologist] appeared to boil down to highlighting those policies and practices in crime control that are repressive and harmful; locating changes in cultures of crime control in historical contexts and pointing to the emergence of new discourses, regimes, technologies and practices of punishment."

First, the existing criminal justice system benefits some actors and disadvantages others, which is related to a more general social (in)justice phenomenon. In the USA, racial bias in the justice system would be one example. One expectation created is that the focus of research and publication on judicial errors will be on subjects related to social (in)justice. The special edition of the *Canadian Journal of Criminology* (Denov and Campbell, 2004) illustrates this point. It is argued here that the interest on judicial errors grew, among others, out of the critical criminology perspective, which may give it a certain context and direction to be considered when looking at the publications by these contributors. In other words, by perceiving judicial errors as injustice produced "willingly" by the state and its institutions to protect the rights and chosen privileges of a minority, the causes and sources of judicial error tend to be attributed to a general malfunctioning of the (criminal justice/social) system, rather than to other explanations. It is felt here that one major drawback of this approach is that it does make amendments rather difficult, as they depend on the political motivation to modify something, which according to the critics has been done voluntary by the institutions, being those in charge of the situation. This view can explain the tendencies of existing research, which is focussing on the social characteristics of the victims of judicial errors and the institutional errors that have been committed. This is especially visible on research in the context of wrongful executions, the worst consequence of judicial errors in jurisdictions which carry out the death penalty (Cooley, 2003a; Cooley, 2004).

Second, according to Cullen and Agnew (2006) and Hil and Robertson (2003) (critical) criminologists can be expected to guard human rights. This raises the questions of how judicial errors relate to human rights and if this guarding human rights perspective allows for a different view of the subject, as opposed to the social (in)justice perspective.

> "Certainly a human rights perspective, as embodied in zemiological analysis, may well provide fertile ground for a more wide-ranging analysis of crime and criminality and it may encourage critical criminology to more adequately address injustices at various institutional levels. Additionally, this may further serve to re-focus critical criminological inquiry in a direction that is not merely descriptive of occurrences in crime control but which also questions forms of political persecution from a human rights perspective. It is arguable that in focussing so heavily on crime control per se, and by retreating from many of the political agendas present in earlier versions of radical analysis, critical criminology has tended towards an overly descriptive analysis of what is rather than more prescriptive positions about what ought to be."
> (Hil and Robertson, 2003, p. 100)

The message from the critics of critical criminology seems to be that more practical recommendations, rather than only analytical and descriptive research should be provided by critical criminologists (or in this context anybody interested in the research topic). This point of

view offers an interesting criterion for reviewing papers proposing amendments to minimise the occurrence of judicial errors due to forensic science and other related topics.

2.3 Methodology and results

Methodology included a thorough research by internet and the following up of references in (key) articles. Priority was given to more recent publications. After an initial overview of publications, research questions (2.1) thought to be interesting were asked. Some answers found by extensive literature review have been summarised in the following.

2.3.1 The status quo of publications concerning the interaction between forensic science and judicial errors and the lack of recent publications by forensic scientists on the risks of errors in forensic science

This section will first give an overview of publications which discuss judicial errors focusing on error sources and remedies. Studies on the effect of judicial errors for individuals wrongfully convicted, as well as for society as such are not included[30] as they usually do not discuss forensic science, the focal point of interest of this research. It starts with general publications and goes down into further details; more specific publications are not listed in the more general categories.

Roughly, two main types of publications can be distinguished: those concerned with single or singular cases and those discussing the subject more generally. The first group can further be subdivided into official, respectively private reviews. The second group includes the following areas: the quality of forensic laboratory work in general (training, proficiency tests, guidelines (lack of), codes of conduct; ethical considerations, newspaper publications); the complex interaction between science and law (misunderstandings, different philosophies, different goals, presentation of evidence, expert witness); specific propositions as to error sources not directly related to the interaction between law and science (observer bias, non validated methods); and other subjects[31] (asymmetrical criticism of science to inculpate and exculpate, European view on forensic expertise, post-conviction DNA testing).

(A) Succinct overview of the status quo of publications

The purpose here is to give a short and neutral overview of a maximum of different topics discussed in the context of judicial errors (and forensic science) including the corresponding

30 The special edition of the *Canadian Journal of Criminology*, 2004 (Volume 46), for instance Huff (2004) gives a good introduction into these issues.

31 Only those most interesting for the research are mentioned.

references so as to allow a quick access to potentially interesting publications. The aim was also to identify some particularly interesting publications, though an overview is never exhaustive. How this collected literature has been understood in this research will be presented under (B), allowing the author a more personal view of the topic.

General reports on judicial errors include either the study of individual or an overview of several cases. Some of the cases will be developed more in part (B) rather than in the short overview. Examples of publications thought to be of special interest will be given for some countries, including those that publish most on the subject.

- Australia: One of the earliest (official) publications on judicial errors comes from this country (Morling, 1987)
- Britain: The focus of studies presented here is on forensic science and official review bodies, such as the CCRC and SCCRC (R v. Ward, 1993; Roberts and Willmore, 1993; SCCRC, 2005; Zellick, 2004)
- Canada: Have among the most comprehensive case studies, both for individual cases or overviews (Bellemare and Finlayson, 2004; Kaufman, 1998; MacFarlane, 2003)
- USA: Have a variety of publications, including official reports on specific cases (Connors et al., 1996; Parks, 2000; Ryan, 2002)

(Nearly) all publications interested in general issues of judicial errors are by non forensic scientists[32]. They have been categorised for a better overview of subjects that seem to be of interest to them.

The (legal) definition of judicial errors, wrongful convictions, and miscarriages of justice

This issue has a certain importance, as it influences the type and amount of cases considered, thus also influencing potential research on the causes and the proposition of remedies. The INRWC (2003) chose a restrictive definition – wrongful conviction, which is also used in the present research - to avoid discussions whether a case is to be included or not.

On the other hand, information on all errors, independently of their consequences could be interesting to know.

(Aebi, 2003; Dongois, 2008; INRWC, 2003; Naughton, 2005; Roberts, 2003b)

Errors in a specific country or justice system (causes)

Publications include descriptions of the situation in several countries, be it the structure of the judicial system (Brants, 2004) and its influence on judicial errors or, for instance, studies on a series of cases for a better understanding of underlying phenomena (Gross et al., 2005). Huff and Killias (2008) are the authors of a book that includes several chapters on this topic.

(Brants, 2003; Brants, 2004; Brants; 2008; Gross, 1998; Gross et al., 2005; Gualco, 2003; Huff et al., 1996; Huff, 2002; Huff, 2003; Huff and Killias, 2002; Jehle, 2003; Killias, 2003; Killias et

[32] Based on different criteria, such as academic background (studies), current enrolment, and journal of publication; this includes (legal) scholars interested in the Criminal Justice System.

al., 2004; Naughton, 2003; Ramsey, 2003; Rattner, 2003; Saks et al., 2001; Scheck et al., 2001; van Koppen, 2008; Walker, 2003)

A discussion on specific error sources (overview)

Error sources excluding forensic science have not been taken much into consideration in the present research.

However, some of these publications do summarise a specific area in a very exhaustive way, such as Gudjonsson (2002) does for unreliable confessions.

(Campbell, 2003; Grometstein, 2008; Gudjonsson, 2002; Rattner, 1995; Roberts, 2003a)

A discussion on (functioning) remedies to judicial errors

This includes publications on propositions on how to minimise errors (Forst, 2003), the study of existing remedies such as the Criminal Case Review Commission (CCRC) (Nobles and Schiff, 2001) or the comparison of such systems (Quirk, 2007).

(CCRC, 2005; Forst, 2003; Koppl, 2005; Nobles and Schiff, 2001; Saks et al., 2001; SCCRC, 2005)

There is a certain amount of publications by non forensic scientists on judicial errors and forensic science. Forensic science practitioners seem (until now) not to have published much on these topics.

Publications by non forensic scientists on judicial errors and forensic science comprise:

General issues

Publications include a discussion of possible error sources in forensic science. Risinger et al. (2002) evoke observer effects, which will be discussed in more detail in Chapter 3. Jakobs and Sprangers (2000) present one of the rare European papers on the subject.

(Bryson, 1985; Cooley, 2003a; Cooley, 2003b; Cooley, 2004; Edmond, 2000; Edmond, 2002; Freckelton, 1989; Freckelton and Selby, 2002; Gans and Urbas, 2002; Giannelli, 2001; Giannelli, 2002; Jakobs and Sprangers, 2000; Johnson and Williams, 2004; Leahy, 2001; Loftus and Cole, 2004; Maguire, 1993; Risinger et al., 2002; Saks, 1989; Schwartz, 2005)

Fingermarks

As mentioned in the introduction, fingermark individualisation has been of special interest as an error source in the discussion of judicial errors, among others due to the Mayfield affair (US Department of Justice, 2006) and the Stephen Cowans exoneration (Cole, 2005), which led to a certain amount of journal articles as well.

(Cole, 1998; Cole, 2005; Cole, 2006a; Cole, 2006b; Cole, 2008a; Cole, 2008b; Dror et al., 2005; Dror and Charlton, 2006; Dror et al., 2006; Epstein, 2002; Gysin, 2005; Lewis, 2001; Meili, 2005; Mnookin, 2004; Mnookin, 2001; Saks and Koehler, 2005)

Newspaper articles

This contains a sample of articles published in daily newspapers on errors in forensic science, showing that awareness of possible errors exists to a certain degree.
(Anonymous, 2004; Daley, 2004; Johnson, 2003; Kershaw and Lichtblau, 2004; McDougall, 2004; Saltzman, 2004; Saltzman and Daniel, 2004; Wax and Schatz, 2004; Zaitz, 2004)

As already mentioned, there are nearly no publications by forensic scientists on judicial errors and forensic science as well as error sources in forensic science. As to errors in forensic science, papers tend to discuss that proficiency tests (Haber and Haber, 2004) have shown worrying results. Other topics are the difficulty of communicating results to non-scientists, especially if related to DNA profiles.
(Champod et al., 1999; Broeders, 2006; Bromwich, 1997; Haber and Haber, 2004; Kerstholt et al., 2007; Langenburg, 2004; Moenssens, 2000; Penacino et al., 2003; Peterson et al., 2003a; Peterson et al., 2003b; Randerson and Coghlan, 2004; Taroni and Aitken, 1997; Thompson et al., 2003; Thompson, 1995; Thompson and Cole, 2005; OIG, 2006; Wertheim et al., 2006)

In summary, a status quo of publications concerning the interaction between forensic science and judicial errors has been established, with special focus on fingermark individualisation as a representative of established identification evidence. It has been observed that publications are nearly all written by criminal justice scholars and not by forensic scientists. The same applies to the discussion of error sources, with the exception of proficiency tests and related topics.
In other words, a certain lack of information on the interaction or relationship between forensic science and judicial error can be observed, which confirms the interest of the present research.

(B) Chronological overview

In the preceding section judicial errors have been summarised in several categories of interest. In the following, an overview of these categories will be given, considering the cases in a comparative perspective, for instance between countries.

Although judicial errors have been known to happen nearly as long as legal sanctions and proceedings have been a part of human society (Otto, 2004), the arrival of scientific evidence, as it is understood today, is rather recent. It is in the second part of the 19th century that its utility started to be perceived more generally and new techniques were discovered and developed (Eckert, 1997; chapter 2). One example of this progress was the comparison of fingerprints left on a crime scene with inked fingerprints of suspects (Ashbaugh, 1999; chapter 2). Shortly after the introduction of scientific evidence in court errors were discovered, and even the existences

of judicial errors due to this novel item of proof were revealed. A famous case example is the Dreyfus affair in France (Champod et al., 1999).

Some problems arise

These cases in the early days of forensic science reveal some problems that will become characteristic for this new type of evidence[33]. The first one is the acceptance of a new type of evidence in comparison with established methods. In the 19th century it is the emergence of scientific techniques which tends to diminish the importance of testimonies, the conventional "preuve par excellence". The growing interest and strong belief in the reliability of scientific reasoning and research findings in society reflects itself thus in the criminal justice system. However, all of these novel techniques had to be tested as to their technical and scientific reliability, as well as their usefulness as legal proof. How well this scrutiny has been done has raised controversial opinions in both the disciplines of fingermarks and DNA.

For fingermarks, the questions of technical and scientific reliability were considered in the early twentieth century. How well this was done is discussed by Cole (1998), Epstein (2002), Haber and Haber (2004; 2008a), Zabell (2005). They claim that the scientific foundations of the discipline are (partially) lacking – among others due to the police background - and that its acceptance in court was based on purely legal arguments.

For DNA profiles, the questions arose in the 1980s. It has been less subjected (until now) to this criticism (Lynch, 2003), as it is often claimed that it underwent a rather close scientific scrutiny due to its "real science background" in biology (Jobling and Gill, 2004) before being accepted in court. However, Aronson (2005) asserts that, when DNA profile analyses were first presented in court, the scientific foundations were not yet very well established, nor the tests judged to be completely validated. In the following years, the "DNA war" clarified a lot of technical, scientific, and legal (admissibility) issues related to DNA profile creation, matching and frequency estimation, as detailed by Thompson (1993). The paper proposes also what can be done in future to evaluate concerns raised in an adequate way. Lander and Budowle (1994), by their joint publication covering the defendants and critics position on the way DNA profiles were used, efficiently quenched the "DNA war" debate. Lempert (1997), on the other side, feels that still some issues have to be solved.

Furthermore, even if acceptance in the relevant scientific community exists, this does not forcibly imply successful and useful judicial use. For instance, DNA profile comparisons might work well, but not for the specific forensic traces examined, or might not respond to the questions the court asked.

[33] It has occurred in the past that some types of "expertise" have been used in court that have later been discredited for not having any scientific/technical foundation and thus legal admissibility and might in some cases even have led to erroneous judgments. Therefore, the initial test of admissibility in court has its importance, but went through some difficulties, even for recognised areas such as DNA and fingermarks.

In other words, in both areas, although at different times, it was the positive impression of the new technique that allowed its use, rather than real legal or scientific testing. In fingermark individualisation, the fact that the visual comparisons made were easy to understand and "reconstruct" in court, seemed to have facilitated its introduction. For DNA, it was rather the perceived potential of the technique that did the trick. Indeed, Aronson (2005) stipulates that it was rather the investigative[34] than the evaluative importance of DNA that was of interest in the two first UK cases, where this type of evidence was first presented. It was therefore not the novel scientific evidence presented in court that was important, but its previous use in the investigation. Nonetheless, the message that got passed was that DNA profiles were an accepted and reliable new type of evidence that had taken its initial hurdle in court.

Similar observations might be true as well for other countries and affairs (Beyleveld, 1997; Gans and Urbas, 2002). Analogous tendencies can be observed in other domains as well, for instance ear mark comparisons (Champod et al., 2002). Therefore, one point to keep in mind is the difference between scientific acceptance of a new method among the "relevant scientific community" (which was, in the case of DNA profiles, not yet achieved at the moment the evidence was presented in court) and the acceptance by the relevant legal gatekeepers. The needs and expectations of both are different and do not forcibly overlap.

If comparing both fingermarks and DNA, the former is criticised more heavily. A further explanation for this difference could be that fingermark individualisation was the introduction of a new technique[35] that allowed gathering novel evidence for trials, whereas DNA profile comparison replaced already existing blood group tests, hair comparisons, etc. The evaluation of the advantages of a new method compared to an already existing one and the need to show the gain of the new method might well be part of this perceived more "scientific" scrutiny of DNA in contrast to the introduction of something completely new, where nothing comparable existed before, as in fingermark individualisation[36].

However, aside from these more general questions of scientific or other foundations for the use of forensic science, its quality has to be and will be questioned as well. Problems range from outright fraud, examples of which are Zain and Fish in the USA, to too affirmative or incomplete reports or simply non-voluntary errors of all type (Cooley, 2003a; Giannelli, 2002; Giannelli, 2007; MacFarlane, 2003; Risinger et al., 2002). They will be enumerated and discussed in more detail later in this chapter. For now, the focus will be on the effect of forensic science in general, whether faulty or not, on convictions and exonerations in different countries.

---

[34] Investigative: forensic science used to advance police investigation; evaluative: forensic science as legal proof in court

[35] Keeping aside the Bertillonnage or its previous use as identification tool of habitual offenders

[36] Bertillonnage has been more useful for finding recidivists than for crime scene work. Furthermore, there is a lot of research going on for DNA profile analysis outside the field of forensic science, partly due to other possible applications. This was less the case for fingermark comparison.

Serious troubles

In the 1970s and 1980s most Anglophone countries have been subjected to important and heavy mediatised cases of accepted and potential judicial errors. The report by Bellemare and Finlayson (2004) *Report on the Prevention of Miscarriages of Justice in Canada* gives a full overview of the most important studies on wrongfully convicted individuals in the USA, Canada, the UK, Australia and New Zealand[37]. Several hypotheses may be advanced as to why these nations seemed to have experienced more cases of judicial error than others[38]. The simplest one could be their greater willingness to discuss the problem – and thus a greater visibility of the phenomenon - followed by explanations considering the legal system and its possible influence on the occurrence of errors (Brants, 2008; Dongois, 2008; Huff, 2003; Huff, 2008; Kessler, 2008; Killias, 2008; Mathieson and Gross, 2003; Zalman, 2008). However, as no comprehensive comparison has been made yet between different countries, this point cannot be clarified at the moment. Even if Anglophone countries have published most, an overview of newspaper articles shows that this topic is universal[39] (see also 2.3.1).

As to the **causes of judicial errors**, the list is long. However, a rough classification into three groups will be attempted here. The first is the gathering of evidence to construct a case. This includes police work, eyewitnesses, forensic science, etc. The second are the legal structures and safeguards in place to ensure the (highest) quality of court proceedings. The third comprises public opinion, be it media or political pressure, as well as socio-historical movements (e.g. witch-hunt against paedophiles (Grometstein, 2008)). They have not been detailed here, as the following authors give a comprehensive overview (Gross, 1998; Huff, 2008; MacFarlane, 2003; Scheck et al, 2001). Furthermore, the focus of the research is on forensic science and not the other topics. Besides knowing why errors happen, it is of interest to know how many of them exist. In the following several countries and regions will be compared as to their incidence and specificities of judicial errors.

As to the **incidence of the phenomenon**, the **USA** count at least 672[40] acknowledged exonerations. Cautious estimations tend to consider that wrongful convictions might happen in 0.5-1% of all cases (Huff, 2003; Ramsey, 2003). Concerning the number of cases that will or should be reviewed due to the discovery of important malpractice in a laboratory for instance,

---

[37] This and the next chapter will rely in part on this study as it gives a good overview of the research done on judicial errors.

[38] More exhaustive lists of judicial errors can be found for instance on the internet http://en.wikipedia.org/wiki/Miscarriage_of_justice or in books (e.g. Otto, 2004), though here the focus is on cases that have been officially been acknowledged as such (and published in academic or official form) and correspond to the definitions used all along this text.

[39] For instance, in China a man had been convicted for the murder of his wife. Eleven years later she reappeared in the village where her daughter lives. She just went away one day to get married to somebody else in another province. It was the finding of a badly decomposed body of a woman that was wrongly identified as his wife, as well as the police interrogation, which "helped" to produce a (wrong) confession that seem to be the key elements in this wrongful conviction (NZZ online, "Folgenschwere Schlamperei der Justiz in China", 06.04.2005)).

[40] This is composed of 328 exonerations mentioned by Gross et al. (2004) from 1989-2003, 135 and 70 exonerations in specific scandals they did not include and 139 affairs from 1900-1986 as mentioned by Bellemare and Finlayson (2004).

this amounts to over 3000 incidents to be reviewed in the case of the FBI laboratory and over 1000 cases for the Houston laboratory alone (Cooley, 2004; Bromwich, 2007), which are not included in the estimate of Huff (2003).

In **Canada**, at least 5 cases have been acknowledged since 1986 and new ones are added continuously. One of the latest cases involved Steven Truscott, being convicted of raping and killing a schoolgirl. The conviction was overturned in August 2007[41].

In **Australia**, the number of acknowledged incidents of judicial errors amounts to at least five as well as at least one case in New Zealand (Bellemare and Finlayson, 2004). However, as the topics of discussion are scarcer in these countries and are similar to the Canadian situation, they will be discussed together.

The **UK**[42] has acknowledged over 200 judicial errors[43]. The Criminal Case Review Commission has referred 370 cases for re-judgement since 1997 (260 quashed, 109 upheld, 1 reserved)[44]. Estimations by Naughton (2003) stipulate that the number of judicial errors could be several thousands per year if "routine miscarriages of justice" (quashed convictions at the Crown Court level) were considered as well. Naturally, this number will depend heavily on the definition given to the terms (see earlier comments). However, the question that still has to be decided, not only in the UK, is whether the cases or the people involved are to be counted. The difference might be quite important, as the IRA bombings illustrate. Indeed, four cases could be counted, which involve at least 18 persons[45].

For **Continental Europe** the situation of judicial errors is more difficult to establish. Considering the cases reviewed by Otto (2004), the following number of wrongful convictions according to the definition given earlier in this overview can be listed: at least four for Austria, at least three for France, at least 13 for Germany, at least one for Ireland, at least two for Italy and at least two for Switzerland (Killias, 2008). The comparatively small numbers, however, can not be seen as an indicator for less judicial errors, but might rather be a sign of lacking research into the topic. It

---

[41] Truscott (Re), 2007 ONCA 575 ; DATE: 20070828; http://www.cbc.ca/news/background/truscott/pdf/2007ONCA0575.pdf; last visited 2008/09/03.

[42] England and Wales, Scotland and Northern Ireland have all their particularities as far as the criminal justice system is concerned. Here they will be all summarised as UK for simplification of use, unless mentioned differently. It is supposed (following Quirk (2007) that similar tendencies exist for all of them and that the difference in legal systems can be overlooked for the study of forensic science. Both England and Wales and Scotland have introduced a CCRC and SCCRC, respectively.

[43] http://www.ccrc.gov.uk/cases/case_44.htm; last updated: 31 July 2008.
154 are the cases mentioned by the CCRC (http://www.ccrc.gov.uk/) and 4+6+7+1 comprise the affairs linked to the IRA bombings

[44] As on the 31 July 2008, the Commission had referred 395 out of 10357 cases completed. 421 cases were under review, and 200 waiting.

[45] Maguire 7, Birmingham 6, Guildford 4, and Judith Ward.

is for this reason that the International Network on Researchers on Wrongful Convictions (INRWC, 2003) has been created and comparative research is undertaken.[46]

Some cases will be mentioned later to illustrate specific points of interest and particularities of the various nations. However, regardless of all differences, there are still some very broad general characteristics, which might be identified as risk factors for the occurrence of a wrongful conviction. Edmond (2000; 2002), Jonakait (1991), and MacFarlane (2003) discuss some of these features. They emphasise the particular distasteful nature of the case (murder linked to sexual offences or bombings of innocent people), a large media coverage which might put pressure on the investigators, absence of an evident suspect from the beginning, and the lack of evidentiary material.

Background to these troubles

Although judicial errors tend to share some common features, there are still some differences between countries. The aim of the following discussion is to stress the particularities of each nation, which to some degree also may explain the different measures that were taken to minimise errors[47].

To start with, in the **USA** the criminal justice system *per se* has been accused of many weaknesses. Accusations of injustice are prevalent in every criminal justice system. However, they seem to play an especially important role in the USA. This can be due to either the existence of a considerable amount of critical points or due to a "good" coverage by critical observers of the system[48]. It could also be that *because* of a questionable criminal justice system there is a higher amount of critical publications. Some of them mention the unequal resources available to the prosecution and the defence, both financially and in number of employees, to give only one example (Huff, 2008). Also, on the procedural side, the common use of guilty pleas (Gross, 1998) is considered a serious problem, as their frequency is not equally distributed within the accused population (nor the type of offences committed). It is thus often argued that (innocent) suspects are not protected nor well represented in the USA legal system. The inequities seem to be linked very strong to issues of race (and poverty). This is visible in the descriptive analysis of Gross et al. (2005) on 328 cases of wrongfully convicted in the USA, where nearly 90% of juvenile exonerates convicted for rape was Black or Hispanic, as

[46] In general it is difficult to know how to compare the amount of judicial errors from one country to another, and this for several reasons: The definition of judicial error will determine the possible amount of cases discovered that fulfil the criteria of the definition; Are the cases or the people involved in the cases counted; What population is to be considered in the denominator, the general population of a country, people convicted, the amount of prisoners, or the amount of serious offences happening; And what timeframe is to be considered?
It is interesting to see how this question is solved in comparative studies on crime rates and what solutions could be applied here by analogy.

[47] It could be suggested that judicial errors linked to political cases might trigger a more "official" reaction than a non-political murder case.

[48] It is here that the concept of critical criminology can have its importance as to the form the research agenda on judicial errors could take.

mentioned earlier. False confessions and interracial eyewitness misidentifications seem to be one source of the problem.

A second factor is the **death penalty**, which distinguishes the USA from all other Anglophone and European countries discussed in this review. The risk of potentially executing an innocent person does fuel the discussions of abolitionists as well as adherents of this sentence. Besides all ethical considerations that have been raised on the topic on both sides, this specificity might state one important fact about the USA criminal justice system: irrevocable decisions, such as the death penalty seem to be acceptable in some States[49]. One way to understand this attitude is to say that errors do not occur, and that therefore there is no need for the option of reversing a definitive punishment. However, it has been shown that errors do and have happened. To deny their existence no longer seems a plausible option. A second interpretation would be that such errors are acceptable in the USA (or any other country with or without death penalty) and represent part of the price paid by society for an otherwise satisfactory criminal justice system[50]. Indeed, even if death penalty is excluded from the discussion, it might be interesting to see which, if any, error rate for judicial errors seems acceptable to society in general and the justice system more specifically. Although this might sound rather callous, errors are part of everyday life and even fatal accidents due to one or another type of error have become quite "acceptable", for instance fatal road accidents. So why should this opinion not be adopted by some people in connection with the death penalty or judicial errors?[51] In short, the death penalty has heavily influenced the way judicial errors are debated in the USA. This is reflected for instance in the significant number of publications on the subject (see for example Cooley (2004)), and this point therefore merits special consideration.

In **Canada**, the focus has been more on vulnerability of suspects to pressures during police interrogations, be it persons with limited mental faculties, ethnic and social minorities or indigents, etc. (MacFarlane, 2003). Indeed, several wrongfully convicted in Canada share some or all of the following socio-demographic factors: their age - very young, their social status - underprivileged (e.g. Aboriginal, as was the case for Donald Marshall Jr.[52]), and/or their borderline lifestyle. However, the number of exonerated people is much lower than in the USA[53]; therefore, it is more difficult to characterise the factors that led to judicial errors. A second aspect has been the role forensic science and police investigations have played in helping to convict innocent individuals. This point will be discussed later in more detail.

---

[49] However, the latest decisions of the Supreme Court seem to represent a change of attitude. Indeed, the category of people "eligible" for execution/death penalty is diminishing, the latest decision fixing the minimal age of the perpetrator at 18 (at the moment of the act?) http://www.deathpenaltyinfo.org/article.php?did=885&scid, last visited 2008/08/05.

[50] A certain choice is made whether a wrong accusation or a wrong "letting free" is more acceptable.

[51] This does not mean that the author shares this or another view. But studying a topic implies including all kind of existing opinions.

[52] He was accused of having stabbed an acquaintance (Sandy Seal) in 1971 when both of them tried to rob a third person. He was acquitted in 1983.

[53] It might be mentioned here that no attempt will be made to establish any error rates, exoneration rates or any other rates due to the aforementioned difficulties of numerical and methodological order.

In **Australia**, there is one case in particular that has been acknowledged as a judicial error, which aroused an important amount of discussion. It is difficult to say whether there is a specific factor important for judicial errors in Australia, as there are not many of them. However, the heavily mediatised case of Chamberlain[54] unites almost every risk factor for a judicial error mentioned by Edmond (2002): a serious crime that attires publicity; the accused proclaims her innocence; the suspect is part of a minority (in this case religious); the trial is heavily covered by media; the suspect is found guilty; several movements form themselves with the aim of proving her innocence, etc[55] (Bryson, 1985).

In the **United Kingdom**, on the other hand, judicial errors often[56] seem to be closely related to political issues, namely the violence linked to the fight for the reunification of both Irelands. The so-called IRA (Irish Republican Army) bombings (since 1972) had to be "logically" perpetrated by Irish; it was therefore easy to see in each inhabitant of the "Green Island" a potential terrorist. These bombings, with heavy civil casualties, as well as a sensitive political background, led to tremendous media (and political) pressure, forcing a rapid resolution of the cases in order to re-establish some sort of peace and order, or in other words, the confidence of the citizens in the government and other official structures. Thus, the demands on forensic laboratories and investigators, as well as the general context of the situation helped very probably to augment the risk (and incidence) of wrongfully convicted Irish, as the Maguire Seven case exemplifies[57].

In other parts of **Continental Europe**, the question of judicial errors has not yet reached the (public and academic) awareness it has in the USA or the UK, for instance. The number of cases known is less important, as shown in the precedent section. Also the focus of interest is different. Thus, for example, Otto (2004), in his overview of judicial error, includes cases of wrongful non-convictions – especially of former NSDAP (Nationalsozialistische Deutsche Arbeiterpartei) members or judges (in Germany) – or of convictions where the mental condition of the accused would exclude his responsibility for the acts committed or where the punishment would go well beyond the facts that could be determined (for instance accidental versus premeditated death). In France, for example, the cases of Dreyfus[58] (Champod, 1999) and

---

[54] In this case a mother has been accused of having killed her baby while camping in the Outback. She always claimed that the baby was stolen and killed by a Dingo (wild dog). The fact that she was part of a certain Christian "sect" seems to have had an important role in creating a prejudice against her (and her family). For a short summary see MacFarlane (2003) p. 12ff and 44ff.

[55] A summary of the case can be found on http://en.wikipedia.org/wiki/Azaria_Chamberlain_disappearance, last visited 07/07/2008.

[56] Though other cases of judicial errors exist, the "best known" still are the political ones. This may have shaped to a certain degree the discussion of judicial errors in the UK.

[57] The Family Maguire (including two boys aged 14 and 17) and friends are arrested in 1974. In 1976 they are charged for possession of nitro-glycerine. The last member of the family is released from prison in 1985. http://www.innocent.org.uk/cases/maguire7/moj.pdf, last visited 2008/08/05.

[58] Alfred Dreyfus was arrested and convicted for treason in 1894; anti-Semitism and scientific evidence are keywords that are linked to the case. He was exonerated in 1906. See http://en.wikipedia.org/wiki/Dreyfus_affair for a short overview; last visited 2008/08/05.

Patrick Dils[59] are some of the acknowledged judicial errors (Dongois, 2008). In the Netherlands, the Schiedam murder case has attracted considerable attention and is considered a judicial error. Van Koppen (2008) gives a comprehensive overview of the case. In Switzerland, there are two cases that are rather well-known, a case in Geneva[60] (Killias, 2008) and the case against Bruno Zwahlen[61]. However, cases of alleged, but not proven judicial errors have perhaps occasionally had a more important impact in society than the accepted ones. Thus, for instance the Seznec affair in France (http://www.france-justice.org/), dating from the beginning of the twentieth century has produced more expert reports, both on scientific and legal questions (at least 14 appeals), and legislative changes than acknowledged judicial errors. In the same way the Outreau affair, according to the definition given previously, not a wrongful conviction, has triggered reform movements in the French legal system. Their efficiency is to be seen. Similarly, the Zwahlen case in Switzerland led to a "reorganisation" of forensic work.

One other factor played a role in all these judicial errors, independently of the country – **scientific evidence**. Several features seem to be rather common problems concerning forensic science. The first is outright fraud. That includes the presentation of a report with a conclusion that does not correspond to the actual results, for instance because the work was simply not done (fictitious autopsies by Erdmann) (Giannelli, 2002) or did not offer the wished result (Zain) (Giannelli, 2001). Quite close is the conscious suppression of parts of the evidence (Fish) or the results in order to favour the prosecution (in most cases) (Giannelli, 2007). This is not far away from drawing for ambiguous results the most favourable conclusion, usually for the prosecution (Maguire, 1993). But often the problem is of a different nature and concerns more the misunderstanding or wrong interpretations of correct results. Thus, for example in the case of Michael Shirley (Johnson and Williams, 2004) it can be argued that it was more the way the scientific evidence was understood that led to the judicial error than an error in its analysis (see also 2.3.5). Sometimes it is the search for certainty in scientific evidence that leads to the *fatal* outcome[62]. Indeed, the risk of attaching more value to scientific evidence than it deserves is an important source of error. This happened repeatedly for blood group tests and hair comparison examinations, where non-exclusion was most often understood as an "incriminating" factor (Connors et al., 1996). Misunderstandings or biased evaluations are then (too) often attributed to lacking explanatory skills of the scientific experts called to present the results (Henderson, 2002). Besides these somewhat voluntary errors there is a considerable amount of non-voluntary errors, which might be simply the inadvertent switching of a specimen or reference sample, or transcription errors (see also 2.3.2). Thus, voluntary and non voluntary errors might

[59] As a 16 year old boy Dils was accused of having killed two young boys in Montigny-les-Metz. After having been interrogated for a long time without appropriate legal assistance he admitted having committed the crime (http://fr.wikipedia.org/wiki/Patrick_Dils; last visited 2008/03/14).

[60] Mistaken identification in a robbery; although the convicted pleaded guilty it was only a chance event by the prosecutor coming across another likely suspect that allowed to redress the case.

[61] Bruno Zwahlen was accused of having killed his wife and put her in the freezer. The first time the police visited the house they did not find the body in the freezer. Bruno Zwahlen was liberated in a second instance appeal, the "demand" of the prosecutor to reopen the accusation against him at the last-instance level was refused.

[62] An idea that might raise some questions as the legal requirement for conviction is "beyond reasonable doubt" or "intime conviction", both concepts not including any absolute certainty.

explain why forensic science led to judicial errors. On the other hand, it was also forensic science, especially DNA profile tests that helped to discover judicial errors (Johnson and Williams, 2004; Saks et al., 2001). What led to this change in attitude and awareness will be discussed next.

The change

In all countries mentioned before, it took a considerable time to admit (publicly) that judicial errors could take place and had occurred, and that this happened repeatedly[63]. However, once the first judicial error was acknowledged, it was difficult to stop the avalanche and further cases were admitted. Two phenomena, coming from initially independent sources, allowed for the profound change concerning the attitude towards judicial errors.

The first depended on changes in society, which could be summarised as the renewed attention on the individual's rights within the State[64]. Thus, for instance in the United Kingdom, it was the important media coverage and pressure, which helped draw attention on judicial errors and "forced" the government to take steps to change the situation. However, it may also be stated, that on the other hand this phenomenon of media pressure perhaps initially helped to create (those) judicial errors (Nobles and Schiff, 2004). Thus, as is equally valid for forensic science in general, the same factor can have both a positive and a negative influence on the occurrence of judicial error.

Of more direct influence is the second event, the discovery of the utility of DNA profile tests for forensic science applications (Jobling and Gill, 2004). Already the first uses in the legal context showed that one of its most interesting potentials, often forgotten for other identification techniques, is the power to exclude somebody as suspect in a case, as happened in the first UK forensic DNA case[65] (Aronson, 2005). Ever since, DNA has been often the key element allowing for the reopening of a case and a subsequent exoneration. Its post-conviction use in the USA has led to at least 213 exonerations since 1989, most of them in rape (murder) crimes (http://www.innocenceproject.org/; last visited 2008/02/21). These cold case reviews (McClure, 2007) are mostly possible because evidence has been kept and is deemed to be relevant because of a well-functioning chain of custody protocol.

The confrontation of the new proficient technique (DNA profile analysis) with former existing tests such as blood group or hair comparisons had several (far-reaching) consequences. The most obvious was the gain in (reliability and/or) discriminatory power. The chances of finding two persons with the same characteristics are much smaller for DNA profile tests than they are for the formerly used techniques. Thus, a better method for testing forensic specimen had been

---

63 It can be argued that an appeal system does provide for the possibility of error. However, this does not reflect the way these errors have reached general public awareness and interest.

64 This might again be reflected in parallel academic movements such as critical criminology.

65 Indeed, in this affair called *Blooding* (Pitchfork affair) the first suspect was excluded, as well as over 4000 men whose DNA or blood group had been compared to the semen found on two raped and killed girls. It is interesting to note however, that it was the investigative use of forensic science which was used first (e.g. finding of a potential suspect) and only then that DNA as scientific proof has been accepted (Aronson, 2005). Further investigation led then to a suspect who had avoided being tested, and when confronted with this information, confessed both crimes.

found. The fact of having an old as well as a new test for the same specimen allowed to verify the results obtained with the first. As not exactly the same characteristics are tested, comparison has to focus on the faculty of excluding possible suspects. If the old technique excludes a suspect and DNA profile testing does as well, then both have found the same result and no problem exists. If the old technique can not exclude a suspect, but DNA profile analysis does, then there are two possible explanations. Either the old technique is not discriminatory enough, or some error happened. If no error happened and a suspect was still convicted, the problem might be found in too heavy a reliance on the scientific evidence presented in court, or a combination of other factors. In this case, no error at all happened and the "problem" lies in the decision-making and threshold used in court to decide on the culpability. This is, strictly speaking, not a forensic problem but one of the uses made of it. If error had happened in the laboratory (or any other step of the process from crime scene to court) then DNA profile analysis was a powerful tool to show it and ask for a re-judgement of the case. It is important to stress that, contrary to common belief, a non-matching DNA profile comparison does not automatically exclude a suspect or render him guilty of an offence. For a more detailed discussion on this point see appendix 9.1.1.

However, this differentiation is seldom made, and the interest focuses on retesting specimen by DNA procedures to exclude a common source with the supposedly wrongly convicted individual. Although this allows overturning mistaken sentences, it is not very helpful in illuminating error sources (see also 2.3.5). Edmond (2000; 2002) does raise similar points in his publications, as to this unequal way of weighting and scrutinising inculpating and exonerating evidence.

What the error discoveries did, however, was to force an open discussion as to the quality, reliability, etc. of scientific evidence in court, both on issues of scientific reliability and legal admissibility criteria - this last criterion especially in the USA. Thus, for a first time both these problems were openly debated, even if not entirely satisfactory solutions have been found, according to prevailing literature (Schwartz, 1997; Kaufman, 2001). Still, some improvements can be observed, though they tend to vary from country to country. How this renewed discussion was brought about and was perceived in different nations will be discussed next.

The era of great reviews

As a reaction to the increasing number of discovered judicial errors and the issuing public debates, private as well as official organisms emerged to cope with this problem. In addition, academic research and publications took up the subject. The way these three bodies – private, official, and academic – are interested in the subject will be discussed by comparing the before-mentioned countries. But first the three bodies will be presented briefly.

Private is understood as the initiatives of lawyers, family members, friends or other persons who felt concerned, angry, worried or anything else in a case and decided to take steps to change things. This affects individual cases as well as critical comments on the criminal justice system as such. A famous example is the "Innocence Project"

(http://www.innocenceproject.org/) in the USA. They have 220 exonerations since 1989 on their credit (data on 2008/09/01). Also, they have been an example for others to follow and to offer hope to wrongfully convicted and their families. Their experiences have been published as a book (Scheck et al., 2001).

Official is understood as the reaction by a government body, by issuing reports in one or several cases of alleged or recognised judicial errors, reports on the functioning of a laboratory, a police force, a court, or a part of any of these institutions, as well as the organisms created or decisions taken to influence or study the question of judicial errors or the quality of forensic science, to enumerate some of them. Examples are the Guy Paul Morin Inquiry in Canada (Kaufman, 1998), the Commission on Capital Punishment that led to the Illinois Moratorium (Ryan, 2002), and the establishment of the Criminal Case Review Commission in the United Kingdom (http://www.ccrc.gov.uk/) and Scotland (http://www.sccrc.org.uk/).

The way private and official measures coexist, the types of measures taken, the direction they take, etc., represents the interest of society and the official organisms in judicial errors, and, more specifically, the role of forensic science in it. A short overview will illustrate how the positive or negative role of forensic science as to judicial errors is perceived, both by private and official organisms. The aim is not to give a comprehensive picture, but to highlight differences between nations. These dissimilarities might be quite important for the understanding of working countermeasures aiming at reducing judicial errors.

A third group to represent the awareness of forensic science as a source of and remedy for judicial errors is **academia**. The amount of people interested in the topic, relevant research done and the number of publications, etc., as well as other signs might be an indication of the increased interest in the subject. One important aspect of the academic attention is the scientific consideration of the topic, allowing the study of it in a comprehensive way and thus gaining a correct view of the phenomenon, thereby facilitating the implementation of effective counter-measures. Indeed, until now, most publications based on data are case studies. Although they do highlight real problems and thus offer genuine information on error sources and possible remedies, they nevertheless are not representative of all judicial errors. A more comprehensive, scientific study will consider various types of judicial errors, thereby increasing the understanding of the phenomenon as well as the effectiveness of the discovered/studied countermeasures. This is discussed in more detail in later chapters (5.1-5.3).

In the **USA**, *private* initiatives concerned with judicial errors and the role of forensic science exist. The Innocence Project (http://www.innocenceproject.org/), relying heavily on post-conviction DNA profile analyses is one of the best known examples. Other organisations offer similar services to convicted offenders who claim to have been wrongfully convicted (http://www.centurionministries.org/; http://www.mojo.freehosting.net/index.html; http://www.law.northwestern.edu/depts/clinic/wrongful/links.htm; http://forejustice.org/index.htm; last visited 2008/08/05).

Official recognition of the importance of forensic science in the legal context can be perceived in the amount of state publications, for instance on the use of DNA profile tests, both for convictions and exonerations. Examples are the NIJ (National Institute of Justice) studies (http://www.ojp.usdoj.gov/nij/publications/welcome.htm, last visited 2008/08/05) and Connors et al. “Convicted by Juries, Exonerated by Science” (1996). Other reports, for example on the FBI laboratory (Bromwich, 1997) are significant as well, as they show that it is possible to review the quality of a laboratory. The Commission on Capital Punishment (Ryan, 2002) is another sign of at least partial awareness of problems linked to forensic science quality, but also to judicial errors in general.

On the *academic* level, many publications concerned with the impact of forensic science on judicial errors are published by researchers in the USA (Cole, 1998; Cooley, 2003a; Cooley, 2004; Giannelli, 2001; Giannelli, 2002; Huff, 2002; Huff, 2003; Huff and Killias, 2002; Huff et al., 1996; Risinger et al., 2002; Saks, 1989; Saks et al., 2001; Saks and Koehler, 2005). Their publications also tend to be among the most critical towards forensic science.

In summary, to characterise the USA situation, one could say that it is accepted that an important number of errors happen. Several private, official, and academic movements are working to discuss and minimise the occurrence of judicial errors.

In **Canada**, *private* initiatives against judicial errors are Injusticebusters (http://news.injusticebusters.com/wordpress/; last visited 2008/08/05) or the Association in Defence of the Wrongfully Convicted (http://www.aidwyc.org/index.cfm/ci_id/1073/la_id/1.htm; last visited 2008/08/05).

On a governmental level, *official* reports on different acknowledged judicial errors, for instance the Guy Paul Morin Inquiry (Kaufman, 1998), have been made to study the causes that led to these errors. Another publication, “The Prevention of Miscarriages of Justice” (Bellemare and Finlayson, 2004) proposes a range of measures, some of them concerning forensic science.

Academic research has been done for instance by MacFarlane (2003) and Campbell (2003). Recently a whole issue of the Canadian Journal of Criminology and Criminal Justice (January 2004) has been devoted on the question of judicial errors (see for instance Denov and Campbell (2004)). The focus was on individuals who became victims of judicial errors. A conference was held in 2005 (http://www.wrongfulconviction.ca/; last visited 2008/08/05) on the same subject. A central topic of interest in Canadian studies is the social consequences of judicial errors, both for the incarcerated person, but also for his family. This includes issues such as compensation and public apologies.

In summary, the way judicial errors are analysed in Canada seems to centre more on the individual rather than on more general issues. This does not mean that forensic science is not heavily criticised, but that the focus is to correct the wrong which has been inflicted on an innocent individual.

In the **United Kingdom**, *private* initiatives include MOJUK (http://www.mojuk.org.uk/; last visited 2008/08/05), but also web pages of family members of wrongfully convicted individuals, as for instance Shirley McKie (http://www.shirleymckie.com/index.htm; last visited 2008/08/05) or Susan May (http://www.susanmay.co.uk/; last visited 2008/08/05). Investigative journalisms, as well as the media in general have an important influence on the discovery of judicial errors.

On the *official* level, forensic science has already been subject to government reports, one of the more important was *The Royal Commission on Criminal Justice - the Role of Forensic Science Evidence in Criminal Proceedings* in 1993 (Roberts and Willmore, 1993), where the functioning of forensic laboratories was studied and practical recommendations were made[66]. Furthermore, the independence of laboratories might grow even more, if the recommendations of the *Forensic Science on Trial Seventh Report of Session 2004-05* will be followed (House of Commons, 2005). Concerning judicial errors in general, the creation of the Criminal Case Review Commission (CCRC), has put into reach of every convicted individual to claim his innocence, as well as for their defence lawyers the possibility to refer their cases to an independent and "extra-legal" review. Forensic science, especially novel developments or findings, play an important role in presenting arguments to reopen a case (CCRC, 2005).

More recently, several hundred cases of children that had died presumably of SIDS (Sudden Infant Death Syndrome) have been reviewed after doubts arose on the reliability of the expert opinion (Goldsmith, 2004). Of all the cases reviewed four were considered to be critical.

On the *academic* level, publications on several topics concerning judicial errors have drawn a certain attention (Walker, 2003; Nobles and Schiff, 1995; Nobles and Schiff, 2001), but it is rather the recent publications that propose novel and interesting research issues. The publications by Naughton (2003; 2005), for instance, open new views of research, though they are not specifically concerned with forensic science. More scientific questions have been raised by Dror et al. (2005), which tend to show that the risk of error in identification evidence, as for instance fingermark individualisation, has been underestimated.

In summary, the problem of judicial errors has been recognised and measures have been taken to minimise their occurrence. For instance, on the forensic science level important changes in the management of laboratories and similar institutions have been implemented, which seem to augment the quality of forensic science presented in court[67]. On a governmental level a forensic regulator has been established[68]. His role is to, together with the Forensic Science Advisory Council (FSAC) (http://police.homeoffice.gov.uk/publications/operational-policing/fsac-terms-of-reference?view=Binary; last visited 2008/08/10), to help guarantee quality in forensic science work by persuading those involved to put into place the necessary measures.

---

66 They were then implemented and led to the "fully private sector classified under a separate initiative to establish market choice and a level playing field for the forensic sector" (Bramley, 2008) of the Forensic Science Service (FSS), one of the steps that had been viewed as imperative in the report. The introduction of quality management measures can be traced both to the heavily official scrutiny the laboratories have gone through, but also to the governmental interest to proceed to real and not only cosmetic changes. One reason given for these changes was the general "arrival" of quality management measures (personal communication, Phil Jones).

67 Accreditation is a key word mentioned in this context

68 See http://police.homeoffice.gov.uk/operational-policing/forensic-science-regulator/reports-publications/ (last visited 2008/08/11) for an overview of reports explaining why the regulator has been established, what he does, how, etc.

A further change is the establishment of an experts list by the council of registration of forensic practitioners (http://www.crfp.org.uk/, last visited 2008/08/11), which has two aims. First, to function as a quality control mechanism for the individual listed, who must fulfil certain criteria to be accepted (e.g. follow a code of conduct) and second, to facilitate the access to experts to the defence, an additional critical point to avoid judicial errors. The National Policing Improvement Agency (http://www.npia.police.uk/; last visited 2008/08/11) has to date proposed nine projects to further improve the quality of forensic science (service and use), which are summarised under the general term forensics21 (http://www.npia.police.uk/en/10432.htm; last visited 2008/08/11). Considerable efforts are made at different levels to improve forensic science. Their practical implementation will show which of them will indeed help to meet the expected goals.

In **Continental Europe**, the fact of summarising all the countries in one general label without specifying them already suggests that not very much can be said about the individual countries, judicial errors are not (yet) making the big headlines. Therefore, it is more difficult to draw a general picture. However, discussing the three points – private, official, and academic - might help to gain insight into why this is so.

Private organisations of individuals or associations interested in judicial errors can be found in France (http://www.soutien-jean-michel.net/action_justice/aj.htm; last visited 2008/08/06), Germany (http://www.gehove.de/irrtum/irrtum_2.html; last visited 2008/08/06), and Switzerland (http://www.peter.zihlmann.com/; last visited 2008/08/06) among others. However, some of these web pages are concerned more with judicial errors abroad, mainly the USA, often connected to the issue of death penalty rather than in cases happening within their national boundaries.

Official reactions to judicial errors are scarce. One example is France, where several governmental studies and reports were mandated and specific laws introduced to cope with the subject of the Seznec affair (http://www.france-justice.org/; last visited 2008/08/06). However, to date the case has not been acknowledged officially as a judicial error.

Academic research has been rather limited, especially if forensic science is to be included. Research published in Switzerland by Killias et al. (2007) shows that forensic science is not very often present in last instance appeals. Jakobs and Sprangers (2000) describe the problems encountered in the Netherlands in a case, where the defence wanted to obtain a counter-expertise. Several difficulties are discussed, namely problems of accessibility, financing and control of the report received. Van Koppen (2008) discusses also the Schiedam park murder case. This might have happened in other countries too, as for instance in Switzerland there is no important legislative source/data/info on the question of scientific expertise. The International Network on Researchers on Wrongful Convictions (INRWC, 2003) is one rather novel phenomenon in this domain. Participants are partners of the following countries: Canada, France, Germany, Israel, Italy, Netherlands, Poland, Spain, Switzerland, UK, and USA.

In summary, nearly all categories of awareness - private, official, or academic – have scarcely been developed. Why this is so is not yet known. Speculative opinions mention the type of

criminal justice system, the possible lower incidence of judicial errors, and the different legal culture prevailing. The comparative research project underway might help in answering some of these questions.

If the different countries were to be compared in function of the measures taken against judicial errors, both in a general way or more specifically concerned with forensic science, the following observations could be made.

Private initiatives concerned with judicial errors are the first to make their appearance in a country. Depending on their influence, for instance creating public awareness through the media, the official reaction will be different. Other factors, such as political sensitivity in the UK, might explain part of their success. Unsurprisingly, the type of measures taken by the government will also influence the type of private measures necessary. For instance, in the UK, the creation of the CCRC by the government has given a chance to any convicted individual to have his case reviewed, if his application fulfils certain criteria. In contrast, in the USA no such organism exists. Therefore, the Innocence Project, as well as other institutions concerned with prisoners does offer some sort of private review possibility. For a comparison of both these countries see Quirk (2007). In France, the "Cour de cassation"[69] has in theory the same role; although in reality "eligibility" criteria are so strict, that hardly any case ever reaches the threshold (Dongois, 2008). In (continental) Europe, the presence of the European Human Right Convention (EHRC) might be seen to fulfil the role of the CCRC. Whether Article 6 of this convention - right to a fair process – does perform this role is discussed in Dongois and Schiffer (personal communication).

In contrast, **official** measures have not been taken in all countries in a similar way. Thus, the UK has started with official inquiries on the weaknesses of the criminal justice system. The establishment of the CCRC, but also reports on the functioning of several organisms implied in the reaching of a (criminal law) conviction, police, scientific laboratories, etc., have not only brought an analysis of the situation, but also have proposed practical recommendations that have been introduced. On the scientific level, the independence of the forensic laboratories and the introduction of quality management systems can be cited as examples. Canada has also had some official inquiries and measures have been implemented; however, they tend to be less far-reaching than in the UK. In the USA, official measures have been "limited" to the analysis of specific scandals, such as the report on the FBI laboratory (OIG, 2006) and subsequent propositions of change (Bromwich, 2007).

On the **academic** level, concerning forensic science, most publications illustrate general or specific problems in an affair based on case studies. In other domains, such as eyewitness testimony, experimental research has been done to learn more about potential error sources and about ways to minimise error. Examples are publications by Devenport and Cutler (2004).

[69] http://www.courdecassation.fr/hautes_juridictions_commissions_juridictionnelles_3/; last visited 2008/08/05.

Although some isolated publications do consider specific aspects of errors in forensic science (Cole, 2005; Dror et al., 2006) (and perhaps even possible consequences such as judicial errors), they are nevertheless not frequent enough. And the more abundant publications by lawyers on problems in forensic science are not very helpful, as they are based primarily on case studies and are not representative. Furthermore, they often avoid proposing operational and practical recommendations, as will be discussed more in detail in Chapter 5.

Comments

In this first part the effort was made to give an overview of the variety of existing publications on judicial errors and forensic science. A chronological and comparative presentation helped organise existing information on the topic. By doing so it was possible to see common factors, differences and observe evolutions that took place, for instance by considering geographical (USA, Canada, Australia, UK, Europe) and social aspects (official, private, academic). The view into the past helped detect tendencies that might lead to hypotheses explaining present observations. This includes the introduction of the concept of critical criminology as an explaining factor for the emergence of publications on the topic of judicial errors. By using this comparative perspective it was also possible to directly evaluate these assumptions. Tackling the topic from different perspectives is in line with the multidisciplinary dimension crucial to propose useful recommendations, as already suggested in the introduction.

A considerable amount of interesting data emerged. The focus here will be on aspects touching forensic science. Depending on the country, it is a topic that is discussed more or less. Thus, forensic science is given a very important role in the discussion of judicial errors in the USA, less so in the UK and Canada/Australia and is rather absent in the discussion concerning European countries. One example illustrating this is the creation of the Innocence Project, solely concentrating on cases where DNA can be used to question the initial judgment (Scheck et al., 2001). Other examples are reports dealing entirely with the failing of whole scientific police laboratories, such as the Houston laboratory (Bromwich, 2007). The UK has shown interest in more general aspects, for instance in the Bloody Sunday Inquiry (http://www.bloody-sunday-inquiry.org.uk/; last visited 2008/07/08) and other following similar inquests, where forensic science played a considerable role, though perhaps not as prominent as in the USA. In Canada/Australia the situation mostly resembles the one found in the UK, as far as questions of forensic science go. In most European countries scientific expertise is somewhat absent among the reasons given for judicial errors or their discovery, except for the Dreyfus case (Champod et al., 1999). One of the possible explanations already mentioned earlier could be the differences in the legal systems, but due to a lack of systematic research and a certain bias in the amount of publications in the countries of interest, it is difficult to know what the influential factors really are.

One observation is that publications on judicial errors and forensic science tend to have a strong US, and, to a lesser extent, UK bias. This could mean that some of the problems found in these

countries may not be as important in other, for example European countries, and the other way round. However, wherever the nationwide differences do not affect cross-national observations on error sources, this could be, tentatively, the sign of a fundamental problem in the field of forensic science. Admitting for the sake of the argument an unproven similar chance of error discovery in all countries, the fact of finding more errors in one jurisdiction rather than another could point to specific, rather than to general problems in forensic science[70]. This distinction between error sources in forensic science as such or in a specific context in one country is important when it comes to proposing working recommendations for improvement. It might be much easier to change the first rather than the second.

It is here where the concept of critical criminology becomes attractive[71]. It criticises the justice system and is therefore, by extension, interested in judicial errors. The type of publications found in this overview and the topics discussed in it are explicable against the background of the movement, especially the focus on social (in)justice. This explains also to a certain degree the rather limited interest in forensic science as such.

But also the recent critical comments (Arrigo, 2001; Hil and Robertson, 2003; Robinson, 2001) on the limitations of critical criminology, namely its overly descriptive nature in detriment of more prescriptive positions are of interest here. To put it simply, the main argument is that in only describing problems instead of trying to find, test, and propose solutions, critical criminologists are neglecting one of their (self-proclaimed) responsibilities, the guarding of human rights. In contrast, if feasible changes are discussed on a practical level and can be implemented, it becomes possible to influence the criminal justice system and thus to do something about the initial source of the problem.

In agreement with this last statement the present research focuses on one specific area within judicial errors “to work on”: forensic science. It seems to have a potential for the testing of possible errors and counter measures. The aim will be, in accordance with the previously presented critical texts on critical criminology, to first give a descriptive overview of the problem and then find practical ways of testing possibilities to decrease the (potential) negative impact of forensic science. This last part will be considered in more detail in Chapter 3 (and 4). In the following a detailed overview of possible errors in forensic science will be given, though not all of them might have the same impact on the quality of forensic science or on potential judicial errors.

Some of the areas within this list of possible errors that seem of special interest will then be developed in more detail. They include the lack of the decision making process in forensic science, the lack of empirical studies on the observer bias, and the unbalanced way evidence to inculpate or to exonerate is handled. That all of these three areas rely more heavily on direct human abilities (and therefore include a strong subjective element) should no longer come as a surprise and will be developed in more detail in the chapters concerned by it.

---

[70] Thus, if only laboratories in Switzerland were to report making important errors in one type of DNA profile analysis, for instance, and no other country performing the same tests under similar conditions does, the problem might be somehow linked to a Swiss specificity and not the quality of this type of DNA profile testing.

[71] See introduction to critical criminology in Chapter 2.2.

Considering the three categories interested in judicial errors – private, official, and academic – it might be theorised that it is only when visibility by the three is achieved that things will change, though this does not imply that they have to be all equally aware of the topic. Taking the example of the UK – having made the biggest amendments – considerable private, e.g. media pressure, was built up without having at first direct consequences. It was only when official notice was taken and (academic) research was undertaken that considerable changes were planned and executed.

Thus, even though judicial error has been observed nearly since the beginning of forensic science – as for instance the Dreyfus affair in France – it needs more than media pressure alone in specific cases to change official attitudes. Another basic requirement that allows development is the realisation that judicial errors are not single events – that is, a concatenation of errors in one specific case – but might well be the product of systemic malfunctioning such as in the UK IRA bombing affairs or the discovery of several judicial errors due to the re-testing of DNA in the USA. When the "such a combination of errors only happens once" reasoning is abandoned, it is then possible to view what is necessary to avoid future judicial errors. The UK report by Roberts and Wilmore (1993) is an example. Contrary to the single case studies, it has the advantage of offering remedies adapted to a "generally recognised phenomenon" thus spreading the net wider and being able to cover/catch more cases. Furthermore, as judicial errors are an accumulation of errors, several control mechanisms (in place) had to fail in order to arrive at the final consequence. There is no use replacing one link of the chain for a better one if the chain will break further along. So, instead of blaming one individual or entity the whole structure can be revised and adapted. A specific risk for one-sided blame exists because of the (psychological) need of individuals to criticise something or someone for an unacceptable outcome (Stewart, 2005). The more tragic the consequences of an event, the more the need increases to blame somebody else for it. If reviewed in this light, it is perhaps easier to understand the critical comments by legal scholars on forensic science as a cause of judicial errors – and even more their reluctance to discuss "their own" errors. In addition, it might be perceived easier to introduce measures for improving a forensic laboratory/police force than improving the more general functioning of the criminal justice system[72].

2.3.2 An inventory of various types of errors (error sources) in forensic science leading to judicial errors

For a better understanding of how forensic science can lead to judicial errors it seems necessary to know what kind of errors can happen. The aim of this section is to give an overview of them, by considering the different stages of the process, but without going into too many details as to specific error risks in specific fields.

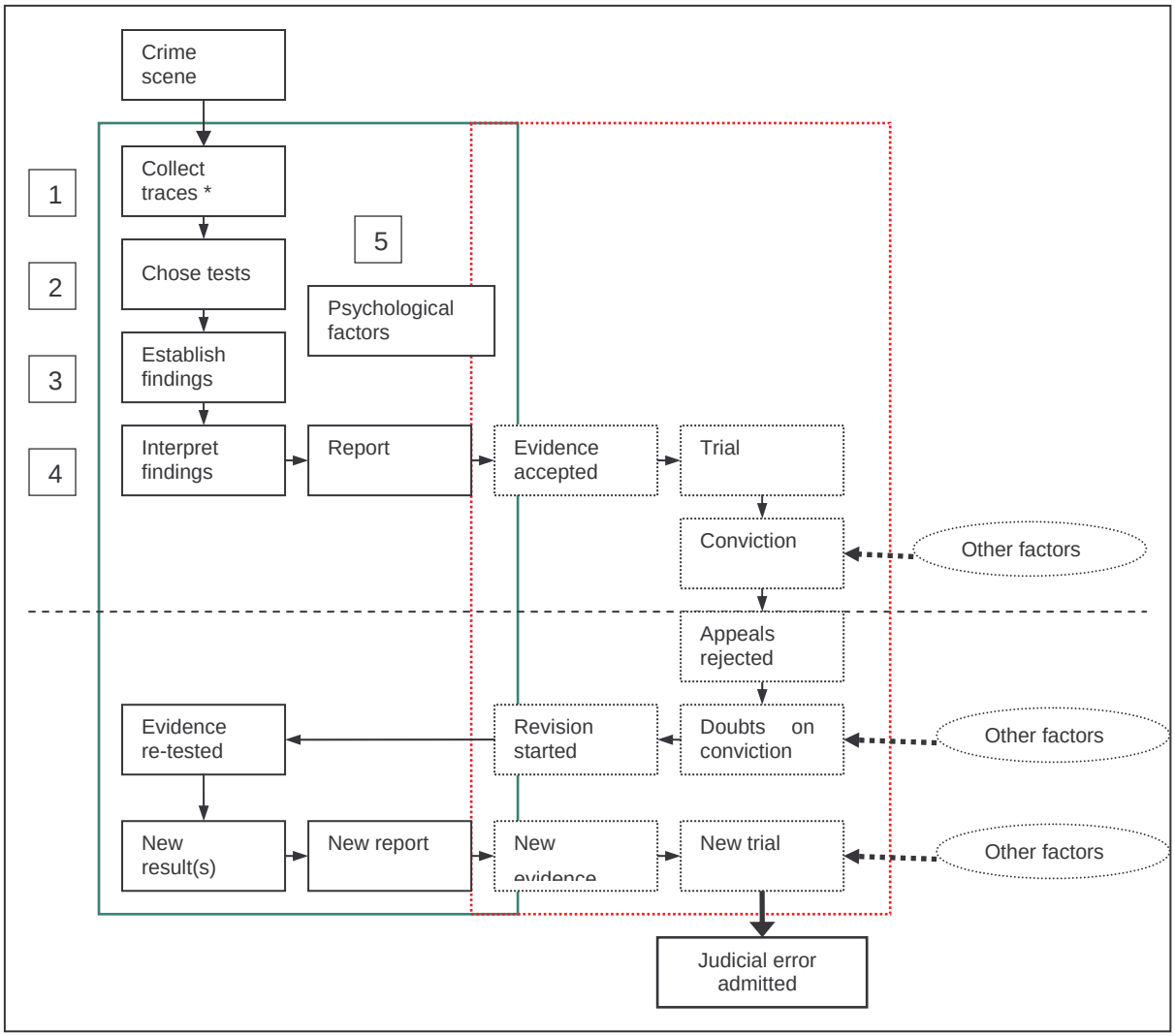
[72] Besides the fact that changing specific elements is easier than trying to modify a whole structure.

As illustrated by Table 1, five different stages will be considered. First, traces and/or specimen which constitute the "raw material" for forensic science work will be collected (2.3.2.1). Second, the tests and analyses to perform, which will depend on various factors such as availability or quality of the specimen, will be selected (2.3.2.2). Third, the establishment of factual findings and interpretation of these findings in a context, which includes for instance the understanding of alleles extracted from a DNA profile analysis, but also their consideration in a wider context so as to be of use according to the needs of forensic science will be made (2.3.2.3). Fourth, the presentation of results to non-scientists will be considered (2.3.2.4). Fifth, the fact that psychological biases can have an influence at all before mentioned stages and tend to be stronger in those that rely more directly on human judgment (2.3.2.5) will be established.
It is only if an error is not noticed at the moment it occurs that it will progress further in the system. If it is not discovered at any of the next stages, then it can lead to a judicial error, together with other factors which may include the absence of other items of evidence, a false confession, etc. In short, for a judicial error to happen, several protective mechanisms in place have to fail. If at any stage the error(s) is discovered, the outcome could be quite different.

Furthermore, it is proposed here that the nature of the forensic science work does not change whether it is presented before or after the first trial and that therefore this distinction is not necessary to understand potential error sources. Therefore, abstraction of the different *use* of forensic science can be made that is, whether it is an investigative[73] lead (Jackson et al., 2006), an evidentiary proof in court, or an exoneration tool once a judicial error is suspected. The focus will be here on forensic science *before* the first conviction has been made. However, there seems to be a tendency to understand forensic science in a different way, depending on its use made a topic discussed under 2.3.5.

[73] see p. 11 for a definition.

Figure 4 The (possible) role of forensic science in the "creation" of a judicial error



(* Detection and collection of traces)

At each stage it is possible that error does not occur and that the case continues normally towards the next stage. On the other hand, errors can happen at each stage. They can also be discovered at any moment. Even if mistakes happen, they do not necessarily have a negative influence on the "life" of a case. Only the "worst-case-scenario" is represented in this illustration.

As already mentioned before, the focus will be on non-voluntary errors. Fraud and wilful manipulation can happen at each stage of the forensic science process and might be a question of general policies in place, such as the selection/training of future forensic experts or the existing control mechanisms and not due to specificities of forensic science.
Furthermore,

> "The more sophisticated the error-avoidance procedures, the more difficult deliberate falsification will be rendered, and the more the temptation to do so will be reduced." (Risinger et al., 2002, p.11).

The interest is therefore directly on forensic science work and not on external factors that are potentially influencing such as a lack of funding leading to important backlogs in DNA profile analyses (Stevens, 2008)[74] or other outside constraints (Edmond, 2002)[75]. As mentioned before, these factors are not specific to forensic science, and will therefore not be developed here.

2.3.2.1 Errors in the collection of traces/specimen or errors (e.g. confusions) in the manipulation of such traces

A forensic investigation starts at the crime scene (Martin, 2002). It is based on the premise known as the transfer principle of Locard (1920[76]). Generally, the offence has been discovered by somebody, reported to the police and verified by police officers, before the arrival at the crime scene of the investigators. Hence, the crime scene is probably most times already altered and/or contaminated before the search for and collection of traces starts. Thus, the initial situation is not perfect. Various factors influence the way the investigation is initiated and carried out. The nature of the crime scene, the type of event, external circumstance, etc., will influence the work which has to be done. Experience and information on the supposed crime, for instance, will guide the work of the crime scene investigators. They have to select and collect relevant traces so as to provide a maximum of valuable information in an adequate timeframe. **Relevance** is a concept not very clearly defined in forensic science. In this context the relevance of a trace will be understood as depending on the trace as such (quality, nature, etc.), the context in which it is found (type of case, place found), and the use that can be made of it (investigative or evaluative)[77].

One of the first things done on a crime scene is to proceed with a **selection/choice** of things to do and to collect. Every selection does not include all potentially latent traces, specimen, etc.; in consequence, not all information will be available (logically, it never can be). On the other hand, it is necessary to concentrate on useful items. In case of a fire in an important industrial complex for instance, it could be helpful to have some idea as to what might have caused the event and/or where its (potential) origin was. Thus, the search can concentrate on electrical supplies or a defective machine instead of working the way through tons and tons of charred and burned material. Without having some preliminary hypothesis on what to look for, it is difficult if not impossible to find something. If one is searching for a specific item, the chances are higher of

[74] This is especially important, as there have been instances where samples have not been analysed and other innocent people arrested in a case or the perpetrator is free to commit further crimes.

[75] Other factors could be : case loading, resource limitations, timeliness pressures, performance targets, abstraction for other tasks (e.g. Management) (Bramley, 2008)

[76] « La vérité est que nul ne peut agir avec l'intensité que suppose l'action criminelle sans laisser des marques multiples de son passage. [...]. Les indices dont je veux montrer ici l'emploi sont de deux ordres : Tantôt le malfaiteur a laissé sur les lieux les marques de son passage, tantôt, par une action inverse, il a emporté sur son corps ou sur ses vêtements les indices de son séjour ou de son geste. » (Locard, 1920, p. 139)
No one can act (commit a crime) with the force (intensity) that the criminal act requires without leaving behind numerous signs (marks) of it; either the wrong-doer (felon, malefactor, offender) has left signs at the scene of the crime, or on the other hand, has taken away with him — on his person (body) or clothes — indications of where he has been or what he has done (approximate translation found in Inman and Rudin (2000; chapter 4). See for an extensive discussion on the topic Crispino (2006).

[77] The author would like to thank Đurđica Hazard for discussions on this subject.

finding it. Thus, if one searches for footwear marks for instance, one might find them. But, depending on the situation/case and their location their importance could be more or less great. For instance, footwear marks on a dirty floor in a burglarised supermarket might be much less relevant as the same marks on the freshly cleaned floor of a burglarised flat. Therefore, the circumstances and the initial hypotheses on the offence will influence the choices made on the crime scene. This selective way of working may be problematic as several types of factors risk interfering with an objective approach. Difficulties linked to biases will be discussed in more detail in Chapter 2.3.2.5 and 3.2. In short, selecting what is relevant on a crime scene already involves choices that can be influenced by different factors and therefore can be prone to errors. What is not recovered in the first place may not be recovered later.

Traces gathered have to be protected from **contamination and/or destruction**, as most of them are fragile. A fingermark on a smooth surface, for instance a wine glass, is easily transformed into an unrecognisable smudge, if handled carelessly. The same applies to footwearmarks, ear marks, fibres, etc. That is why special care is applied and every item is photographed and registered before lifting. Specific adapted techniques and materials are used to gain an utmost amount of information and to retrieve traces as carefully as possible. Again, if an item is contaminated[78] with DNA for instance, or destroyed, valuable information is lost. The fact that information was lost may also cast doubts on the other parts of the work performed. Thus carelessness in one area may create doubts on well done areas as well.

The next step is a correct **labelling** in order to find, even years after the fact, the necessary information as to who collected which specimen; how, when, and where on the crime scene it was found; and what was done to it or with it. The item is then transported to a laboratory or a storing facility. To keep track of all these movements, a **chain of custody protocol** is established. This allows for proof that the specimen presented in court is the same as the one analysed by the laboratory, which is, in turn the same that was found on the crime scene at a specific place. This protocol should also help to protect fragile traces, for example biological specimen, and ensure that they have been handled and stored in a manner stipulated by guidelines. The more data is available, the more information can be extracted, for instance on the relevance of the item treated. Gaps in a chain of custody, even small, can easily undermine the credibility of the work performed and render a useful piece of information useless.

Obviously, at each of these steps, as well as in most subsequent laboratory work, it is possible to inadvertently (or willingly) switch (crime scene or reference) specimens, label them incorrectly, and/or lose them as they are transferred from one place to another. The only way to minimise these risks is to work in an organised way, to avoid doing several things at one time, and to have parallel mechanisms to verify the steps that have been taken. That is one of the important roles of photography (Martin, 2002) which demonstrates, in addition to the written protocol, where each specimen was found and lifted.

[78] Initial contamination before discovery and collection can not be controlled. But any further contamination has to be avoided.

Concerning the abovementioned points, two main areas that seem to be creating some discussions lately will be shortly presented. Both of them are ultimately linked to the question of the quality of crime scene work and thus of potential errors and error avoidance procedures. The first is **relevance**, as it has been defined here. What kind of use can be made of forensic science for instance is one area of discussion, the question being whether it can be integrated directly into the police investigation as an additional investigative lead or if it keeps the more traditional role of evidence in court. Another tendency observed is to replace some types of forensic science for others, in most times traces such as fingermarks, ear marks, tool marks, etc., are replaced by biological traces offering the often perceived as individualising DNA profile (Roux et al., 2007). This concerns very directly the relevance that is attributed to a trace.

The second topic is **accreditation** of crime scene work. Whereas many areas of forensic science work in the laboratories have undergone accreditation or other forms of quality management procedures, this has been less so the case for crime scene works until now. Practical difficulties, such as for instance which norm would be applicable tended to be topics of discussion, but seem solved now[79]. Though accreditation is not synonymous with quality, a non-accredited entity being capable of producing very high quality, it can be understood nevertheless as a sign of interest on the quality of the work performed. Both accreditation and usefulness of a trace do therefore shape the quality[80] and also the type of work performed on a crime scene, and thus indirectly also influence the kind and quantity of possible errors.

2.3.2.2 Errors in the choice and performance of technical tests

Once a trace or reference sample has been collected and/or registered, it will be processed for further extraction of information. This happens through various kinds of tests and examinations. This includes first categorising an item into some broad classes to have an idea of its general nature. Then, further tests will render these categories smaller until only items coming initially from the same source are in the same class. This is called identification (Kirk, 1963). It is at this moment that items coming from the crime scene (in the broadest sense) will be compared to sample(s) coming from a reference or comparison source, which can be linked to the case or a reference database.

For (nearly) all types of forensic science specimen, several methods for testing, comparing or evaluating exist. It is common to use more than one technique to analyse a specimen, mostly because of their complementarity, though other factors might influence the choice as well. The simplest criteria refer to the nature of the specimen, for instance its state, size, quality, etc. A

---

[79] **UKAS**: http://www.ukas.com/Library/downloads/Information_Centre/Brochures/Forensic%20Accreditation.pdf; last visited 2008/08/11; **ENFSI**: http://www.enfsi.eu/page.php?uid=45; Standards for Accreditation, last visited 2008/08/11; and **SECO**: Guide for the assessment in the field of forensic evidence recovery ISO/IEC 17020:1998; http://www.seco.admin.ch/sas/rm_kt/00145/index.html?download=NHzLpZig7t,lnp6I0NTU042l2Z6ln1ad1IZn4Z2qZpnO2Yuq2Z6gpJCEdoF2fmym162dpYbUzd,Gpd6emK2Oz9aGodetmqaN19XI2IdvoaCVZ,s-&lang=en; last visited 2008/08/11.

[80] But do not guarantee the quality of the result, in the same way as procedural rules should guarantee a fair process but not a “correct” or “just” outcome.

second aspect is the availability of the technique or method, often a question of financial affluence and education/training of the examiners using them. Furthermore, if comparisons are envisaged, reference samples have to be available as well and to fulfil the same (or if possible higher) quality criteria. These are points to worry about before any analysis starts. Afterwards, several additional factors have to be considered, which will be presented next.

Once the sequence of examinations has been decided – based on preliminary (optical) observations of the traces and the reference samples - the items to be tested are selected. If, for instance, several paint chips are recovered, one choice could be to submit only part of them to some type of tests, considering their size, quality etc. Thus, if not the entire specimen is tested it is always possible to do the same tests again. The faculty to replicate a test is one criterion of scientific work and will allow verification, especially if doubts arise. Even though there is an ongoing discussion as to whether forensic science qualifies as science or not (Crispino, 2006), it does not seem wrong to try to keep as close as possible - whenever feasible - to scientific standards[81]. It is also possible, that due to insufficient quality (e.g. partial fingermarks) an examination will not be able to produce any result at all.
Each of these steps demands a subjective evaluation of the possibilities and is therefore vulnerable to mistakes. One of the specialties of forensic science is the varying nature of specimens that arrive at the laboratory. Even if the type of traces itself is always the same - for instance paint chips from a car - their condition, quality, etc. is quite variable and thus makes it difficult to have a strict protocol to ensure quality and consistent testing.

Both the specimen and the techniques available influence each other. The specimen determine, among other things, which techniques will be employed. On the other hand, the analyses performed modify the specimen and their characteristics (sometimes permanently) so that additional tests might not be possible any more, especially if the specimen have been destroyed in the process. The risk seems greater for advanced/sophisticated analytical tests. This is one reason for an essential forensic premise, to proceed always from non-destructive to destructive techniques. This means to preserve the specimens' integrity inviolate as long as possible and start with methods that do not change its physical and chemical properties.
The destructiveness of certain methods is also in direct violation of the principle of being able to retest traces in order to check on earlier results. This could be necessary in three types of situations. First, the results of an examination do not correspond to the expected and/or usual findings. Without "back-up" material, it will be impossible to determine whether this is due to this specific specimen or whether some sort of error, of human or technical origin, is involved. Second, if there is a doubt about the quality of the work performed, for instance if the case is

[81] See Goodstein (2000) and Crispino (2006) for an overview of different ways to understand Science (in the context of forensic science). It can be inferred therefore that the term Science is not as self-explaining as might be assumed. This seems to be, at least partly, an explanation for (potential) misconceptions emerging between non-scientists and scientists, but also within the scientific community (as such). Examples for this are the discussions that took place in the context of the Daubert/Kumho hearings (Haack, 2003; Kaufman, 2001; Schwartz, 1997), where the definition of science plays an important role.

reviewed, it will not be possible to confirm or refute this hypothesis. Third, if in the future a new technique should replace the one used, additional tests would not be feasible either.

To avoid destructive examinations in the first place would mean to exclude several techniques and a certain amount of precious information. However, there is one way to mitigate the negative effects of the "one-test-only" situation without having to exclude destructive tests. The simplest way is to split the initial specimen and keep one half for future reference and testing. This might not always be possible, especially if the specimen is very small, a typical feature in forensic science work. Another way is to keep complete and precise laboratory notes. Although these do not replace retesting, they, if done correctly, allow for a complete follow-up. It might be therefore possible to reconstruct the work done, and as a historian does from historical material, construct a hypothesis as to what occurred. Through this, the cause of the problem could probably be established (long) after the initial testing and might be sufficient for a third person to decide on the reliability of the results. The worse initial examinations have been, the easier it will be to show that the tests did not fulfil the requirements in place at the time of the analysis[82].

Even under these ideal(ised) conditions two further problems can not be discarded (at least not under normal circumstances). These concern the mixing-up of specimen in the laboratory or the contamination – with items of the same or another case – as well as the wrong execution/performing of tests, due to technical or human error (or other unavoidable causes).

Cook et al. (1998) present a model for formalising decision-making in the testing stage. This development was triggered by a new customer-oriented management style of their organisation and raises the following points. First, some advanced techniques might be time-consuming and thus lead to problems in the general investigation of the inquiry. Second, the techniques and their sequences must also be selected in function of their capability to answer the questions raised by the customer of the report. Indeed, sometimes tests will not help to elucidate the points the police and ultimately the Criminal Justice System wants to have answered. This could be especially worrying, as the understanding in court could be compromised if many tests with no clear purpose are presented to non-specialised persons[83].

If summarising two aspects of this section, error sources and error avoidance measures, it can be observed that the "human factor" is again very important. At each stage decisions have to be made as to what tests to perform. Though for some areas this is rather straightforward to determine, as very specific protocols are in place - for instance for DNA profile analysis - this might be less the case for other areas. This lies in the nature of forensic science work and can not always be controlled. Then, the technical possibilities available and the knowledge that is necessary to implement them are fundamental as well. As to error avoidance measures the need for correct documentation is always a basic requirement. Sometimes it might be possible to keep part of a specimen for reanalysis, and if this is possible, it should be done. It is also clear that at this stage nearly all errors mentioned before can still happen, such as mislabelling,

[82] During the review process for the Birmingham Six bombing a special emphasis was given to establish whether the methodology applied at this time was following standard practice and state of the art recommendations existing at the time of the analyses. Indeed, judging from today's perspective the quality of earlier analyses could lead to hindsight, one category of observer bias (as discussed in more detail in Chapter 3).

[83] This is one of the admissibility criteria in the USA (Thompson, 1993).

confusing, contaminating, choosing the wrong item to be analysed, etc., and that for their avoidance the reliance has to be placed on the human being performing the work.

Comment

Besides these factors mentioned, other more general errors can prevail as well. For instance, the results obtained in the laboratory have to be transferred into a report, thus simple clerical errors might be responsible for wrong results. The transfer and/or the interpretation can equally be lacunal or erroneous, as it is not necessarily the same person who did the analyses that also writes the report. Thus, for example the laboratory assistant doing the DNA profile testing might have some doubts about one peak representing a specific allele. However, when transmitting the information to the person in charge of writing the report, this detail might get lost. Also, even though rigorous laboratory notes are common standard in most scientific laboratories, they cannot contain 100% of all information that might be (considered) necessary afterwards.

2.3.2.3 Errors in the establishment of factual findings and interpretation of these findings in a context

Factual finding focuses on understanding the results of a test or an analysis, while interpretation of these findings is defined here as concerned with the interest of those results in forensic science. In other words, analytical results (factual findings) from a chemical test will not vary much whether they are made with forensic specimen or other samples (besides the differences in sample size already mentioned and the need for comparing traces and and reference samples), while the interpretation might consider to a lesser or bigger extent the specific forensic science context. First, the focus will be on establishing factual findings and then on interpretational issues.

The results of a test can take different forms, such as a chromatogram for instance. They need to be interpreted to extract useful information. This involves identifying peaks of interest by comparing them to reference samples, for example. Or it includes evaluating the presence, type, and quality of minutiae observed on a fingermark and a comparison print.

One first aspect to consider when discussing the establishment of factual findings of the evidence is the **nature of the test, indicative or confirmative**. A specific “drug” spot test might tell something about the presumptive presence of “heroin derivates” in a street sample seized by the police. While this might be enough to confiscate it and send it to a laboratory for further testing, it is not a definitive proof of the presence of heroine in the item, as other substances could also produce the same reaction. That is why it is called an indicative or presumptive test. In contrast, once tests have been performed in the laboratory, by GC-MS (gas chromatography coupled to a mass spectrometer) for instance, the presence of the substance “heroin” and its derivates can be confirmed. Both methods are accepted and have their reasons to be. Difficulties arise if the results of an indicative test are presented as if coming from a confirmative test. Several explanations can be found for this to occur, ranging from ignorance or misunderstandings to deliberate falsification or great investigative pressure (for instance, the

indicative result from a spot test is understood by investigators as a fixed value). While there are different reasons, the outcome of this confusion is often the same. In most cases, it ends with attaching too much weight to the results (see the case of Barry George (2007), or not explaining the risks of false positives, as for instance the case of the Maguire Seven in the United Kingdom (Maguire, 1993)). To erroneously link a reference sample of a suspect to a specimen coming from the crime (scene) is the most obvious error leading to judicial errors (or is at least perceived that way).

While in the example given above the technique used was not able to differentiate one category of interest, heroine derivates, from other substances producing the same reaction, some techniques cannot differentiate items within a same category but not coming from the same source. In other words, the discriminatory power of a test will allow the distinction of glass from bottles and windows, but not between all different windows. This means that when comparing glass fragments found on a suspect from a burglary with a sample of glass from the burglarised window no analytical difference is found, but comparing the same incriminating specimen to other, non-burglarised windows could also lead to an observation of non-difference. There is no failure in not being able to differentiate items coming from different sources; it is a technical limitation that has to be known to correctly assess the value of the result. It can lead to a **false positive**[84], that is giving the same reaction as the one desired, but by a non targeted substance.
Another example of a technically correct result but not a correct inference nonetheless is when specimen are (inadvertently) switched. For instance, if the comparison samples of several suspects (A-E) are mixed up, a correspondence between the crime scene specimen and the comparison sample from suspect B could be found, though in reality the sample labelled B corresponds to suspect E.
A further situation is when the test or technique should be able to highlight existing similarities from crime scene specimen and reference samples coming from the same source, but is not able to do so. This is a **false negative**. In a scientific context both the false positive and the false negative have a similar importance. It will be the nature of the test that determines whether the false positives or false negatives are acceptable and to what degree. In the legal context, however, and by extension, in forensic science, the importance of the false negative tends to be mitigated by the following reasoning. The famous “it is better to let free ten guilty men than to falsely condemn one” attributed to Blackstone is often quoted. This view somehow represents the legal principle of *in dubio pro reo*, stating that if there is a “legitimate doubt” about the culpability of the accused, he will not be convicted, but should “profit form the doubt” (le doute profite à l’accusé).
Taking a “scientific” point of view, a false positive (to convict an innocent) and a false negative (let guilty free) are equivalent, as both describe a situation where a difference to the true outcome is observed. It is an error. However, the consequences of these errors can be different and it is here that the aforementioned quote on ten guilty going free (or *in dubio pro reo*)

[84] See for a slightly different evaluation of its importance Inman and Rudin (2000) Chapter 10.

represents the choice of society to protect the presumption of innocence. Although it might be argued that in the context of judicial errors this error seems not to be of importance, as no innocent will be convicted, but rather a guilty one will/might go free, there seems to be one serious drawback to this argument. If the real offender is excluded by (forensic) science – false negative - then the focus of the investigation will be on somebody else. And as this person is factually innocent, errors do have to occur in most cases to be able to find scientific or other support for his "guilt". Thus, false negatives can be, indirectly, as much a source of judicial error as false positives[85]. One of the reasons for this confusion is that a "not guilty" verdict can state doubts on the safety of the conviction[86] and not manifest a belief in the factual innocence of the suspect. Risks seem to be highest in cases with very limited evidence, scientific or other, where the interpretation of factual findings may be decisive; hence the risk of overestimating the meaning of the analytical results.

As to the second point - interpretational issues - statistics have been introduced early into forensic science, as for instance in the process against Dreyfus (Champod et al., 1999) but have gained considerable momentum with the introduction of DNA profile analysis. One of the reasons was that it became necessary to have an estimate as to how frequent the observed characteristics, in this case alleles, would be in a given population in order to assess the probability to find per chance the same combination of characteristics in the same population. In this context the concept of likelihood ratio has established itself in forensic DNA profile analysis (Taroni and Aitken, 2000) although conceptually exposed in other forensic fields. It then also spread into other areas, first those having a stronger analytical background, such as the analysis of glass, and gradually also those having been historically less associated with it, such as fingermarks (Neumann et al., 2007).

One of the basic difficulties is associated with obtaining useful data for the statistical and probabilistic calculations. These do not diverge from any other statistical studies carried out in medical or social sciences fields, where for instance having a representative population and a given distribution of the characteristics observed are factors to be considered. Contrary, however, to these fields where the aim is to reach a generalised view of a phenomenon, for example the chance (or probability) of getting ill with measles given a certain frequency of the phenomenon observed in a given population, forensic science is interested in knowing what the probability is to observe a specific combination of characteristics in one specific case.

Several recurring errors in understanding statistical significance have been observed, for instance the **prosecutor and defence fallacy** (Thompson and Schumann, 1987). The main explanations found for these errors tend to show that probabilistic concepts are rather abstract and often counter-intuitive to human reasoning. Very often people confronted with them, for instance jurists or even biologists do not always have a strong statistical or mathematical background. However, this so-called likelihood ratio based reasoning also applies in a more

[85] And there is the risk of recidivism creating indirect victims in consequence of the decision of letting him go.

[86] Considering the possible "costs" for society of letting a guilty person free one might wonder whether in cases where doubts arise on the safety of the conviction and less so on the factual innocence question it would be justifiable to keep an eye of suspects acquitted. This would depend on the accepted compromise between the individuals' rights and the need or wish of society to be protected (point raised in a discussion with Bernard Wirz).

general way, thus helping to present forensic science results in a way that has been stipulated to be more balanced. In sum, likelihood ratio based reasoning, which will be developed a bit more later, clashes with a more traditional way of presenting forensic science results often labelled "**expert opinion**"[87] (Cole, 2008b).

This other way of presenting results tends to argue that the evaluation of the work done by the forensic scientist happens in the laboratory, is based on training and experience, may include statistics, but will not try to formulate this in a explicit way. Finally, it is out of the observations made, the experience, etc. that a personal conviction as to the result arises which is then presented in the report or orally in court[88]. In contrast, in a likelihood ratio framework, at least two competing and mutually exclusive hypotheses, generally attributed to the position of the prosecution and the defence, are formulated and the analytical results are assessed under both of them. Depending on the type of data available, including the amount of context information, this is considered in a hierarchy of propositions which is divided into source, activity, and offence level (Cook et al., 1998), increasing at each level the amount of context information considered.

Heated discussions exist as to the respective merits and disadvantages of both approaches (Cole, 2008b), which will not be developed here. Their significance here lies in the types of errors that have been attributed to them. "Expert opinion" has been accused of being subjective and lacking transparency, thus often leading to an overestimation of the value attributed to the results. Likelihood ratio-based reasoning on the other hand, has been accused of unnecessary complexity by the difficulty for non-scientists (and of forensic scientists as well) to understand the statistical and probabilistic models advanced. It has also been accused of trying to "impress" with numbers and thus helping to overestimate the value of forensic science evidence. In fact, research on the perception of statistical evidence has shown otherwise (Kaye and Koehler, 1991). In short, both sides tend to accuse the other of overconfidence in their results, one side because of their subjective view, the other because of their objective view.

This summary of the ongoing discussion is very succinct and does not consider all aspects. One inference can be drawn very easily however: that the evaluation of the value to be attributed to forensic science results is not easy, nor is the presentation of those results, a topic discussed in the next chapter.

2.3.2.4 Errors in the presentation of evidence to (non) scientists

There seem to be various problems connected to the presentation of scientific evidence to non scientists and lay people less familiar with this kind of approach. A clash of cultures between scientists and lawyers (or in some countries also the jury) can be observed. This has been

[87] Depending on the jurisdiction, the term "expert" can be very narrowly defined; that is why it has been put into brackets, though here it will be used mainly as a concept opposed to likelihood ratio-based reasoning.

[88] In some ways this concept seems to be similar to the "intime conviction" of a judge in a inquisitorial legal system, where after the consideration of all evidence submitted, he will reach a verdict, intimately convicted; here the "evidence" are the examinations performed directly by the forensic scientist himself and concern the result observed, whereas the judge will evaluate what is presented to him and take a decision as to convict or not.

discussed repeatedly by different authors, such as Lucas (1989), Lynch and McNally (2003), or Rudram (1996), among others. It exists not only for the intersecting area between forensic science and law but also for other types of specialised evidence, for instance medical opinion.
Reasons given for the clash vary depending on the point of view. Thus, lawyers tend to argue that scientists talk (and write) in a much too scientific and complicated way, without considering their target audience. Scientists, on the other hand, tend to argue that lawyers (and juries) lack the basic (scientific) knowledge required to understand whatever they present. Both sides are right and wrong at the same time. There is a **language problem** because the same words might not necessarily have the same meaning in the others' field. This seems rather obvious for legal matters, where even a simple word will have a specific meaning in a specific context[89]. However, it is not always simple to understand the difference between the day to day use and the specific use of a term in the legal context. Exactly the same problem is true for scientific terms.
A simple way of understanding possible different meanings is looking up a common word in a dictionary, where its use in different areas is explained. If defining "**error**" for instance, the online edition of the Oxford English Dictionary[90] gives several meanings for the word, the most interesting in our context being the following:

> "[…] 3. a. The condition of erring in opinion; the holding of mistaken notions or beliefs; an instance of this, a mistaken notion or belief; false beliefs collectively. […]
> 4. a. Something incorrectly done through ignorance or inadvertence; a mistake, e.g. in calculation, judgement, speech, writing, action, etc. […]
> b. A mistake in the making of a thing; a miscarriage, mishap; a flaw, malformation. […]
> c. Law. A mistake in matter of law appearing on the proceedings of a court of record. […]
> d. Math. The quantity by which a result obtained by observation or by approximate calculation differs from accurate determination […]"

Exactly the same word can have different meanings, depending on the field and context in which it is used. Furthermore, each of these meanings includes a different conceptual background that might not always be known to everybody using them, one of the reasons that can cause misunderstandings.

This can further be illustrated by the following example of the meaning of a measurement result. If the length of a table is measured, some value will be found, for instance 120 cm. This will be understood quite differently by a scientist and a non-scientist. The latter will perceive the 120 cm as a fixed value, whereas the first is trained to see the 120 cm as the best approximation of the

---

[89] See the example of the different (legal) meanings the term "issue" takes in Dorothy L. Sayers' detective novel "Unnatural Death" (Sayers, 1995b) which defines the whole plot of the story.
[90] http://dictionary.oed.com/cgi/entry/50077695?query_type=word&queryword=error&first=1&max_to_show=10&sort_type=alpha&result_place=1&search_id=n6Dv-Evcvw0-2015&hilite=50077695; last visited, 2008/08/10)

real value, as several measures have to be taken to get near the real value. An acceptable error rate will be defined as well, for a better description of the measured length of the table. Thus, although the scientist says 120 cm, he actually means: considering the *n* measures made, and the (estimated or calculated) error rate, the real value is best described as being 120 cm ± 2cm and thus 120 is a simplified way to express this. A full and adequate Bayesian estimation has been left out to simplify the issue. Nevertheless, for the day-to-day needs it is quite correct to see 120 cm as such. For instance, if the table is to fit into a room then there is no need to worry about approximation and error rate. This small example is a good illustration that the same expression might be understood quite differently, depending on the background of the person and the needs of the circumstances.

Awareness on both sides of such communication difficulties might add much to improve the situation. What can be done to minimise the gap and increase mutual understanding has already been subject of various discussion - among them publications by Bird (2001), Frankel (1989), Haack (2003), Hauck (1996), Jackson et al. (2006), and Lynch and McNally (2003) - and will not be treated here.

The problem of unrealistic **expectations** by non-forensic scientists will be introduced next. It is advanced here that the arrival of DNA, as an example, encouraged the view among non-scientists that new technologies might help to resolve the difficult questions of guilt or innocence and thus minimise the cases where considerable doubts on the subject persist. For instance, DNA is much more accurate than earlier technologies, the risks of error seem therefore smaller and the possibility to link an individual to a crime scene trace to the exclusion of the rest of the human population on Earth seems more likely. These developments are reinforced by the important success of television series and books featuring forensic scientists and forensic science laboratories (Starrs, 2005), where science works miracles to convict the suspect (seldom to exonerate an innocent) and where nearly everything is possible. The expectations of lawyers, jurors, and the general public are influenced, and do surpass the real scientific possibilities. It already seems more difficult to convict suspects, as the decision makers expect forensic evidence to be present, and do not convict if absent (Starrs, 2005). However, tests by Schweitzer and Saks (2007) seem not to reinforce this hypothesis (see also Cole and Dioso-Villa, 2007). This trend creates problems for forensic scientists, as real life conditions are never as perfect as in the pre-orchestrated television series. Besides the fact that real life is never adequately represented in a book or on screen, there is another, much more important underlying problem. The wish to rely on forensic science can be perceived as the need to have something reliable to evaluate the case and get rid of uncertainties.

Although it is reassuring to observe that the quality of evidence presented in court is an issue, there is another side to that. It can equally be perceived as a wish to transfer the burden of responsibility for a judicial decision from the decision makers (judge, jury, etc.) to some sort of "solid and reliable" external standard: (forensic) science[91]. Yet, as explained earlier, (forensic)

[91] Depending on the judicial system this might have a bigger or smaller impact, but as explained in the introduction, the aim of this work is not to compare adversarial versus inquisitorial justice systems.

science can not offer a guarantee of always obtaining the correct result, nor to answer questions of responsibility. It only proposes to treat information or data in a predefined way, according to the scientific standards applied to it, without the guarantee of a zero error rate. Additionally, the assumptions and underlying theories are based on humans as well, and therefore might - and certainly do to some amount - contain errors. Consequently, (forensic) science cannot replace the human decision-making process in the legal context (or other areas) nor the moral and ethical responsibilities that are linked to these decisions. Furthermore, on a more down-to-earth view, the role of (forensic) science is to offer all possible help to make a decision. However, it is the duty, the right, and the privilege of the legal actors to decide on the propositions at hand and it is not to the forensic scientist to provide such answers but only to contribute to the decisive process providing supporting information[92]. This (blind) reliance might - to speak frankly - substitute the heavily criticised dependence of forensic science from the prosecution by shifting it to the judge.

The discussion has, until now, focused more on the understanding of evidence than on the presentation itself. However, this point also has its importance, as several factors can intervene at this stage. One form of presenting the results is the written expert report. What exactly is included, how it is formulated, which details are given, and whether different interpretations of the same result are discussed or not, have been subjected to discussion by various authorities. Depending on the jurisdiction, experts are more or less often invited to explain (orally) their report in court.

Research has also been done to evaluate the influence of the presentation of evidence in court on the judgment or the severity of the sentence (Enescu, 2008). Several factors might influence it: the use of audio-visual representation media (Douglas et al., 1997; Feigenson, 2003), the criterion influencing the credibility of the expert such as an academic title (Cooper and Neuhaus, 2000), or even the order of testimony presentation (Kuhn and Enescu, 2007). The control of these factors is not only in the hands of the experts presenting the evidence, but also, or even more, in the hands of the legal “participants” who establish criteria as to what is evidence, and therefore might also easily say how it should be presented. The debates on admissibility in the USA could be cited as an example, as mentioned before.

Summarising the points presented here, the key word is communication. This includes basic problems such as misunderstanding of vocabulary, but also of underlying concepts or on the weight to be attached to the messages. While the focus here has been on inter-human difficulties, the next chapter will consider specificities inherent to any human being and how they relate to forensic science.

[92] There could be many explanations for its presence; there might for instance be the possibility of error in the analysis process, etc.

2.3.2.5 Errors due to various psychological reasons, such as observer bias (Risinger et al., 2002)

As already mentioned, one of the main arguments advanced for criticising the current quality of forensic science work is the subjective way things are done. This includes the way work is organised (is the technique scientific?), how it is performed (is the person trained in a scientific way?) and if the findings are interpreted in an objective way, without being influenced by cognitive biases, for instance.

These **psychological aspects** are related to the human condition and are not specifically linked to forensic science or judicial errors. This is of interest for two reasons. First, the subject has already been treated in other fields, has been analysed, and solutions have been found that have and/or can be adapted to the present specific context. Second, the fact that the discussion is about human characteristics minimises the "misconduct" linked to these psychological effects. It is difficult to perceive somebody as wicked because he or she is simply acting erroneously due to his/her human condition.

But what exactly are these psychological effects and what do they do? The answer to the last question is simpler than to the first one. They influence the observations and decisions made by an individual (Busey and Dror, in press). The same happens for decision-making (Tversky and Kahneman, 1974), which is discussed in more detail in the next chapter (2.3.3). The main reasons for these biases are believed to be a combination of evolutionary adaptations of the human being for higher efficiency (survival) and learned behaviours. For instance, the observation of some dissociated lines will, with help of the memory, experience, etc., be instantly completed into an entire image or picture. This process considerably accelerates the time needed for an observation and subsequent decision making (to run away from a lion, for instance). Another example is the ability to recognise somebody who is far away only by the person's shape. The same mechanism is involved in the decision process. If, even for the resolution of a simple problem with few factors, all of those factors as well as their interactions have to be analysed, valuable time is lost. Humans therefore make a selection of factors to consider for a given decision. For instance, with a quick glance we decide whether the person we meet at two in the morning in a deserted street is dangerous or not (for our safety) without stopping to think about all possible risk indicators. In summary, humans use shortcuts that help to accelerate observation and decision-making processes.

These shortcuts work well for most common cases, but will fail for some exceptions. And it is within these exceptions that potential error sources in forensic science can be situated. In nearly all previously listed stages where errors can happen, these possibilities have already been illustrated.

However, one real example might just recall some of these biases (Donzé, 2005). In Bern, the police were called to a "suicide scene". Allegedly, a woman had jumped from the roof or balcony of a house. The neighbour who called the police reported that she had seen Mrs. X jump from the roof/balcony. The dead woman was then visually identified by her daughter in the morgue.

Several days later Mrs. X returned, astonished to see that she had been declared dead. In fact, she had been on holidays for some days. The initial question: *How could this happen?* might be transformed into *what biases were present?* The policemen who first arrived at the scene were informed that the dead woman was Mrs. X. Expectation as to the identity of the victim was created. This could be named an expectancy effect or bias. They then rang at the flat. Nobody answered (as indeed the inhabitant was on holiday). They felt confirmed in their hypothesis of the victim's identity. At the morgue no identification methods were employed (DNA, fingerprints, odontology) and the daughter of the supposed dead was "invited to say good-bye to her mother" (and visually identify her mother). This is a (perfect) example of a confirmation bias. It was only the "return of the dead" that proved the initial hypothesis wrong. Thus, an initial untested hypothesis, led to the final result, a wrong identification. Furthermore, all along the potential safeguards existing for "checking" the initial hypothesis were missing. This includes not considering a second hypothesis as well as neglecting a correct identification procedure.

These kind of biases have to some extent already been studied in areas such as expert testimony (Devenport and Cutler, 2004), line-ups (Malpass and Lindsay, 1999), and the evaluation of the credibility of a witness/suspect by police officers with/out additional training (Meissner and Kassin, 2002). Moreover, abundant research in different mock jurors' experiments exists (Cooper and Neuhaus, 2000; Feigenson and Dunn, 2003; Kerstholt and Jackson, 1998; Koehler, 2001; Koehler and Macchi, 2004; Spackman et al., 2002), which may also include evaluating the impact of colour versus black and white photographs (Douglas et al., 1997).

Studies have been much more limited in forensic science (see Chapter 2.3.4 and 3.2.3) though observer effects have been suggested to "undermine to some degree the reliability of virtually any form of expertise." (Risinger et al., 2002, p.5) This is why a considerable part of this research (Chapter 3) focuses on studying such factors empirically.

In summary, an inventory of non-voluntary errors in forensic science has been presented. Problems may occur at the moment of evidence collection, the choice of tests in laboratory, in the analytical and statistical interpretation, and in the presentation of evidence as a report or in court. Psychological factors such as observer effects/biases seem to contribute as well to errors. The questions remains what can be done to minimise these kinds of errors.

2.3.3 The lack of literature on the decision making process in forensic science (laboratory work and evaluation of results)

As seen in the previous overview of possible errors in forensic science, at each stage minor and major decisions are taken in a conscious and unconscious way. They are, because of the heavy reliance on human observation and judgment, for instance in fingermark comparison, crucial to forensic science and have been often attacked as being unscientific and subjective. In other

areas where (subjective) decisions are important (e.g. analysis of radiographies), the mechanisms underlying such decisions have been studied at least to some extent. This has been possible because decision-making/taking has developed over the years as its own field (Mellers et al., 1998) of interest with applications in different areas. A good introduction to its (possible) use in the legal context is given by Howard (2002).
In forensic science, as summarised by Bernard (2005), existing publications tend to focus on DNA and Bayesian networks, where it is not too difficult to build an abstract model. The interest has been so far less on deciding what test to do or how to take the decisions. The aim here is not to give an overview of what has already been done, but just to illustrate that existing research tends to limit questions of decision-making to the interpretation of analytical results and not to ask itself on what decisions those tests and results are based in the first place. One explanation for this situation might be that for some areas of forensic science everyday work is based on empirically made experiences and not on carefully researched methodologies developed over years. So it is the basis, or the formalisation and understanding of these processes that might be questioned and perhaps even questionable, rather than the use of established procedures[93].

To know about the exact way decisions are taken might not be essential in every situation nor for all tasks performed by forensic science personnel. Thus, a laboratory technician following a strict analytical protocol in DNA profile analysis does not necessarily need to know details on underlying theories and concepts in the field. On the other hand, the senior scientist in charge of the laboratory should know all reasons and justifications for the way decisions are reached.

An existing example of formalising this process is case pre-assessment (Cook et al., 1998). In this approach traces have been collected and the background of the case is known, the possible analyses are studied and potential results anticipated. The added value of these examinations is then pondered, in order to find the best tests and sequence of examinations, etc. for the present case. The nature of the tests will depend on the type of specimen, the presence of statistical data, for instance the frequency of the observed characteristic in a given population, and the way the results will be presented. One way of doing this is by using a likelihood ratio based framework. Before any examination is made, available information on the case and samples is used to “speculate” on what kind of results could be obtained. This is done by estimating the likelihood ratio of the potential outcomes given the propositions at hand. This implies an evaluation of any “analytical” results or observations under at least two competitive and mutually exclusive hypotheses, as explained earlier. Then, it is decided if and what examinations are made. Following that, the values obtained are fed into the previously constructed model and the new likelihood ratio estimated. This can help deciding on the interest of doing additional tests or directly help estimating the value attributed to the results.

[93] There is nothing wrong with experience-based knowledge, as long as its advantages and disadvantages are clearly understood.

This way of proceeding offers several advantages, for instance that automatically more than one hypothesis is evaluated. By this, if done correctly, the danger of confirmation bias (2.3.2.5) should decrease, as more than one explanation for the occurrence of any event has to be found. To assess a likelihood ratio, statistical data is (most often) needed. To have it, such data has to be collected and analysed first, thus already forcing systematic acquisition of data in the field. Furthermore, instead of presenting the results based on one's own experience, however large and exhaustive, the reliance on structured data allows for a more balanced evaluation. Indeed, instead of considering only the cases one has been confronted with - which might be biased by the fact that only a certain category of cases will be retained, for example - systematic data acquisition should consider a more randomly distributed population, thus giving a better representation of the actual features of the population in question. Finally, the consideration of a logical framework of evaluation such as the likelihood ratio based approach forces the evaluator to think about a given set of factors, thus minimising the risk of forgetting an important one. He will for instance consider all possible methodologies and techniques for all specimens and samples before making a selection, both in the number and type of specimens and samples to analyse, but also on the methods to be used.

The idea of pre-assessment as described, gained more weight when the FSS (Forensic Science Service) went independent and started a client-based approach towards the police. As their costumers wanted cost-effectiveness, the need was to find the best and most beneficial forensic specimen analysis to get the most informative result with the best selection of efficient techniques in order to obtain the expected data. The testing (and processing) is pre-evaluated and adapted to the type of result expected and agreed upon with the customer. The main interest of Bernard's research project (2005) in this context is the effort it shows to formalise the decision-making process in forensic science. In other words, the same decisional framework should be applicable to any choice and decision made. This does not imply following a strict (analytical) protocol, bur rather to use the same criteria each time for a similar decision to make. The "protocol" will consist of a list of questions that need to be answered and not of actions to take. They will consider several key elements of the process, by taking into account the values attributed to factors judged important. In consequence of this choice, an outcome will take place, which according to the criteria of the evaluator should correspond to the best possible outcome. Furthermore, by applying a formalised framework, a certain guarantee of repeatability of decisions should be reached[94]. As already mentioned earlier, to consider a structured decisional model should minimise the influence of subjective factors, and thus the risk of (observational) biases in the work performed. This had not been tested empirically here, but seems to be an interesting element to be tested in future research.

[94] As seen previously, repeatability is one criteria of scientificity.

2.3.4 The lack of empirical studies of the observer (bias) effect (Risinger et al., 2002) in forensic science

Why observer effects are perceived to have a negative influence on the quality of forensic science work has been described previously and will only be summarised shortly here. The subjective nature of the work seems to make it particularly vulnerable to influences of human factors that can include observer effects. Special risks are seen in expectation and suggestion (Risinger et al., 2002) for areas relying primarily on human observation and judgment, for instance fingermark individualisation.

However, while this risk of observer effects has been suggested, there are very few published examples of the actual influence of such observer effects in forensic science, or more generally, in judicial errors. Exceptions include research by Dror et al. (2006) and Kerstholt et al. (2007). They will be discussed in more detail in Chapter 3, which studies observer effects and biases empirically. Here the focus will be on the (quasi) absence of such studies in forensic science.
The issue of observer effects in forensic science gained momentum with the publication “The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion" by Risinger et al. (2002). By comparing forensic science to other fields it demonstrates that these influencing factors could have non negligible consequences in daily work. Risinger et al. (2002) put it as follows:

> “that there are certain factors which, when present, undermine to some degree the reliability of virtually any form of expertise. Further, we will show that the extent to which reliability is undermined depends not only on the presence of such factors, but on the characteristics of the expertise at issue, most particularly the degree to which it depends on subjective human judgment. Moreover, we will show that there is an entire established constellation of expertise, celebrated in popular culture and heretofore generally admissible, in which such factors form a rampart and uncontrolled part of normal practice.” (p. 5)

They go on to say:

> “the factors we refer to are primarily expectation and suggestion, which drive much of what is globally labelled “observer effects” in social psychology and research methodology. And the constellation of expertise is “forensic science” in general, and especially those forensic science practices utilizing subjective human judgment as their primary instrumentality, and not based on techniques derived from normal science methodology.” (p. 5-6)

What influence those factors have will be, at least partially, be studied with the empirical tests in Chapter 3.

Why (empirical) research on the topic is absent will be developed next. There seem to be two main explanations: there is a lack of interest in the topic and/or a lack of feasibility of the research. Into the first category falls the missing realisation that subjectivity in forensic science can (negatively) influence the outcome of the examinations and tests performed and that this is a topic that merits research. Into the second category fall the difficulties of implementing such research projects, for instance by having access to a relevant population. On a more general note, the question might be whether by doing some research it will be possible to minimise subjectivity altogether or not; or on other words, to do something about the observer effects. Next, the aim will be to explain why the study of decision making is important in this context and why empirical research can make a difference on the question of subjectivity and help minimise its (negative) impact. Whether such tendencies exist already in forensic science will be discussed as well.

A decision is made considering several factors that might have a variable "weight" and thus impact differently on the decision (Howard, 2002). To know what those factors are and which are decisive for the choice to be made will explain how the decision is made. Experience, observation, and logic can help construct a canvas containing those factors. However, to know if this proposition reflects reality it has to be empirically validated. To do this, structured data has to be collected. It is here that empirical tests will help to understand how decisions are taken. And if it is known how decisions are made it is also possible to see what influences them or what is needed to keep them constant. In other words, it is suggested here, that it offers a tool to minimise subjectivity.

There are some indicators that tend to show that many processes in forensic science include a subjective element and that this has, at least implicitly, been recognised. For instance in the area of fingermark individualisation, the tendency has been (and still is) to abolish the numerical standard of minutiae for identification[95]. One reason is (very) probably the often mentioned absence of statistical data to justify one threshold rather than another (however, in DNA the amount of loci considered has also varied with the years and has evolved). Another aspect is the considerable variation observed in the annotation of minutiae by experts (Evett and Williams, 1995). While some of them can be explained by a different understanding of what a minutia is, etc., discrepancies seem to be greater than could be expected by these.

Why variability is so important would be interesting to know. Are there people born with greater ability to compare fingermarks? Or is training/education/experience essential for acquiring these abilities? Are there technical factors that influence this process, for instance the quality of the image etc.? Or are there human factors that are central, for instance the current mood? Is there an influence of factors possibly creating observer effects? Or are those questions secondary to,

[95] "No valid basis exists to require a predetermined number of characteristics to exist between two fingerprint impressions in order to establish positive identity." (IAI, 1973, p. 14)

for instance, finding a good method/technique for developing fingermarks on a tape, one of the more challenging surfaces for fingermark enhancement?

Wertheim (1996) summarises factors that according to him make somebody an able fingermark comparison professional. Training, experience, talent, motivation, and daily variables are discussed. All of these factors are intuitively appealing. Presenting them as resulting from individual experience however allows only to say that they might have an impact and incidence, but not how this would manifest itself for a given fingermark comparison. In other words, Wertheim does say that training is important, but not how this training should be structured.

Nevertheless, it would be very interesting to have answers to these questions. Why this is only partially the case is perhaps also explainable by the fact that the discipline is rather young and focused more on “what works” and less so on “why does it work or not”. Furthermore, disciplines that have asked more of these questions are either related to other non-forensic science research areas (e.g. DNA) or the findings they have are rather basic. This means that more complicated topics have not found that much attention yet, perhaps also because their testing is more complicated. To find out what factors determine a good fingermark expert is more difficult to study than to know whether the new reagent reacts better or not than Indandiones to latent fingermarks.

In other areas with similar tasks, for instance the interpretation of radiographies (Obuchowski, 2003), some of these traits have been studied. Whether they are of interest for forensic science work will only become clear if similar studies are made. This is one of the reasons this research includes experimental studies of such factors, as presented in Chapter 3. One of the main interests of those studies has been the fact that specific elements influencing the interpretation of radiographies for example, could be isolated. In consequence, it is also possible to make specific, empirically based recommendations for change. Looking at the lack of studies on observer effects from outside forensic science, there seems to be missing research on factors that could have a considerable impact on the work performed. From the inside, it might be just one more facet that needs to be researched, together with a lot of other under-developed areas. One of the reasons why this has not been given more priority might be the difficulty of the task. But perhaps it is only the tacit acceptance that subjectivity is an important and unavoidable component of forensic science that explains part of the lacking research. Why try to understand differences if you know that they are due to subjectivity? Will knowing more about what contributes to subjectivity change it? It might be argued here that given the results in other fields it will not be possible to eliminate the subjectivity of the process. However, it might be feasible to delimitate it and, after having understood the specific factors forming it, make practical recommendations that can be implemented and be tested for their usefulness to minimise or at least control its influence.

2.3.5 The fact that people tend to be less critical of exculpatory than of inculpatory evidence, especially when a wrongful conviction is supposed. Therefore there is a need for symmetry (Edmond, 2002)

This topic has been discussed in the form of a publication, which will be briefly summarised next. The complete paper can be found in appendix 9.1.1.

Schiffer, B. and Champod, C., (2008). "Judicial Error and Forensic Science: Pondering the Contribution of DNA Evidence", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 33-55.

Forensic science has a double role of, on one hand, being identified as a source leading to judicial errors, but also being used as remedy for discovery, for instance by DNA profile analyses to exonerate innocents. Edmond (2002) argues that scientific evidence to inculpate may not be perceived in the same way as the same kind of evidence used to exonerate a convicted person. Using the example of DNA profiling, the “new gold standard” (Lynch, 2003) this proposition is studied considering two British case examples. The first case is that of Michael Shirley, arrested in 1987 for the rape and murder of Linda Cooke. The evidence collected included a blood group test (of semen), a footwear mark, and bruises and scratches on the suspect. The second case is that of Mark Dallagher, arrested and charged for the burglary and murder of 94 years old Dorothy Wood. Based on incriminating evidence involving ear prints, Dallagher was sentenced to life imprisonment in 1998. In both cases sensitive DNA tests, called LCN (Low Copy Number), excluding the defendants played a critical role in the decisions that led to free both of them in 2003. Both analyses show that the strength of the DNA evidence, although extremely useful to the fact finder, should not be considered sufficient in itself to exonerate a defendant. The paper also examines why the difference in perception exists as to forensic science before or after trial. As mentioned before, it is the difficulty of distinguishing forensic science “as such” from the “use made” of it that is the reason for this unequal treatment observed. Indeed, whether DNA profiles are used to inculpate or exonerate does not modify their nature, however the interpretation made may vary considerably. This reflects communicational issues as raised also in Chapter 2.3.2.4 rather than “scientific” errors.

2.4 Comments

Initially, a structured literature review had been proposed (Fink, 1998). However, given the amount and type of publications found, the systematic gathering in a given time-window was abandoned in profit of an exhaustive research. Some of the questions asked were considered only in a small number of publications and too strict a limitation would not have allowed for a more comprehensive overview. At the start it was also proposed to consider different kinds of publications separately. This has only been done for newspaper articles, for the previously mentioned reason of limited material. Other factors however, such as the background of the

authors, as well as the content of the publications, have been considered when answering the questions asked. Finally, the findings have been presented as a descriptive, qualitative literature review.

Because not much on the interaction of forensic science and judicial errors is known or has been published, the aim here was to give an extensive overview on as many aspects as possible. In order to avoid getting lost, some questions of special interest were used to structure the information extracted from the literature review. While this risks sometimes the creation of artificial separations or provocation of repetitions as for instance when discussing errors due to various psychological reasons (2.3.2.5) or the lack of empirical studies on observer effects (2.3.4), it helps nonetheless to give a clear structure to the overview.

The question on the status quo and the small amount of publications on errors in forensic science by scientists and non-scientists has been answered by a comparison of the publications found and their respective authors. Several explanations for the lack of publications on forensic science as a source of judicial errors can be found. One of them might be that forensic scientists are not being as critical towards their own work as necessary.

For some of the topics, another way of representing the results could have been chosen as well. The inventory of possible errors for instance, could have been structured in function of the potential actors (who did them), the potential reasons (various types of exterior and interior pressure, etc.) the potential results ((not) likely to have an influence), their occurrence (exceptional or regular), etc. Given the existing data sources, the chronological approach presented the most comprehensive way of reviewing them. Indeed, empirical data on the incidence of various types of errors, their gravity and influence, etc., is non-existent.

The nature of forensic science makes it impossible to establish an exhaustive list of all potential error sources. Each specimen collected is different, though the techniques and methods used for examination and analysis are, to a certain degree, standardised. Certain areas are difficult to formalise, such as for instance evidence collection on a crime scene. It was therefore thought more useful to give a conceptual view of the life cycle of a typical specimen rather than list all possible variations for each specific for all kind of traces.

This means that error types rather than errors as such are described. The process from the search for traces on the crime scene to the presentation of evidence in court has been divided into several stages. They characterise important steps of the forensic science process. This division does not always correspond to all real-life situations, but allows a more organised overview of presenting the life cycle of traces. Furthermore, depending on the legal and forensic structures in place, many of these elements will vary to a greater or lesser extent[96]. Besides the theoretical consideration of potential error sources inherent to forensic science, case reviews and studies on judicial errors due to forensic science have been consulted to establish and

[96] In general, the chronology chosen is based both on the Swiss situation but also on the “best practice” recommendations, as far as they exist and have been accepted by the “relevant scientific community”.

complete this overview. To take into account only the reviews based on discovered errors would offer an incomplete picture of the situation.

Furthermore, errors have a tendency to cumulate, for instance by ratifying a first error by a second (Huff et al., 1996). The chronological presentation does take into account this phenomenon, not so the other possible approaches. It has been thought necessary to complement the sequential view by emphasizing one error source in particular, psychological biases. The gain is threefold. First, one specific error is pursued across the chronological view and offers thus a further and different perspective on the subject. Second, psychological biases as potential error sources in forensic science are a rather new topic under discussion in the field (Risinger et al., 2002) and merit special attention because of its “subjectivity”. And third, they unite several interesting characteristics “typical to forensic error”, namely: that it is recognised in other existing domains (such as for instance medical research, eyewitness testimony etc.), that it does not only have to be perceived as negative (typical “judicial bias”), that it has not been studied empirically so far, and that it is independent of the will of the acting individual, or in other words, non-voluntary.

The introduction of the concept of critical criminology helps understand the nature and scope of many existing publications on the subject, especially their focus on the unfairness of the justice system. On the other hand, the need to become more proactive - according to some commentators one of the weak spots in current critical criminology - coincides with the same necessity in forensic science to be more aware of potential errors and error avoidance mechanisms. It is in this context that the recurring question of the place of subjectivity in forensic science is of special interest.

In order to overcome the limitations of the comprehensive review on potential errors in forensic science, one single type of error - psychological biases, especially observer bias - has been selected for a more detailed study. This has been done both in an abstract and an empirical way, which will be presented in the next chapter.

In summary, this chapter tried to give an overview of aspects important to the discussion of forensic science and judicial errors. This included a chronological and comparative summary of events having influenced the topic, but also a comprehensive inventory of potential error sources at various stages of forensic science work. The special focus on one possible error source, observer effects, has been explained. Furthermore, the different perception of forensic science to inculpate or exonerate (published as a chapter in a book) has also been included in this overview of the topic.

3. Empirical tests

"'It is all well, Poirot, but I think you are falling into the habit of despising certain things too much. A fingerprint has led sometimes to the arrest and conviction of a murderer.'
'And has, without doubt, hanged more than one innocent man,' remarked Poirot dryly."
(Christie, 2001, p. 24)

This chapter focuses on the empirical testing of observational biases, which according to some authors might lead to misleading forensic statements that in fine can lead to judicial errors (Risinger et al., 2002). A short introduction will explain why questions of subjective elements influencing the forensic science process are of interest in the context of judicial errors, by the example of fingermark individualisation. A short theoretical outline will introduce fingermarks and observer effects. The combination of both will be presented next, focusing on the factors that have been studied with experimental tests. Two papers, one of them published, summarise the findings of the experimental tests on various factors potentially influencing the ACE-V (Analysis, Comparison, Evaluation, and Verification) process. Finally, some comments on methodological and theoretical issues close this chapter.

As already introduced in the literature review, the question of the role and impact of subjectivity in forensic science is gaining more importance. One way this subjectivity presents itself is the question of the influence of observer effects or observer biases (for the definition see 3.2.2) on daily routine work, as for instance fingermark individualisation. While the question has been asked in a rather theoretical way (Risinger et al., 2002), there is not much experimental research available evaluating if and how these observer effects could or do influence forensic science work. An overview of existing research, by for instance Dror et al. (Dror et al., 2005; Dror and Charlton, 2006; Dror et al., 2006; Dror and Rosenthal, 2008), will be presented under 3.2.1.
Some instances of error, such as the misidentification of a fingermark by the FBI in the Madrid Bombing case (Mayfield) (OIG, 2006) or the case of Stephan Cowans (Cole, 2005) seem to show a vulnerability of the individualisation process, one of the "forensic science practices utilizing subjective human judgment as their primary instrumentality" (Risinger et al., 2002, p. 6). Fingermark individualisation having been the old "gold standard" (Lynch, 2003) of forensic science evidence makes it especially important to study its possible failings.
While such questions have been raised, mainly by Risinger et al. (2002), not much experimental data exists to further explain or corroborate this hypothesis, which will be illustrated in the literature overview under 3.2.1. The research presented here tried to study some of the factors that seem to influence the work done, adding thus a subjective dimension to forensic science. These factors will be presented next.

3.1 Research questions asked

Based on the literature review (Chapter 2), and as summarised in 3.2, some factors or points of interest emerged. The aim was to evaluate them experimentally whenever possible:

I. Observer bias exists in forensic science (Risinger et al., 2002) and can be established empirically, but "it is easier to accuse someone of bias then to actually establish that a judgment is biased" (MacCoun, 1998, p. 263) (Ehrlinger et al., 2005).
   (1) It exists at various stages of the ACE-V (Analysis, Comparison, Evaluation and Verification) process.
   (2) It also exists for various types of identification evidence (e.g. fingermarks, footwear marks etc.) although not all of them will be tested empirically in this research.
   (3) Various factors/situations do/might have an influence: multi-trace case, superfluous information on the suspect (Kerstholt and Jackson, 1998), gravity of the offence, etc.
   (4) The "forensic" approach per se might lead to a certain type of bias (confirmation bias)[97].
   (5) Education/training can augment/diminish the risk of being influenced by those biases.

II. Problems/mistakes will occur much more easily in dubious cases (partial, difficult, contaminated traces) than in straightforward ones (defined by the technical difficulty of the comparison).

III. Forensic science alone cannot "create" a judicial error. It needs a set of other factors/errors/circumstances, a concatenation of error for the final outcome.

Out of these aspects, a selection was made as to the factors to be studied empirically, which will be presented in the experimental design (3.3) and commented upon (3.4) later. Why these factors have been selected will be explained in 3.2.

3.2 Theoretical outline

3.2.1 Fingermarks

Extensive information on general and specific aspects of the comparison of fingermarks can be found for instance in Ashbaugh (1999) or Champod et al. (2004). Only some specific points useful for this research will be summarised here. Papillary ridges form before birth. Extended comparison practice has shown them to be specific for each human being. Except for profound skin damage (scars, specific illnesses, etc.) they remain the same throughout a person's life and are thus considered immutable. Due to "perspiration" or contact with other substances, the pattern of the ridges can leave marks on contact surfaces. Depending on various factors, such as the finger, the "ink" or substance used for transfer, the type and nature of the receptive surface, the type and time of contact, etc., the quality of the fingermark may vary considerably.

[97] Comparing crime scene (trace) material to suspects is considering a limited population where the expectation of finding correspondence might be higher than the contrary.

Development and/or enhancement of the fingermark can be achieved by using a variety of techniques and methodologies, such as ninhydrin and cyanoacrylate fumigation (Champod et al., 2004, Chapter 4). Most times the fingermark is then photographed and/or scanned and the comparison process is made by using this two-dimensional representation of the original trace. From the initial deposition/apposition of the fingermark on a surface to its photographic fixation, several factors influence its quality and thus its subsequent study. A good knowledge of possible intervening influences is necessary to analyse a fingermark correctly. For instance, distortion modifies to some extent the mark left by the three-dimensional fingertip on the receptive (perhaps two-dimensional) surface. A further element is that although papillary ridge patterns are considered as "unique", this does not automatically apply to (partial) fingermarks left on a surface[98]. Indeed, a small but important step of inference is needed to bridge this gap.

Despite the longstanding and well discussed lack of scientific foundations of fingermark individuality, as proclaimed by several authors (see Table 1), fingermarks have been widely accepted as evidence in court for more than a century. The criteria for individualisation - or often even misnamed identification - do vary from one nation to another, though the individualisation process is most times similar. It comprises roughly a protocol called ACE-V[99] (Analysis, Comparison, Evaluation, and Verification) or some less structured variety of it (Ashbaugh, 1999, chapter 4; Triplett and Cooney, 2006). The form given to the examiner's statement changes throughout jurisdictions, some claiming the necessity of a numerical standard (e.g. formerly 12 points in most parts of Switzerland[100]), whereas others argue that holistic reasoning offers a more scientific approach to this question (IAI, 1973). Whichever side one takes, the fact remains that discussions on the quality of fingermark examination, be it the already mentioned claimed lack of underlying scientific research, the discovery of errors in various cases, or the question of admissibility (especially for the USA) are increasing (see Table 1 for an overview of publications). The most famous case example is the FBI misidentification of Brandon Mayfield (OIG, 2006). Cole (2005) gives a more general overview of cases, as detailed below. As to the debates on admissibility, instances of refusal have happened in the USA (Llera Plaza, discussed by Mnookin (2003)), though the decision has been revoked afterwards in most cases (see for an exception: State of Maryland vs. Bryan Rose, 2007; State of Maryland vs. Bryan Rose, 2008). One specific point gaining importance is the risk of observational bias, which will be discussed in the following section.

---

[98] The information might be limited on the partial mark, for instance only three minutiae are visible. While a high amount of visible minutiae might allow individualisation, it does not forcibly follow that the observed combination of three minutiae is sufficient for individualisation as well.

[99] **Analysis**: Observation of papillary ridge details such as general pattern (arch, left/right loop, whorl; also called first level characteristics), minutiae (ridge endings, bifurcation or dot, also called second level characteristics), and sometimes shape of ridge edges and pore position (third level details) on the fingermark and separately on comparison prints.
Comparison: The features observed before are compared between the fingermark and the comparison print(s).
Evaluation: Weighting of the features compared before. Sometimes this and the previous step happen at the same time.
Verification: This step is supposed to allow checking on the process, objectivity, and the result(s) obtained by a peer review process (Ashbaugh, 1999, chapter 4).

[100] Recently (November 2007) Swiss fingermark comparison experts decided to move from a numerical to a non-numerical standard for the identification of fingermarks.

Table 1 Selection of publications on misidentification and admissibility questions in fingermark individualisation

Misidentifications The few existing publications into the potential influence of observer effects on fingermark individualisation are limited to the post-error study of cases or are interested mainly in the late stages of the individualisation process.	
OIG (2006)	This report investigates the "misidentification, investigation, and detention of Mayfield" (p.1). The causes of the misidentification are presented shortly (p. 6-12) and then developed in more detail. The analysis of the fingermark misidentification is found on p. 29-66, 95-126, and 127-194 (concerning different aspects).
Broeders (2005)	This general paper summarises actual problems of the perceived (un)reliability of forensic science in a comprehensive way. It also summarises the cases of Mayfield, McKie and Cowans. It is one of the few/only publications by a European author.
Cole (2005)	This paper discusses errors in fingerprint identification in a very comprehensive way. It includes an overview of 22 cases of fingermark misidentifications.
Gysin (2005)	Swiss newspaper article on 22 misidentifications (based on Cole, 2005).
Meili (2005)	Based on US cases (and Cole, 2005) the Swiss situation is studied, interrogating various Swiss interview partners.
Zabell (2005)	The Cowans and Mayfield case are discussed, scientificity, ameliorations, context effects mentioned, and DNA is offered as the good example of how to proceed.
Admissibility questions[101] Admissibility according to Daubert/Kumho standards[102] and others are often discussed. Most publications infer that the criteria are not met yet and propose recommendations how to reach "scientificity", based often on research proposals.	
Cole (1998)	Historical overview on how latent fingerprint identifications started and how latent fingerprint examiners (LFPE) have defined their field and work in it.
Cole (2006)	This paper discusses the lack of validation studies in fingerprint identification, as well as several fallacies which have not been solved, according to the author.
Cole (2008b)	This paper summarises the tendency to describe fingermark individualisation as an opinion rather than a fact in order to "cope" with recent critical comments and misidentifications.

[101] The discussion of admissibility in this form is found mostly in US publications, one of the reasons being the way evidence is handled there in comparison to other countries.
[102] Daubert criteria: relevance for the case and reliability of scientific method which is based on four points: 1. empirical testing of the method; 2. peer review and publication; 3. error rate; 4. general acceptance in the relevant field.

Epstein (2002)	This paper discusses how and why fingerprint identification does (not) fulfil the Daubert/Kumho legal standards, though it has been accepted for over 90 years.
Haber and Haber (2004)	Existing anecdotal data from various sources such as proficiency tests are discussed and are judged unapt to establish error rates, as needed in the context of admissibility questions.
Haber and Haber (2008a)	This paper focuses on the question if the ACE-V methodology is valid by first describing it, then showing where it is failing according to the authors, proposing necessary experimental tests to validate it and finally, refuting recurring arguments that it is valid. Champod (2008), Cole (2008), Haber and Haber (2008b), Koehler (2008b), and Mnookin (2008) comment this paper in the same edition. The different points of view on the subject thus represent an overview of the current of the discussion on this topic.
Koehler (2008a)	This paper discusses what error rates exist in fingermark individualisation, why today proficiency tests give no satisfactory answer to questions of diagnosticity, reliability, and error rates. It makes recommendations how these tests could be improved by for instance “calibrating” the difficulties of the fingermarks used and double blind testing.
Mnookin (2001)	The historical debates on admissibility of fingermark individualisation and DNA are presented; recent challenges on the first are presented and solutions proposed.
Mnookin (2003)	By using the two decisions in United States v. Llera Plaza, the limitations of fingermark individualisation validation are discussed, explaining it within the historical development of admissibility (Frye, Daubert) and the more critical view of DNA profiles.

3.2.2 Observer effects

Observer effects are studied in cognitive psychology, which is interested in learning more about the way internal mental processes such as memory, thinking, and language work. The interest here is mainly on perception and decision making, more precisely on factors that structure and influence both these processes (Dror and Mac-Kenzie, in press). These effects or biases are in general named either in function of the element that provokes them or according to the resulting consequence. An example for the first would be the *recency effect,* which is to weight recent events more than earlier events (Enescu, 2008), whereas an example for the second would be *hindsight bia*s, the tendency to believe that events that happened were more predictable than they actually were when they happened or before they happened (MacCoun, 1998). Terminology used for such terms can be variable, one of the reasons it is necessary to agree on

them beforehand[103]. For the present work the focus will be on observer effects which are part of the larger family of cognitive biases, as mentioned before, and the definition used is the one given by Risinger et al. (2002, p. 6).

> "An elementary principle of modern psychology is that the desires and expectations people possess influence their perceptions and interpretations of what they observe. In other words, the results of observation depend upon the state of the observer as well as the thing observed."

The same authors explain also that "context and expectations influence an individual's perceptions and interpretations of what he observes" (Risinger et al., 2002, p. 12) and that some of these effects or biases have already been observed in other research fields. An example would be the possible influence of testimony order in court judgments (Enescu, 2008). Those effects are suspected to influence other forensic science fields as well, for instance DNA profile interpretation (Krane et al., 2008).

A further distinction can be made between observer effects and observer biases. The first are randomly distributed whereas the second happen systematically[104]. The focus here is on observational biases or, in other words, systematic error. One example is confirmation bias, the tendency to search for or interpret information in a way that confirms one's preconceptions. Parmley (2006, p. 46 ff) explains this bias in the context of diagnostic decision making in psychopathology. The explanation given for the occurrence of biases is that they happen in the context of cognitive short-cuts or heuristics which are used to make a judgment or reach a decision in a simpler way (Fariña, et al., 2003)[105].

Literature on observer effects in forensic science and/or the way observations are made is very limited compared to other decision fields (e.g. Malpass and Lindsay, 1999). There are publications in other domains, for instance in the medical field (Obuchowski, 2003; Whiting et al., 2004). Some research exists for fingermark individualisation which will be presented in the next chapter after a short overview of existing ways of studying such phenomena.

Research tools in (cognitive) psychology comprise a range of qualitative and quantitative methodologies inherent to various social sciences, including also criminology. To study phenomena in the context of (observational) biases and effect experimental tests have been regularly used, as illustrated in research in the medical field for instance (Obuchowski, 2003;

---

[103] Thus for instance, the "experimenter effect" defined by Risinger et al. (2002, p. 19ff) seems to correspond to the observer effect as named in other social sciences (McCall, 1984), and is understood as the influence of the observer/experimenter on the behaviour of the (living) subject he is observing.

[104] "In standard parlance, the term "observer error" refers to errors that are randomly distributed and therefore self-cancelling over the long run, while the term "observer bias" refers to errors that are not random but systematic. Note, however, that random observer error may be a serious problem in any process which is not cumulative, but which relies on validity in regard to each individual result." (Risinger et al., 2002, p. 13)

[105] In other words, the decision will be taken using an approximate "model" of the problem which considers only a limited amount of factors, thus constituting the cognitive shortcut.

Whiting et al., 2004). Tests that have been used in the domain of fingermarks will be presented in the next chapter.

3.2.3 Observer effects in the fingermark individualisation process

The interest here is to see how the observer effects mentioned by Risinger et al. (2002) do affect the process of fingermark individualisation, especially the decisional stages (Johnson-Laird, 1999).

One way of representing observer effects – based on Dror et al. (2005) – is to distinguish top-down and bottom-up factors which can influence the task at hand. Top-down means the influence of the observer (and all he includes) on the object he is studying and bottom-up refers to the way the object (and its properties) influences the observer, or rather the observation which in some way is the link between both of them. These two constituents are interrelated and do offer multiple stimuli for cognitive biases. The constituents and the stimuli will be presented in more detail after an overview of publications on experimental tests related to fingermark individualisation in Table 2.

Table 2 Research publications on fingermark annotation and influence of observational and context effects

Fingermark annotation Not many studies exist that focus on the way fingermarks are annotated (by humans).	
Evett and Williams (1995)	Lack of consensus between experts as to the amount of minutiae observed, abolishment of the 16-points rule in the UK “in consequence”.
Langenburg (2004)	This paper compares the amount of minutiae found by latent print examiners/trainees and “lay persons”, the first seeing more minutiae.
Wertheim et al. (2006)	Experimental test with (non)experienced participants as to the comparisons made, errors of various types, etc. The paper discusses in detail the advantages, but especially also the limitations of the research and future projects.
Experimental studies on observer or context effects in fingermark individualisation Research is limited to publications by nearly only one group of authors.	
Dror et al. (2005)	Twenty-seven participants (not fingerprint experts) made a total of 2484 judgments on non/match for fingermarks. Three top-down influences (emotional background stories of crimes, explicitly disturbing photographs from crime scenes, and subliminal messages) were used to induce bias. The more difficult the decision on match/non match, the more influential the contextual information became.

Dror et al. (2006)	Five fingermark examiners were presented the same fingermarks they evaluated as a match in a "biasing" non-match context in a within-subject[106] design. Four participants changed the previous match decision they made five years ago.
Dror and Charlton (2006)	Six fingermark examiners made each eight comparisons which they had in the past already examined. It was studied whether resubmitting them (without their knowledge) and adding biasing contextual information would change their previous decision. This was the case in six out of 48 comparisons, both if contextual biased information was present or not.
Dror and Rosenthal (2008)	The paper asks whether experts can be biased and/or are reliable. It also describes why research is difficult in this field (e.g. availability of participants), limiting the statistical significance by small sample sizes. How this last problem can be solved is demonstrated next by using a statistical tool ($r_{equivalent}$).
Kerstholt et al. (2006)	Expectation and experience did not affect the comparative judgements made by 12 examiners on 8 shoe print comparisons.

In summary, existing research has revealed several difficulties in fingermark individualisation. Some of these seem to be inherent to the process ("natural" differences between participants (Evett and Williams, 1995; Langenburg, 2004)); others might be explained by different factors, such as observer effects (cognitive biases). For the first, published explanations are difficult to find. For the second, the theory has been presented in the preceding chapters. Its application to fingermark individualisation will be the focus of this chapter. Most of it is based on Dror et al. (see Table 2).

In this situation, the observer corresponds to the fingermark examiner and the object to the fingermark of interest. As mentioned before, these elements are interrelated and linked by the observation, which in a certain way is the "stage" on which those factors will have their influence (Figure 5).

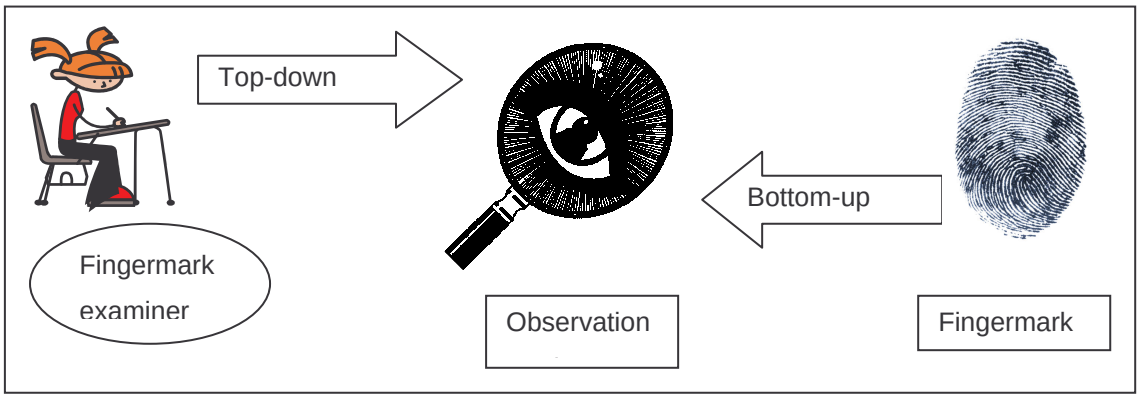
Bottom-up factors (Dror et al., 2005) can be summarised under the term of "quality of the fingermark". This includes features such as contrast, clarity of the mark, amount of details visible, etc. However, the distinction is not that clear-cut, as the fingermark and the characteristics to be observed depend also on the definition given to them by the observer. For instance, whether something is a smudge or whether it is a fingermark is decided by the observer. The same applies for the definition of ridge ending, etc.

Top-down factors (Dror et al., 2005) include all inherent (learned) and external (specific to the moment, such as expectations, context, etc.) abilities of the observer. The first include prior

[106] Within-subject design means that two conditions or factors have been studied "using" the same experiment participant, in the present case by submitting twice the same fingermark comparison in a different context, but with a temporal difference. The aim is to minimise possible third variables influencing the result, such as factors linked to the tested person.

experience and training. The second comprise the actual mind set and the cognitive biases introduced before.

Figure 5 Illustration of top-down and bottom-up factors influencing fingermark individualisation



The fingermark examiner will determine the top-down factors and the fingermark will determine the bottom-up influences. Both converge in one point, the observation.

It can be argued that the eye and the related sensorial systems, nerves etc. have to be considered here as well, as they can be trained to a certain degree to perform better in certain areas. It should also be remembered, that due to evolutionary developments, human vision has fixed some priorities which influence the way things are perceived. Thus, for instance, when presented with two objects in red and green, (same hue, luminosity etc.), then red will be considered to be “more visible” as the eye is more sensitive to reds than to greens. Some of these characteristics are purely physiological; others however are part of day to day conditioning. Dotted lines, for example, can be perceived as being one connected line (due to the “wish” of our brain to simplify information, but also because day to day experience learned that this is the most recurrent event). To distinguish between internal and external abilities of the observer, as well as the physiological faculties of the individual is not always easy. Some might have been acquired by the observer, such as experience, others can be part of the working environment, such as constant “efficiency” pressure, or might be specific to one case – for instance emotional involvement or media pressure, etc.

The combination of all these factors will shape the observation process and the resulting output. According to Dror et al. (2005), the stronger the bottom-up factors, the higher the impact of (subjective) top-down elements, which correspond to the observer effect. In other words, the more difficult the fingermark (bottom-up) individualisation, as for instance in the Mayfield case, the more weight other non-trace related factors will gain, as for instance context information available. Research conducted by Dror et al. (2005) showed that it is possible to study bottom-up and top-down factors by experimental tests to gain a better insight as to their perceived and real influence on fingermark individualisation. Experimentation is the usual approach for testing hypothesis in social sciences. The aim is to isolate one or several factors to be studied. They are tested in a framework that to some extent simulates the “natural environment” of the initial

research question. In order to be able to observe results directly related to the research question, the design of experimental tests tries to minimise the influence of any other unrelated variable. Tools used are repeating the measure taken enough times to obtain a statistically interesting value and confirm that the result is not random. A further strategy is to make sure that only the factor of interest varies among two conditions tested. That can be achieved by either testing the same person under the two conditions (Dror et al., 2006), or making two or more randomly selected groups (supposed to be "identical"), submitting each of them to one of the predefined conditions chosen (Enescu, 2008). The advantages and disadvantages of those methodologies are discussed in (methodological) introductory social sciences research books (Hy et al., 1983; Seale, 2004). More specific discussions are found in the abovementioned publications by Dror and others (Table 2), which consider the specificity of the present topic and the choice of methodology (See also 3.4).

Yet, cognitive biases affecting the way observations are made are often considered only as far as their negative influence is concerned. Thus, the impact in the fingermark individualisation process focuses on the possibility of wrong identification[107] and possibly, judicial errors.

However, not all errors do "produce" the worst case result[108], as they might, for instance, be discovered within the normal review process (if existing). Furthermore, it is also possible that a certain degree of bias facilitates comparison work in fingermark individualisation and is necessary. An example would be to choose from the list of 20 AFIS[109] hits one fingermark to start the comparison work (it is not always the first hit in the list that is the right source). Consequently, background information, viewed by some as one of the more perilous influences on forensic science work (Saks et al., 2003), might be a useful tool for more efficient work. Though one possibility would be to follow the ranking of the software, information on the crime type (e.g. burglary) and the presence of a known burglar in the region of interest to the investigation might justify starting with his comparison print, though only placed 17th in the AFIS ranking. Time and money could be saved in contrast to a context-exempt and purely fingerprint-based approach. Seen in this light, biases might be *necessary* to insure a better performance, one of the reasons why these cognitive shortcuts are part of the human observational and decisional framework. Experimental tests are therefore an essential tool for answering such questions.

---

[107] Note that in a real case it is not so easy to prove beyond "reasonable" doubt that a fingermark comparison was erroneous, as a supplementary, independent source of information will have to prove it (there is for instance still an ongoing discussion in the Shirley McKie case), as the real source of the fingermark is generally not known (see Mnookin, 2008, p.134, inf. 28 on this point).

[108] See Chapter 2 for more details.

[109] Automated Fingerprint Identification System.

3.3 Methodology and results

The methodology chosen is described in more detail in both papers presented next. It is based mostly on research protocols found for similar research, such as Dror et al. (see Table 2). They have been adapted to the needs of the present research. Concerning the two publications, one has been already published (Schiffer and Champod, 2007); the second will be submitted soon. Both papers can be found in appendices 9.1.2 and 9.1.3, respectively. Both research projects were financed by the Swiss National Science Foundation (SNSF).

The first paper, "***The potential (negative) influence of observational biases at the analysis stage of fingermark individualisation"*** (Schiffer and Champod, 2007; see also appendix 9.1.2), summarises the reasons why the study of observational biases is of interest in fingermark comparison, namely due to cases of erroneous identifications and questions of admissibility. It proposes a way of testing those effects through experimental research and applies it to one specific example. The research project includes, among other parts, two experimental tests presented here 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.

Factors potentially influencing the analysis of fingermarks, 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) were studied. For all tests students were given twelve fingermarks 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 (useful for comparison purposes) and identifiable. The aim was to see how the fingermarks 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 twelve fingermarks before and after having followed specific training in fingermark individualisation. The aim was to see how training/education impacts the analysis of fingermarks. For Test II, participants were given eleven fingermarks 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 an 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. This showed that the type of experimental research design chosen furnishes results, even if they seem intuitively obvious (effect of training). Furthermore, it was shown that for “normal” fingermarks under low-bias conditions the analysis stage of ACE-V was resistant against tested factors potentially inducing bias.

The second paper, ***"Is ACE-V vulnerable to contextual biases?"*** (See appendix 9.1.3), is presented next. Based on the same basic research interest and protocol as the first paper, this study focuses on the influence of contextual bias (availability of comparison print) and situational pressure (stress). It includes the three first stages of the ACE-V process in the test (Verification was not studied) in contrast to the first paper only interested in the analysis stage. The population used for these experiments were students enrolled in the forensic science program of the University of Lausanne (BSc and MSc level). Two factors were tested. First, it was tested whether (only) analysing one fingermark (Analysis stage only in a low/no bias context) would differ from reanalysing the same fingermark in a ACE process being exposed to considerable pressure created by an assessment situation and availability of the comparison print. A possible difference was expected to manifest itself by for example an increase of minutiae from the analysis only to the ACE situation.
Second, it was tested whether if under considerable pressure students were tempted to "complete" their findings of minutiae on the fingermark by looking for information on the comparison print (circular reasoning). To test this, one half of the student test population (24) were given the Madrid bombing fingermark with the comparison print of Daoud, the other half (24) with the comparison print of Mayfield. Again, they were expected to follow an ACE process and to pronounce themselves as to the result of their comparison. A possible bias was expected to result in a difference of minutiae observed, and in the value attributed to the result.
For the first test condition, results show that an important difference exists between analysing fingermarks as an exercise (normal context) and doing it during an assessment under pressure and doing comparison and evaluation as well. Indeed, the quantity of minutiae observed increases considerably when under pressure and having the comparison print available. For the second test condition, it is observed that the amount of minutiae does not vary significantly between the two groups having respectively a matching or a non-matching comparison print (Daoud or Mayfield). This can be interpreted as an absence of an tendency to complete the fingermarks with data observed on the comparison print. As to the opinions formed, one participant excluded wrongly the corresponding Daoud print, whereas five participants (out of 24) agreed, tentatively, to a "tends towards individualisation" with the Mayfield comparison print, though no identification was pronounced. As already mentioned, no verification stage was included in these experiments.

In summary, based on the abovementioned results of all tests, contextual biases seem to be observable in the Evaluation stage of the ACE-V process if situational pressure is high. The same external factor and the availability of the comparison fingermark do not influence the Analysis stage of the process if participants do document their findings in writing.

3.4 Comments on methodological and theoretical issues

The first question is whether the sample was adequate for the purpose of the test (see Wertheim et al., 2006 for a similar discussion). Some elements already presented in the research project will be briefly summarised. It is difficult to obtain a sample of correct size in the domain of forensic science, be it identification evidence or other areas. Practitioners are not very easily reached and available, nor very numerous (Dror and Rosenthal, 2008). Forensic science students are more available and more numerous. The main contrary argument is that they are not comparable to practitioners because of a difference in education and experience. Generally speaking, most research projects resorting to empirical tests in psychology do draw their samples from student populations (see for example Enescu, 2008) and agree that results are valid.

It is supposed here that practitioners perform better than students, an argument that might readily be accepted. This means that if students did their analysis, comparison, and evaluation correctly, there is no reason to believe that this would not be true for more experienced practitioners. On the other hand, results obtained tend to show that for the Madrid bombing fingermark professionals did not perform better than students. While this can be a single occurrence, it is suggested here that their performances can be compared.

The second question is whether the research design agrees with established methodology and has been executed correctly. The description of how the tests have been prepared and executed can be found in Chapter 3.3. They have been structured in accordance with existing experimental tests by other authors in similar areas. Due to the specificity of the domain, some adaptations have been made.

The size of the sample and the statistical values obtained limit the generalisability of the results. It should not be forgotten however, that the aim of the tests has been twofold. One of them was to examine the feasibility of experimental tests in fingermark individualisation. This goal has been met. The second was to obtain results which would allow a study of the hypothesis posed at the beginning, an aim that has also been reached.

The third question is whether the data obtained allows us to answer the questions that were of interest for the research. Several results have obtained and have been discussed in Chapter 3.3. In a more general perspective, results tend to show that the type and “strength” of factors inducing observational biases is an important element to consider. Thus, in a first row of tests, to give a low bias background did not affect the analysis of fingermarks while doing the same task in a neutral context but under situational pressure did influence the result. A further element to keep in mind is that not all stages of one process may have the same vulnerability to various possible factors. As the results show, analysis tends to be more “bias-resistant” than the more decisional evaluation stage.

Answers obtained are rather specific to one area - fingermark individualisation - and to the chosen factors studied. Results tend to agree with other existing research on similar fields or

similar factors (Dror et al., 2005; Dror and Charlton, 2006; Dror et al., 2006; Kerstholt et al., 2007). However, to decide to which degree the results can be generalised is not easy because a lot of lacunae persist in this area of research (See Chapter 5 on this). On the other hand, it is possible to reach practical recommendations and/or further specific and useful research directions that will hopefully help to answer those questions in more detail. Indeed, to minimise errors it is not necessary to have the correct and comprehensive answer to the problem, but interesting and feasible information as to how to ameliorate the situation. To know that the following of a strict protocol (ACE) led to encouraging results that can have a practical impact.

The fact that results obtained do not “expose” observational biases as much as could be expected after the literature review can either be explained by the research methodology chosen (discussed above) or that the specific area chosen, more focused on observation and less on decision, is less vulnerable to them. Indeed, stimuli inducing observational biases tend to be strongest (according to literature) in decision-making moments. It could also be that the actual influence of those effects has been overrated (for a discussion on this see 5.3).

4. Semi-structured interviews

In this chapter, semi-structured interviews will give an overview of the opinion of forensic science professionals on forensic science and judicial errors. Questions that were asked comprise, among others, their definition of judicial error, the most typical errors encountered, the measures in place in their laboratories or units to avoid them and their suggestions for minimising error in forensic science work. The chapter will be structured as follows. After a short introduction, the questions asked will be presented (4.1). Then, following a brief theoretical outline (4.2) on existing publications, forensic science institutions and methodological issues, the methodology used and the results obtained (4.3) will be presented. They are structured the same way as the questions asked. If necessary, a short commentary will be added to clarify some details. More general comments (4.4) on methodological and theoretical issues will close the chapter.

The literature review in Chapter 2 gave a very general overview on judicial errors, but also a quasi-exhaustive list of potential error sources in forensic science. Here the focus is on the opinion of professionals regarding these subjects. One of the observations made during the literature review was that the point of view of forensic scientists on these questions is lacking. The interviews try to fill this gap by directly collecting the opinion of professionals. To get this input it is important for several reasons. First, it allows for obtaining new data that might challenge and change the established picture on error sources and on promising remedies. Second, by comparing the opinions of individuals from different countries and backgrounds, it might be possible to differentiate two categories of error sources: those linked to a given situation in a given country (e.g. Swiss Criminal Justice System) from those being of a more general nature and therefore probably directly linked to forensic science. Third, it will also be possible to confront the error list as established by the literature review with a more experienced-based error list. This might help thinking about the visibility or chances of discovery of various error categories. Furthermore, after the specific focus on observational biases in fingermark individualisation in the previous chapter, here the question of errors in forensic science is reconsidered in a larger context. Thus both empirical approaches presented are complementary.

4.1 Research questions asked

The literature review (Chapter 2) helped establish several points of interest which are presented next. By using the adapted research tools, those questions were asked to the participants and their answers “re-codified” to fit these categories (4.3).

1. The status quo of awareness of forensic science as possible source leading to wrongful convictions in Switzerland and compare it to other countries.

2. An inventory of “narrow escapes”, where something went wrong but was “corrected” before having a negative impact or, because its unimportance was established before it could be used (case examples).
3. Reasons that led to the discovery of the near-mistake (internal review; police investigation; counter-expertises).
4. Mechanisms in place in order to avoid such problems (e.g. guidelines, peer review, etc.).
5. Possible explanations for the occurrence of such problems (error sources) and their relative weight at different stages[110].
6. The role of education, training, and experience to anticipate/avoid, recognise, and correct such problems.
7. What would be needed to solve these problems, if they occur.
8. Whether (forensic specimens) are kept beyond the moment a judicial decision becomes final.

4.2 Theoretical outline on organisations and research methods

As seen in the literature review in Chapter 2, the subject of mistakes in forensic science, or difficulties with scientific evidence, seems to be much more prevalent in Anglophone countries in comparison to, for example, continental Europe. As discussed previously, this has been laid down, among other things, to the legal systems whose functioning and expectations are quite different[111], but also to a difference in awareness. However, there is in a general way a lack of information on the opinion or point of view of forensic scientists as to the role of their work in judicial errors or of errors per se in forensic science (see 2.3.1).

This observation applies to Switzerland as well. Anecdotal reports retain some occurrences of judicial errors, without specifying exactly forensic science as a/the main source of error (Killias, 2008). Recent research on last instance appeals at cantonal level barely reveals any case where forensic science was present or debated (Killias et al., 2007). Crime scene work and related tasks (which, in Britain[112] are done by the SOCO (Scene of Crime Officers)) are performed by the technical service of each of the 26 Swiss cantonal police force units and/or laboratories. Some of them may comprise as few as two or three, others up to 90 members (SOCO, laboratory personnel, and administration comprised). General tasks, not needing expensive equipment or very specially trained personnel, are performed by each technical police unit. More specialised tasks are done, depending on the field, by the two Zurich police

[110] This last factor will subsequently be compared with the literature review error list and analysed in the discussion (4.5). A first presentation on the topic has already been made at the 6th ESC conference in Tübingen (Schiffer et al., 2006).
[111] Adversarial versus inquisitorial criminal justice system. See Kessler (2008) or Killias (2008) for a comparison on general judicial error elements and criminal justice systems.
[112] Interviewed forensic practitioners were from England. To agree with previously used language, the terms British and UK will be used, as it is supposed that the points discussed in this research do not vary considerably between England and the rest of the UK (see also Quirk, 2007).

forces (WD[113] city police and KTA cantonal police), the new centralised criminalistics centre in St. Gall, the legal medicine institutes, and/or the Institut de police scientifique, Lausanne, attached to the law faculty of the University of Lausanne (Rivola, 2004). Despite being a small country, each canton is allowed to organise its forensic work independently, though efforts of collaboration and/or centralisation persist. The Swiss federal police do not have any special forensic laboratories[114]. If in need, they "delegate" scientific work to one of the abovementioned laboratories or a police force.

Concerning the other countries surveyed in this research, both Sweden and Finland have a central laboratory, where most of the more forensic tasks are performed and where they keep their specialised staff and equipment. Each SOCO Unit is supposed to send all the traces they will not process themselves to the central laboratory. France does have a similar system, though they have two police forces and associated laboratories, the *gendarmerie* and the *police nationale*. The UK has evolved towards a system of private[115] forensic laboratories, one of them being the Forensic Science Service (FSS). Thus SOCOs do the crime scene work and the police forces choose the laboratory they want for processing their crime scene traces. Australia has a Federal police with a federal laboratory, the functioning being, in its broad outlines, comparable to several continental European countries, though some states have their own laboratories, such as NSW (New South Wales).

Discussions on the relative advantages and disadvantages of interviews, as chosen here, or other social science research tools, such as various kinds of surveys or the use of official statistics, can be found in introductory works to social sciences[116] (Hy et al., 1983; Seale, 2004) and criminology (Killias, 2001). They will not be repeated here. The selection of a given research tool will depend on the topic to be studied (e.g. sensitive or not), the type of research questions asked, the type of data expected (quantitative or qualitative), the size, and the accessibility of the population to be studied.

Interviews can take different forms, ranging from very structured to very free collections of narrations. The more structured, the easier a quantitative exploitation. On the other hand, the more qualitative, the easier to take into account the information received, which will not be "pre-classified" such as in the more structured forms. For the present research a median between the two was chosen, the semi-structured interviews. They do allow a certain comparison between the opinions of experienced forensic scientists while allowing to take into account original and new topics introduced by participants.

For the present research, the topic and the questions asked were judged to be sensitive, as it included questioning the quality of the work performed by participants. Furthermore, as no published empirical research collecting the opinion of forensic scientists on error (sources) has

---

[113] WD: Wissenschaftlicher Dienst; KTA: Kriminaltechnische Abteilung

[114] Only for explosives, which are a federal competence, do they have a dedicated laboratory WFD (Wissenschaftlicher Forschungsdienst), which for practical reasons is attached to the WD.

[115] Private sector classified but with the Home Office as a shareholder

[116] Including sociology and psychology

been found, an exploratory approach seemed justified, this being a usual way of proceeding in social science research. Finally, forensic scientists are a small and overworked population. By resorting to face to face interviews the chance of obtaining data seemed higher than, for instance, using a paper questionnaire.

4.3 Methodology and results

4.3.1 Methodology

Participants

Participants were a total of 11 forensic scientists, most of them in charge of a forensic laboratory or a forensic science unit, some of them retired. Two of them were approached in the developing stage of the interviews; the others were contacted later during the 4th EAFS (European Academy of Forensic Science) conference in Helsinki in 2006 (6 interviews), or shortly afterwards (3 interviews). Criteria for their selection were their position/function as a former/current forensic science "manager", an interest for questions linked to the quality of forensic science, their willingness to be interviewed by a PhD student and their faculty to express themselves fluently either in (Swiss)German, French or English.

For the Swiss part of the sample (4 participants), the selection tried to consider representatives of forensic units of different size, language, and geographical position (Rivola, 2004). For the international part of the sample, European countries with various forensic science systems such as the UK, France, Sweden, and Finland were considered. The Australian participant was included to have an outlook at another, more global level.

Participants were (former) heads of forensic science laboratories in France (1), Finland (1), and Sweden (1); (former) chief forensic scientists or responsible of Research and Development in Australia (1), France (1), and the United Kingdom (2); and heads of forensic science units in Switzerland (4).

Materials and design

A Dictaphone with audio microcassettes was used to tape the interviews. Each of them has been transcribed, analysed, compared, and evaluated. Questions were framed in a semi-structured way, as proposed by Blanchet and Gotman (2001) and Kaufmann (2001). They covered the fields of interest presented in section 4.1. All these topics were initiated by triggering questions in one of the three languages mentioned above. An example would be "if you hear 'miscarriage of justice', what comes spontaneously to your mind?" (See appendix 9.3 for a prepared "interview grid" in French).

Procedure

Participants were asked whether they would be available for one hour of interview. They were guaranteed anonymity. In 4 cases the interviews took place at the participants' office or selected

room. In 7 cases, advantage was taken of their presence for a conference (6 in Helsinki, and 1 in Switzerland), and a quiet area was chosen for the task.

Participants were (most often) asked to present themselves shortly, and then a general, introductory question was asked. In function of topics discussed by the participants further questions were asked, so as to cover all abovementioned topics. The background and the personal interests of the forensic scientists did influence the interviews. Thus, not all questions have been directly answered by all participants. For instance, forensic scientists less involved in crime scene work answered questions of error risk in this field more sketchily than those for whom it is daily work.

4.3.2 Results

In a first part of this chapter, the questions stated under 4.1 will be explored, considering the data extracted from all interviews. All opinions of all participants are represented, reproduced in a reorganised way to give a comprehensive view of the answers given. National or other specificities revealed will be mentioned and commented. Finally, additional results not directly related to general questions will be mentioned and their interest discussed[117].

4.3.2.1 The status quo of awareness of forensic science as possible source leading to wrongful convictions in Switzerland and compare it to other countries

Swiss interview partners (4) agree that judicial errors do happen. However, they tend to feel that they happen abroad, for instance in the USA, and less so in Switzerland. One interview partner mentioned the case of Bruno Zwahlen known as the "Mord in Kehrsatz"[118], and thus accepts the occurrence of judicial errors or rather mistakes in forensic science work in Switzerland. Another participant feels that the risk of judicial errors in Switzerland might be increased in serial crimes such as burglaries. He estimates that the risk of "adding" wrongly a non-related case to a series is higher than to err in a murder case, where special care is given at all stages. All Swiss forensic scientists agree, though, that the risk of a judicial error happening due to forensic science alone is very small, as its role in court is small compared to other evidence. They furthermore feel that there are several safeguards in place against judicial errors. One of them would be the tendency of the prosecutor only to accuse if he feels that there is a high probability of being able to convict the accused.

---

[117] Furthermore, it must be stressed that all elements presented are based on data extracted from the interviews and might not agree with opinions of the author as expressed in other parts of this research. The aim here is to give a neutral overview of opinions extracted from the interviews.

[118] In 1985 the wife of Bruno Zwahlen is found in the freezer of their house. Suspicion against Bruno Zwahlen quickly awakened, especially as he was having an affair at this time. He was judged and found guilty in 1987. His trial awoke a huge public and media interest. Journalists started a campaign claiming his innocence and discussing various errors. At a retrial in 1993 he was acquitted.

British interview partners were aware that judicial errors have happened in their country and they both felt that further improvements are necessary to avoid any further happening, both in the forensic science area and in general.
Continental European participants tend to mention the British Birmingham Six case and US cases as first examples of judicial errors. They have sometimes experienced cases, where things have not been done as good as possible. However, forensic science has seldom been the (only) factor leading to the final result. Only the French participants tend to mention the Outreau case first[119].

4.3.2.2 An inventory of "narrow escapes", where something went wrong but was "corrected" before having a negative impact or, because its unimportance was established before it could be used

Not all participants were asked about narrow escapes or answered this question directly. Some participants did answer it in a more abstract way; others discussed errors very openly. Again, differences in task/function, etc., influence the answers given.
Discovered and mentioned errors both in the experience of participants or observed in experiences of colleagues include organisational issues at the crime scene, problems in the laboratory, clerical errors, and report writing difficulties. Participants also mention that to organise and cover the work at a crime scene is not easy and gets more complicated when the amount of people involved increases. This includes also geographically separated crime scenes or serial crimes, which can be further complicated by different jurisdictions. Another difficulty can be contamination issues, especially with sensitive DNA profile methods, the labelling of specimens (wrongly assigning a specimen to a case), both at the crime scene and the laboratory. At this second stage errors can happen if equipment is not maintained correctly, if blanks are not analysed, if sampling errors occur, or if concentrations are calculated wrongly, so as to influence the result of an analysis. There might also be an underestimation of false positives or the limits of a technique. Clerical errors might happen as well. One example is loading up the wrong DNA profile to a database due to switching numbers. This can also occur at the report writing stage. The interpretation of the results is subjective and will/might vary between two colleagues. The writing down of the results can happen in a subjective way, if the language and working hypothesis are not objective enough. If nothing about the case is known, it is also possible that the correct and useful work has not been done nor the answer to the "real" questions of the police/investigators found.
More general errors at all stages can happen due to contamination or due to reliance on non-verified information. Thus, for instance, reference DNA samples have to be verified for their source. A further issue is the relevance of a trace (specimen). Not everything collected on a crime scene is necessarily relevant to the offence committed. Nor is there a guarantee that this specimen will answer the questions of interest in a case.

[119] Which according to the definitions used here qualifies as miscarriage of justice, but not wrongful conviction.

4.3.2.3 Reasons that led to the discovery of the near-mistake

Not all participants were asked this question specifically, and some of them answered them in a more abstract way[120]. One of the often mentioned points that helped discover near-mistakes were formal mechanisms of quality management or quality assurance, which will be discussed in more detail in the next section. Here the focus will be on other factors. Forensic practitioners mention that it is often the unexpected nature of the result - mostly after the laboratory stage - which triggers a retesting/reanalysing of a specimen. This can both happen after internal or external review.

Internal means, because the result surprises or because verification by a second person questions the result. External means, because the result/report raises questions by the investigative (police/prosecution) instances. A further possibility arises if a counter-expertise is asked. In Switzerland, the closeness of investigative forces (police) and forensic science units (also part of the police forces) allows also for a (mutual) checking of preliminary results. If they do diverge strongly, an unofficial re-examination might be executed in the laboratory. This seems less feasible in other countries, as forensic science and the police investigation tend to be much more separated. Indeed, very often crime scene and laboratory work are divided into two different units, which is not the case in Switzerland.

4.3.2.4 Mechanisms in place in order to avoid such problems

The range of structures in place varies considerably from one laboratory to another. When comparing Swiss forensic science units, their size determines what work is done by them and what is sent to a laboratory. The smaller the unit, the less forensic science areas are covered and the more the focus is on crime scene work and identification evidence. This sometimes goes hand in hand with a less “formal” QMS or QAS (Quality Management System; Quality Assurance System), which complies with internal, but not always with external standards. One example of internal QAS would be the four-eyes-principle in one of the services, which implies that each examination/result is verified (to a certain degree) by a second person. DNA is changing this perception, due to standardised externally imposed QMS measures in place in the DNA analysis laboratories (in Switzerland separated from police forces and attached to legal medicine institutes) somewhat pushing the need for more externally recognised mechanisms. One Swiss laboratory has implemented more formal measures, such as the introduction of a sophisticated LIMS (Laboratory Information Management System) and other QM measures.

In England and Wales, the question of quality has been discussed since the late 1980s and gained importance when the FSS became a “business-driven” institution which started charging for its work. After national guidelines had been followed, ISO 17025 was adopted for the laboratory work. According to the interview partner(s), this helps both to avoid errors by following established and trustworthy procedures and protocols, as well as to identify quality failures before they “leave” the system.

---

[120] Only participants that mentioned near-mistakes (without or with being asked) were asked this question.

Another participant felt that to do case audits to learn from one's own errors/experiences would be useful. Reviews are often limited due to cost. Quality, efficiency, and cost are, according to him, somehow interrelated. The Australian Federal Police (AFP) for instance, has installed quality assurance measures in their (analytical) laboratory according to ISO 17025, which implies written protocols, blind tests, etc.

Comments

Most laboratories will have internal controls, though not all take the structured form of QMS such as ISO 17025. One of the reasons is that this norm applies to laboratories only and therefore seems not to apply to most Swiss forensic science units, which are more focused on crime scene and identification evidence. Countries with a longer QMS tradition, such as the UK, have not accredited/certified their SOCOs, the equivalent of the Swiss forensic science units. Movements exist, for instance in Switzerland, to introduce ISO 17020 (inspection) for SOC work. Opinions on the usefulness/gain of this approach vary considerably. A second tendency is to guarantee the "quality" of the forensic examiners, as proposed by the ENFSI (European Network of Forensic Science Institutes) (Fereday and Kopp, 2002; Malkoc and Neuteboom, 2007). For further discussion on accreditation and certification see Stauffer and Schiffer (2007), Schiffer and Stauffer (in press), and Stauffer and Schiffer (in press).

4.3.2.5 Possible explanations for the occurrence of such problems (error sources) and their relative weight at different stages

According to participants, error sources can be described in several ways. One consists in analysing the chronological events from the crime scene, through the laboratory, to the report. At the crime scene, the risk is highest of not finding material of evidence interest, of losing it or of confusing it/mixing it up. In the laboratory stage, the correct use of a validated technique under controlled circumstances is crucial. If this is not the case, incomplete, contaminated results can occur. At the interpretational or report level, the work done has to be evaluated. Special care has to be given to language, to guarantee a transparent and objective report. Another way of proceeding is to separate errors into voluntary and non-voluntary mistakes. They can also be divided into technical and ethical[121] errors. The first includes the misapplication of techniques, while the second consists of an overstatement of the results or the abovementioned willing modification. The first are thought to be easier to solve than the second. More general difficulties are communication problems, especially the interfaces between the different stages and the various partners implicated in forensic science work or related/linked areas. This starts already with the arrival of the first police officers at the crime scene and their reporting to the forensic scientists/SOCOs. It continues with the cooperation between SOCOs, forensic scientists, investigators, etc. An additional risk is if police officers feel that they can, by

[121] Named this way by a participant

themselves, already proceed to first evidence protecting measures. Though well-intentioned, they may destroy more than protect.
Interview partners mention that further potential error sources include human weaknesses such as the need to establish a working hypothesis (already a selection), the positive/negative influence of experience, the tendency to confirm rather than infirm hypothesis, and the lack of data verification, especially if external pressure augments.

Of the three stages of crime scene, laboratory and report writing (collect traces, process traces, evaluate and present results), the most critical seems to be the first one. If work has to be done quickly, due to budgetary limitations, things can be overlooked or forgotten. Indeed, what is lost at the beginning can not be recovered later. This opinion is shared by all interviewed forensic scientists, and this commentary has been made for all types of forensic services, the crime scene work being delegated to some other unit or not. However, those working only/mainly in the laboratory do stress this point more than the others. (It might be that it is very difficult for them to control the quality of this work).
The second critical place is held by the report-writing stage, usually linked to the "metamorphosis" of analytical results or observations into a formal statement that could be used as evidence in court. The subjective and sensitive part of this task has been stressed by most participants. It is also felt that the risk consists not in writing down false data – though clerical errors might happen – but rather in over- or underemphasising the evidential value attached to the set of data. It has also been mentioned that a lack of information on the background of the case will not allow the selection of the most efficient work to do nor evaluation according to useful hypotheses, the prosecution and defence position for instance. This means that relevant questions might not be asked, nor answered. It has also been stressed that to exclude case information from forensic science work is counterproductive and unrealistic.
Most confidence is expressed by participants in the laboratory process of the work. Arguments given for this position are the ease with which standardised protocols can be/have been established and work can be controlled. Some participants feel that most of the efforts for ameliorations until now have gone into this stage, which explains to some degree their confidence[122]. It has also been mentioned that areas relying more on subjective judgment or experience are more vulnerable to over/underestimating its value.
In summary, the less routinely an activity can be described and performed, and the more subjective choices are included, the smaller seems to be the confidence in the work done at a specific stage.

[122] But perhaps it is simply the work done by most participants, so they feel more confident, as they know what is done and what the risks exactly are.

4.3.2.6 The role of education, training, and experience to anticipate/avoid, recognise, and correct such problems

Several aspects have been mentioned by participants. One of them is the need to keep a general perspective of forensic science, even while being a specialist for a given domain. This general information should be shared, at least to a certain degree, by all participants in a criminal investigation, so as to understand the possibilities and limitations of forensic science. This includes the first police officers arriving at the crime scene, but also the magistrate leading the investigation[123]. They should however not confuse information received on the subject with education, which is their own faculty to act correctly. Several interview partners have stressed the importance of experience as well. As an interviewee put it: “Experience is a succession of previous errors, and because one learns mainly due to errors…one has perhaps more “alarms”[124]. Other inputs include the thinking about the basic selection and training of forensic scientists, the importance of training in the context of QAS (one of its backbones), and the need of discussion and education on ethical aspects of forensic science. Some participants have also insisted on the difficulty of teaching core values of forensic science, which are not of a technical nature, as well as the complex task of training young professionals in work ethics and a proper scientific attitude, such as admitting errors. It is felt that only the immersion in a real-life setting fulfilling these conditions can achieve this task. On the other hand, one participant insisted on the need to teach integrity already at elementary school, as part of general education.

In summary, education is seen as crucial to maintain and augment the quality of forensic science by reaching both forensic scientists and related people. Experience is seen as a further important element for doing good work. And, forensic science is not only about technique, but also other aspects, such as core forensic principles and ethical questions.

4.3.2.7 What would be needed to solve these problems, if they occur

Participants were invited to present their ideal solution without considering real-life constraints. Elements that were mentioned repeatedly were having a well-functioning QMS/QAS with the best-trained and experienced people, as well as the solution of communication problems. In France, the introduction of a “criminalistic coordinator” is studied[125]. British interview partners stressed the need for an “integrated model” both within the “large” forensic science community, but also within the criminal justice system. According to them this means that, for example, the toxicologist should be able to collaborate directly with a forensic scientist if needed. It also includes the immediate collaboration of forensic scientists with the police investigation, but also at the presentation stage of evidence. A further point is access to ideal “logistic situations”[126]. In addition, in Switzerland (Zurich) a seminar on scientific evidence has been held in June 2007 in

[123] in a continental law system

[124] “L’expérience est une succession d’erreurs préalables, et puisqu’on apprend principalement sur les erreurs … on a peut-être plus d’alarmes.»

[125] His task would be to keep track of all forensic science work and coordinate all participants involved in this effort.

[126] Not having the necessary material/techniques at hand might lead to a loss of precious information.

order to discuss the needs and questions of all implicated institutions. It brought together prosecutors, judges, police officers, forensic scientists, forensic pathologists, etc. Participants raised the importance of discussing ethical issues together with the need for internal regulation. “(A) community of professionals, a community where truthfulness and honesty are appreciated more strongly than now. And where misconducts done knowingly would be punished more strongly than they are now”. It has also been mentioned that for high-profile cases the resources are found for doing the work correctly. However, it is as important to put resources in everyday crimes. Generally, forensic science could be used more often than it is, if its possibilities were better known and its quality recognised. This would also help attract good (and qualified) people to this kind of work. Finally, it is stressed that the larger and more varied the different evidence is, both scientific or not, the smaller the risk of judicial errors will be.

4.3.2.8 Whether forensic specimen are kept beyond the moment a judicial decision becomes final

According to participants, in general, forensic traces are kept either during a fixed time, for instance ten years or until all appeal/revision possibilities have been exhausted. How long they are kept will also depend on their nature, storage possibilities, and the type of offence. Chemical and physical limitations have to be considered. On the other hand, biological traces (e.g. for DNA profile specimen) tend to be stored apart under adequate conditions.

Items of “historical” cases may well be kept beyond official or legal deadlines. However, there have been some cold cases were traces were not available any more and where this has been regretted. Case files and notes are generally kept longer than specimen.

In Britain, the time during which items/specimen are kept will depend on the type of court the case went to, how long the appeal period is, what type of legal rules there are and what agreement the police and the forensic science service have found. Generally, the more serious the case is, the longer evidence will be kept. For murder cases that might mean 20 years or more, depending on the aforementioned factors.

Some participants agree that if specimens are kept it is easier to reopen and review a case than if they have already been destroyed. Opinions are less in agreement as to whether this factor explains the differences in the amount of judicial errors observed in continental Europe in comparison to English-speaking countries. Generally, it is difficult to evaluate the influence of one isolated factor, as many other aspects have their impact as well.

4.3.2.9 Additional points of interest revealed in the interviews

Among other questions participants were asked was their definition of judicial errors[127]. Most participants agree that, both wrongful convictions as well as guilty persons going free constitute a malfunctioning of the criminal justice system. Some added the non-discovery that a crime had occurred in the first place as a third phenomenon. One participant felt that to agree that a judicial error had taken place, it is necessary to show convincingly and comprehensively that the suspect/convict is not the author of the offence. Another participant has also specified that he does not feel a judicial error took place if a new method allows now for more "detailed" and interesting results than an older technique earlier on. A rather broad definition of judicial error was given as "the lack of dialogue between the different participants in the legal system and the selection of a wrong direction in the lead of the investigation, which can lead to error, sometimes by not seeing the most evident items."

Several explanations were given for the difference in the amount of judicial errors recorded in English-speaking compared to continental European countries. Some participants feel that statistical considerations (population size, crime rate, high-profile cases, etc.) could be part of the explanation. Thus, the high amount of "domestic murder cases" in Switzerland, where the author is most times easily determined, seems to create less vulnerability to errors than more "anonymous" murders. Another aspect could be the (social) attitude towards the criminal justice system. If a high level of confidence exists, there is a smaller risk of questioning legal decisions. In contrast, if the criminal justice system is under severe scrutiny, more attention is brought on (high-profile) cases. That might especially be true for the USA, where the suing culture opens the door for a critical attitude towards justice, according to the opinion of some participants.

The type of legal culture can also create a more critical attitude – as in the adversarial system – which might facilitate the discussion of error as such, as well as the discovery of errors, by challenging judicial decisions. It is also felt that the possibility of counter-expertises and the quality of experts are not negligible factors. Finally, the USA and British criminal justice systems are often perceived as being more "polarised" than continental Europe criminal justice systems, thus perhaps leading sometimes to considerable pressure on forensic science work and its presentation in court. The presence of a jury has also been perceived as a disadvantage by a participant. He also feels that to link professional success to the amount of people convicted/cases solved is dangerous as well.

[127] Participants were asked about miscarriages of justice, but to keep up general terminology used in this research the term judicial error will be used in the following. At the moment the interviews took place the distinction now made in writing was not yet so clearly defined. Judicial errors have been traduced as "Justizirrtümer" in German and "erreurs judiciaires" in French.

4.5 Comments

Methodological issues

The aim of the research is to learn more about several sensitive topics such as error rates in forensic science work. As not much empirical research exists, as presented in Chapter 2, an exploratory approach was chosen. Several possibilities exist for this kind of research. Among them, criminological studies are often based on interviews, questionnaires or direct observation if possible. Given the subject, the last is not possible as the chances of directly observing errors are not very high. Questionnaires have the advantage of allowing us to collect data on a sample of considerable size, as for instance Ramsey (2003) did when questioning Ohio criminal justice professionals (law enforcement, prosecutors, defence attorneys, and judges) as to their estimation of judicial error rate. On the other hand, response rates tend to be a limiting factor. Poveda (2001), discussing the respective merits of official statistics and a self-reporting survey on USA prisoners to estimate a judicial error rate, covers some aspects of this subject[128].

In sociological or criminological research, interviews are often used for exploratory studies and data gathering on sensitive subjects – such as in the present case. In its exploratory function it is possible to approach a new subject in a comprehensive way so as to establish tentative interesting topics, which might not have been covered by existing research or the imagination of the researcher. Thus, for instance, one participant proposed that a difference might exist between high-profile cases of judicial errors and wrongly attributing an additional (wrong) case to the serial offences of a serial burglar, both on the reasons leading to the judicial error, but also on the consequences for the convicted person. This aspect has been found neither in published literature nor has it occurred to the author. In its function as a tool to explore sensitive subjects, interviews have been used successfully in various areas, for instance victimisation surveys (Killias, 2001, Chapter 2.3.3.2).

According to the results obtained, it seems that though 1-2 participants might not have disclosed in every detail their perception on all subjects, it was judged that others were very open in their approach. The context of the research, for instance having been made from an academic institution in forensic science was certainly a supportive factor. The semi-directive structure of the interviews and the sample size limit the generalisability of the results. On the other hand, to augment the sample size for the qualitative approach chosen does not forcibly add much new information, but tends rather to confirm already gathered data, as had been observed with the last interviews done.

The limited external validity of the results can be, to a certain degree, counterbalanced by the sample selection. It is argued here that the international, comparative view does add weight to the results. Both the Swiss and international participants have, however, not been randomly selected, as the aim was to obtain results for a sensitive topic. Only participants who were

[128] With the self-reporting survey a wrong-imprisonment rate of roughly 15% is obtained. Several explanations for this high result compared to other estimates which range around 0.5-1% are given, though the more obvious seems lacking, namely the rather widespread difficulty to obtain valid results when questioning adults, as has also been observed for self-delinquency reports.

thought to agree to discuss the topic were selected. The results therefore represent the view of forensic scientists who agreed to discuss the topic. Language was also a limiting factor as only those "managed" by the interviewer and interviewed were considered. Furthermore, the functions/positions of the participants, though similar, can not forcibly be compared directly. Considering the results (and the scope of the research), however, which tend to show a high degree of conformity on the answers given by participants on most topics, the small sample size seems not to be a very important limitation.

The quality of the results has been judged high and the information gathered very rich and enlightening. Expectations as to potential results have been fulfilled. Language problems during the interviews were limited. As to translation problems, as they have been done all by the same person, the same interpretation will be present for all work done. Sometimes the lack of knowledge on specific legal structures or police/forensic science units/laboratories needed additional explanations during the interview. Generally, it was possible to answer all questions, though one of them had to be modified following the first two test interviews (evaluative versus investigative stage), as it was not possible to answer it[129]. It has been replaced by the question of the reliability of the different stages. In addition to the questions asked, several further results were obtained, that have been explored as well (4.4.9). It was therefore not only possible to get the "planned" information, but also to obtain extra data.

Theoretical issues

For the first time the role of forensic science in judicial errors has been studied qualitatively, by interviewing senior forensic scientists in a semi-structured way. Questions as to their definition of judicial errors, the risky spots, the potential solutions, and other more detailed points have been asked. How these results fit in with existing research and what their interest is will be discussed directly as a paper (see appendix 9.1.4). Here, a short summary will be given.

The paper, forensic science and judicial error: comparing the point of view resulting from a literature review with the opinion of interviewed forensic science practitioners as to error sources and remedies, uses two questions – error sources and solutions proposed – to exemplify and analyse differences found between the information resulting from the literature review and the interviewed forensic science professionals.

As far as it could be ascertained through the literature review, it is the first time forensic science professionals are directly asked about the possible influence and impact of their field on judicial errors. Eleven participants of 6 European countries and Australia were questioned using a semi-structured interview grid including, among others, questions on error sources and possible remedies. While it was observed that through different jurisdictions and countries the interview

[129] As forensic science is mostly used in its evaluative and not investigative function, most participants could not answer the question. Furthermore, investigative aspects concern the police investigation and forensic scientists do not necessarily get a feedback on this and therefore do not know the results of the investigative input of their work nor any possible errors attached to it.

partners seemed all to agree on both these points, their view is in contradiction with the information that has been extracted in the literature review. For example, in literature, the laboratory is seen as the most critical stage, followed by the evaluation of results. The crime scene is rarely mentioned. In contrast, practitioners feel that the higher error risk is on the crime scene, followed by the evaluation of the results and the laboratory. They justify their view explaining that most quality management efforts have been made in this last area, whereas crime scene work is much more difficult to standardise and control. A similar contradiction exists when studying the results proposed. In literature, the focus is on proposing further quality management measures and the independence of crime scene laboratories from law enforcement units. Sometimes increasing competition between laboratories is seen as a necessary incentive to increase their quality. Practitioners, on the other hand, say that one of the main problems is the separation and lack of collaboration of the different actors in the criminal justice system and they suggest that an integrated model is necessary to minimise those interaction problems. They further propose that efforts should go on improving the quality of the crime scene work rather than the laboratories. In addition, they stress the importance of adequate training and education of forensic science practitioners.

Reasons for these differences found between literature and interviewed practitioners are discussed. One of them is that literature tends to be USA focused, whereas the interview partners were mainly working in the European context. A second explanation centres on the need of practical knowledge of day to day working experience to be able to make useful recommendations for change. Given the important differences of opinion observed it seems imperative to include the professionals' point of view to be sure that proposals will indeed bring the necessary change advocated and help minimise judicial error.

Part III – Connecting and discussing previous results

Part II gave an overview of the three main pillars of the research – literature review, empirical tests with fingermarks, and semi-directive interviews with forensic science professionals. They have been commented upon separately and discussed as to their respective individual interest. Here, in Part III, the aim will be to put the different (empirical) threads together to evaluate their combined interest for studying the relationship between forensic science and judicial error.

"I don't know what to think. The fact is, we've got dashed little material for thinking with. It's probably too early for thinking."
(Sayers, 1991, p. 227)

5. Discussion

In this chapter the results obtained in Part II are discussed by exploring three questions. First, I examine whether the initial working hypothesis could be tested or not (5.1). Second, I study how the results presented here fit in with existing research (5.2). The focus is on the suggestion that existing explanatory models have their limitations and an alternative model is proposed. Third, the lessons learned and an outlook are given (5.3). Methodological comments made in the previous chapters are not discussed again.

5.1 Is forensic science a source of judicial error?

"It is a capital mistake to theorize before you have all the evidence. It biases the judgment." (Conan Doyle, 1981, p. 27[130]*)*

As illustrated in the literature review of Chapter 2, forensic science is seen as one possible contributor to judicial error. Academics and forensic science professionals disagree on what exactly is leading to these errors or allows minimising their occurrence (Chapter 4). Furthermore, tests in the field of fingermark individualisation (Chapter 3) allowed enunciating hypotheses on specific error sources such as observational biases in the evaluation stage of the ACE-V process. In short, to infer that forensic science is a source of judicial error (see Chapter 1.3) is not as simple as is often assumed in scholarly publications (see Chapter 2.3.1). Why this is so is developed in this chapter, using the results obtained respectively in the literature review, the tests on observational biases in fingermark individualisation, and the interviews with forensic science professionals. Key words of the initial research hypothesis such as incidence, impact, source, remedy, and practical recommendations are used to analyse, compare, and evaluate the three tools set up to approach the question of the relationship between forensic science and judicial error.

Note on an aside, that the quality of forensic science (evidence) has been debated, classifying it in categories such as correct, wrong, manipulated, exaggerated, etc. (Cooley, 2004). However, in the present context the factor that seems important is that forensic science is present[131], that it is used in court, and that it has an influence on the verdict[132]. Ideally, to consider a possible link to judicial error it would be preferable to have only forensic science and no other factors to study causality. But taking real case examples this selective approach does not seem

[130] A Study in Scarlet
[131] It could also be interesting to debate whether the absence of forensic science when it could or should be present might also be seen as a potential source of judicial error.
[132] This excludes a lot of other possible options that will not be discussed here.

feasible[133]. Thus it is argued here that, even when trying to study possible risk factors quantitatively, as proposed by Harmon (2001), it is rather difficult to establish direct causality between forensic science and judicial error[134].

As to judicial errors, how this concept is defined and differentiated from wrongful convictions, miscarriages of justice, and other terms will also have an influence as to which cases will be considered to study the abovementioned causality. Problems of definition have been discussed in more detail in Chapter 1.2 and need to be at the forefront of any study in this area.

A further differentiation that has been proposed is to distinguish between forensic science "as such" and the "use" that is made. Technically speaking, whether a DNA profile is analysed to find potential suspects through a database (investigative use), to help to prove a connection between a suspect and a crime scene (evaluative use), or to show a non-match between a convict and an evidentiary specimen (exonerative use) will not change (much[135]) the analysis made nor the result obtained. It is therefore argued here that the use made, investigative, evaluative or exonerative, can be - for the purpose of the discussion at least - separated from forensic science "as such"[136]. This seems an interesting approach because it allows the study of what types of errors can or do happen in forensic science ("as such"), independently of possible consequences of these errors[137]. By "cutting into half" the problem, it might be easier to find partial answers that will then help to build a more general model of causality. This approach has been used for studying experimentally the effect of observational biases on fingermark individualisation (Chapter 3) by decomposing the "problem" into smaller bits, by for instance considering some stages of the ACE-V process separately and/or studying individually one or several factors potentially influencing the process.

The same approach has been used to study the more general subject of judicial error and forensic science, by examining smaller bits of the topics separately, which is reflected in the three chapters of Part II. In the following, the results of the literature review, the experimental tests, and the interviews are discussed together after shortly summarising their contribution to this research.

The literature review allowed the inventory of publications on the subject and elicited some shortcomings. One observation was, for instance, that a considerable part of the publications are by authors from the USA who draw on their experience, which might not be transferred to other (legal/social) contexts. Giving an overview of the subject also helped to identify interesting

---

[133] Even though one of the specificities of judicial errors is that the amount and variety of evidence present is smaller than in comparable non judicial error cases (Edmond, 2002).

[134] What possible solutions to this problem could be is discussed towards the end of the chapter and in the next chapter.

[135] One could imagine that for other types of items of physical evidence a differentiation of the type of test indicative or confirmative could be made in function of the result needed (that might consider time and cost-efficiency). Indicative versus confirmative have been discussed in chapter 2.3.2.3.

[136] As mentioned in Chapter 1.1 forensic science "as such" encloses the gathering and processing of items and the "use made" its role as evidentiary proof in court.

[137] And to minimise errors in forensic science is a goal in itself, independent of its possible consequences such as judicial errors.

research questions and to see what needs to be tested empirically, as for example the observer effects studied in Chapter 3.

For the experimental tests, a first overview allowed us to see if and what publications exist on non-voluntary errors in forensic science (Chapter 2.3.2). As it is impossible to study either all fields or all topics, a choice was made. Fingermark individualisation was selected as a field with long tradition of legal acceptance and as an example for identification evidence[138]. Experimental tests were made that helped to understand how observational biases can affect the ACE-V process.

The aim of the interviews was to collect data from forensic science practitioners as to their awareness/views of the interaction between forensic science and judicial errors and ask further questions to help better understand the topic. One point stressed was for instance the importance of a good communication culture between all stages and actors of the criminal justice system to minimise information loss.

Considering these three approaches for the study of the ***incidence*** of judicial errors of forensic science allows the following observations. The literature contains limited information on this aspect (Harmon, 2001; Naughton, 2003; Risinger, 2007, Poveda, 2001), as well as for forensic science errors as such (Haber and Haber, 2004; Haber and Haber, 2008a). One of the reasons is the lack of experimental research[139]. Observer bias in forensic science, for instance, is difficult to discern as awareness is missing. Its non-voluntary nature makes it even harder to realise that something did not work out the way it should and to find out what exactly went wrong[140]. To establish a judicial error rate, however, though also fraught with methodological difficulties, seems easier to achieve. Interviews showed that participants are aware of the subject. Those having been exposed more to the topic, for instance because of the occurrence of judicial errors in their country (e.g. the UK), tended to show a greater alertness than those less exposed. Doing more empirical research is the best way to clarify the question of incidence of errors, both in forensic science and in judicial proceedings. This is the only way to stop reasoning based on anecdotic single cases and instead discuss systematically acquired data. This data could include error rates derived from structured quality management reviews of cases or any other questionnaire/survey specifically designed to study error rates in any forensic field or to evaluate the performance of forensic scientists.

---

[138] Other reasons for this choice have been discussed in Chapter 3, such as its reliance on individual observation and the fact that some empirical research has already been done in the field of fingermark comparison. The importance of observational effects/biases as source of non-voluntary error has been stressed by Risinger et al. (2002), who discuss this subject at length (for further details see Chapters 2.3.4, 3.2.2, and 3.2.3). The same authors say that the more the expertise relies on (subjective) human observation and judgment, the higher the risk of those effects/biases having an influence.

[139] Furthermore, cases where such types of errors have been acknowledged are rather rare. The Mayfield case (OIG, 2006) is one example. The limitations of case studies for building a structured data set have been described earlier in this research.

[140] For example, result different than expected due to bias.

The literature review showed that it was difficult to measure the ***impact*** of forensic science errors on judicial error, as existing case studies have their limitations[141] and only allow seeing part of the problem (Leo, 2005). One example is that exonerations can depend on the availability of evidence to be retested in contrast to evidence that cannot be retested (see Chapter 4.3.2.8). The experimental tests illustrated that a suggested important impact, as initially stipulated by Risinger et al. (2002), and somehow verified by the Mayfield case, may be less important when studied empirically. In a similar way, participants to the interviews readily admit the incidence of forensic errors, but are more careful when considering its impact on judicial error. Indeed, they stipulated that errors committed may not have (negative) consequences, as it depends on the use and weight attributed to forensic science at trial and subsequent proceedings. The impact of errors is estimated as being rather low by forensic practitioners[142]. Seeing the contradictory views, it is important to include several points of view by any research in such a complex field. When it comes to the matter of incidence and impact it is fundamental to know how many errors happen, where, and what effect they can have. Then only it becomes possible to know where to act.

The literature review helped also establish a list of ***error sources***, which is somewhat in contradiction with the results of the experimental tests and of the interviews. Experimental tests performed[143], for instance, (Chapter 3) show more nuanced results than what has been feared by Risinger et al. (2002). Research done illustrates that in the ACE-V[144] process, not all stages show the same vulnerability, the evaluation (E) being the most sensitive, probably because it is also most open to external influence and of a more subjective nature than the other stages. The results allow us also to pinpoint specific error sources. By adding to this observation the knowledge on the underlying cognitive mechanisms, it becomes clear what factors exactly impact on the fingermark comparison process. In a nutshell, a lack of data regarding sources of error has been observed across all forensic fields combined with an alarming corpus of papers claiming large perceived risks such as contextual bias without empirical demonstration. The only valid proposition to be able to start discussing real risks is to launch a programme of empirical research focused on the forensic science fields of interest.

As for the ***remedies and practical recommendations*** proposed, once more, they differ to a certain degree between the literature review and the other two approaches. As long as there is no agreement as to what causes the problem it is difficult to imagine how to solve it. One of the things to do is to continue research on topics of interest to find out more about error sources. Empirically obtained results in contrast to single case reviews should lead to more efficient solutions. Furthermore, it is important to distinguish between perceived and real risks. This

[141] See the differences in the evaluation of the same set of data first presented by Saks and Koehler (2005), and critically commented by Collins and Jarvis (2008).
[142] See also Collins and Jarvis (2009).
[143] And publications by other authors (see Chapter 3).
[144] The Verification process has not been tested here.

helps to establish factors that may effectively play a role and thus make specific recommendations.
A practical example in this research has been experienced in the field of fingermark comparison. From results obtained in controlled experiments it transpires that being forced to document and justify, each step taken is a safeguard against biases already during the observation and decision process. The approach adopted here tries to avoid the case-specific, fit for purpose solutions that have been systematically proposed following famous cases of misattribution. This contrasts with error avoidance propositions advanced in the literature for example in one specific case such as Mayfield (OIG, 2006). Case specific recommendations can rarely be adapted to make real fundamental changes. A next step will be to try to build (general) models of causality as discussed in the next chapter (5.2), and thus to evaluate the research hypothesis on causality in a more transparent way than published so far. Furthermore, the subject should be studied from a forensic science perspective, since this is lacking until now.

When considering the overall picture, it appears that the initial assertion of a theoretically and empirically established link between forensic science and judicial error cannot be confirmed, the principal reason being the highlighted lack of information, but also the lack of a more general causality model (Leo, 2005). Some other inferences can be drawn from the results. They are presented next.

The results of the experimental tests have revealed the following:

- First, they showed that it is feasible to conduct experimental research in fingermark individualisation to answer questions regarding the influence of observer effects and biases. This observation can probably also be transposed to other areas, or directly on the more general level of forensic science and judicial error.
- Second, it is conceivable that, adding future research, the process will be better understood so as to extract practical recommendations on the following issues: how to minimise bias by adapted procedures/methodologies; how to select adequate future practitioners; and how to adapt their training to make it more efficient.
- Third, with the contribution made here it is not only possible to observe cognitive biases – as done by Dror et al. 2006 (see Chapter 3.2.3) as well – but also to stress the importance of factors that seem to diminish the impact of those influences. “Sequential unmasking” (See Chapter 3.4) and written documentation are specific examples. This means that the interest of applying ACE-V can be tested and confirmed, and thus further discussion on the subject would be based on structured data and not on (personal) opinion.

Furthermore, the interviews with practitioners have established discrepancies with the existing view on the subject concerning forensic science and judicial error. It is rather surprising, that

such an (empirical) approach has not been chosen before. While it is easy to understand that to propose a general model (hypothesis) - explaining judicial errors and the possible contribution of forensic science - is a difficult task (see Chapter 5.2 on that), it seems less understandable that few (or no) efforts have been made to collect the point of view of those directly held responsible for producing errors, namely forensic scientists (see also Leo (2005) on this point). One of the reasons might be that when starting to look into details for one error source, for instance forensic science, it becomes quickly clear that simple solutions will not solve the issue, since the problem has many layers and facets. Another possibility is that it was not feasible to gain access to this specific population, which in its turn may raise other questions. Alhough case reviews (Bromwich, 2007; Rawley and Caddy, 2007) have been done with the help of forensic scientists, they are nonetheless specific to one context. Limitations of that have already been discussed. Here, the international pool of experts discussing the subject more generally, gives a different dimension to the results, since similar answers were given across different jurisdictions. It is suggested here, that this is a better justification for the results found and that this more general view allows us to focus on other types of errors less represented in the cases inventoried in the literature review.

This might be one of the reasons why the human factor seems a crucial element for the interviewed forensic scientists. External review may establish that the assistant scientific officer should have reported her difficulty with item APR 60 in the Damilola Taylor case (Rawley and Caddy, 2007) to the superior Reporting Officer, and that internal manuals should be amended to include recommendations of this type. However, to understand why this specific "error" happened may need inside knowledge of what were the "human" challenges of the work. Those can seldom be regulated or eliminated by quality management measures or directly observed by the usual review mechanisms. To seek the experience, knowledge, and propositions from practitioners seems also important because ultimately they will be the ones to implement and live with the changes. To impose modifications on somebody's work without consultation will meet probably with more resistance and less efficiency than if change is initiated by the practitioners themselves, because they have understood the reason for doing so and were able to propose how to introduce change. To argue that forensic scientists are not open to change, without having consulted them, could be understood as representing a preconceived view by those advancing this opinion[145].

In conclusion, the three tools used for evaluating the initial hypothesis (Is forensic science a source of judicial error?) are complementary and offer helpful insight in different fields and at different levels. None by itself can cover the subject, nor do they do together, because the topic is vast. But by combining their individual contribution, it is possible to gain a better insight than would have been possible by just considering these three research parts as separate contributions to the subject or by developing only one of them in more detail. Furthermore, a

[145] It is also important to keep in mind that "forensic scientists" is a large pool including many different people with different opinions and where "internal" disagreements exist as well.

certain cross-validation of the results obtained seemed possible, because the data input comes from different sources, though it might be argued that they might not be completely independent.
All three methodologies used are based on social science research tools. As presented in the comments for each of them, they proved to be helpful to answer the questions asked, that is to test the initial hypothesis of causality between forensic science and judicial error. It seems therefore justified to infer that criminology not only offers the necessary research tools for the task, but also is a relevant framework for discussing the topic. This was already proposed in the introduction of the thesis (Chapter 1).

5.2 How do the main results fit in with existing research?

"'What can be the matter then? Can you suggest no explanation?'
'I have devised seven separate explanations, each of which would cover the facts as far as we know them. But which of these is correct can only be determined by the fresh information which we shall no doubt find waiting for us.'"
(Conan Doyle, 1981, p. 323[146])

In the previous chapter, findings obtained were discussed in view of testing the initial research hypothesis on the possible relationship between forensic science and judicial error. Here, the results are considered in the more general framework of existing publications on the topic, especially those proposing explanatory models for judicial error (and forensic science). Their limitations and advantages are presented next before proposing some other ways of building additional, possibly more adapted models of causality.

5.2.1 Introduction

Model: "A simplified or idealized description or conception of a particular system, situation, or process, often in mathematical terms, that is put forward as a basis for theoretical or empirical understanding, or for calculations, predictions, etc.; a conceptual or mental representation of something."
(Oxford English Dictionary, last visited 2008/07/10[147])

To explain the occurrence of observations, models are constructed. They help to represent information in a more organised way. Sometimes different models can be used to explain the

[146] The Adventure of the Copper Beeches

[147] http://dictionary.oed.com/cgi/entry/00313038?query_type=word&queryword=model&first=1&max_to_show=10&sort_type=alpha&result_place=1&search_id=dove-eggQD8-1329&hilite=00313038

same phenomenon, as they may be based on different levels of detail, schools of thought, theories or utilities. Examples in criminology are the different theories explaining criminal behaviour based on individual, societal, and/or biological influences (Jones, 2006). This does not mean that one of them is more correct than others; it might just be that their function or goal varies. Sometimes it is not possible to test them comprehensively, thereby several propositions may coexist.

In another area, physics for instance, several models exist to describe forces acting upon bodies. They include classical mechanics, relativity, and quantum mechanics. One of the reasons why classical mechanics (Newtonian) are still used though "newer and better" models exist, is because it describes phenomena in everyday life quite correctly and in a simpler way than relativity or quantum mechanics. Approximation[148] is accepted for the sake of simplicity of application.

Models concerned with judicial error and forensic science should help evaluate incidence, impact, sources, remedies, and allow practical recommendations. According to Leo (2005), Nobles and Schiff (1995), and Savage (2007), as well as the literature review in Chapter 2, a certain "lack of modelling" can be observed. Reasons for that are, as argued by Leo (2005) insufficiencies in the focus of (empirical) research such as was illustrated in the previous chapter (5.1), with differences observed between empirically and non-empirically obtained propositions for resolution.

Second, there are practical difficulties encountered when building a hypothesis of causality between forensic science and judicial error. Some of them will be illustrated next to help understand the limitations of existing models, but also the possibilities of building more appropriate ones.

To begin with one has to consider whether forensic science is present or not in a given case. It could be argued that the absence of forensic science, when it could or should be available as evidence in court can also produce a judicial error. However, the focus here will be on situations where forensic science is present. In this case two main situations can occur. Either forensic science results are correct or are wrong[149]. Then, a further element is crucial. Forensic science is used or not. In the second case the situation is the same as if forensic science was not present. This means that initial forensic science error does not automatically lead to judicial error. Furthermore, several factors can or could filter it out with, for instance, quality controls.

First, the forensic science item (evidence) under discussion can be faulty. However, this can be discovered before or while the evidence is presented in court. In this case the element will not be used further. An example would be the discovery of missing results on a test routinely performed for this kind of specimen.

---

[148] This means that even in fields that are perceived as being very scientific a subjective and imperfect model is accepted. Therefore why should this subjectivity and imperfection be absent in other areas where human (judgment) has an even more important place? This does not mean that nothing is true and everything is allowed, but that uncertainty is part of every moment and that approximation as accepted elsewhere should also be acceptable in forensic science and in court. However, this is not to be used as excuse for sloppy work.

[149] This is a simplification of the question, but it is enough to show what can raise concern.

Second, if the forensic evidence presented is faulty, it will still depend whether the information obtained by it is also wrong or not, and has any influence. One could imagine an error in manipulation, resulting for instance in attributing a lower concentration of an active substance (purity) in a specimen presumed to contain drugs due to a wrong calibration. Even if the correct purity is higher, the result obtained can nonetheless still be higher than the limit that determines the difference in seriousness of offence for seized cocaine and heroine specimens, which will determine the sanction applicable in Switzerland. The wrong result will have neither an influence on the determination whether an offence took place (yes, because an active illicit substance is present) nor on the severity of the offence, since no difference would be made in the present case. Analytically speaking, however, the quantitative measurement is still erroneous.

Third, even if it is erroneous, it can be regarded as evidence and may play a role in the judgment or ruling. One could imagine for instance that results are judged not relevant for the case. Under these conditions whether the evidence is correct or not plays a minor role. The result can also play a marginal role, for instance the discovery of a shoe wear mark in the entrance hall of a building can be attributed to the wrong shoe and therefore, by inference, a link to a wrong person can be made. However, it can also be that such information is neglected if other evidence is present, for instance an eyewitness testimony.

Fourth, while for now the focus was on situations where the potentially erroneous character of forensic science would not play a role, there might be situations where forensic science is wrongly understood, used, and interpreted though the analytical result has been correct. The most obvious case is when forensic science evidence is present, but for different reasons it is not used at all, though it could help for a better understanding of the questions to be solved by the court. An example would be the case of a suspicious death where no or incomplete crime scene work was performed because the first working hypotheses assumed a natural and/or accidental, explainable death. When going back to the crime scene later, in most circumstances a loss of important “perishable” traces will have taken place. Another example is the over- or underestimation of the factual findings, as happened in the Barry George (Barry George v. Regina, 2007) case. The evidentiary value of a single FDR (firearm discharge residue) was re-evaluated following a change in interpretational doctrine.

Fifth, the most obvious situation is when forensic science is erroneous, is used (correctly), and in consequence leads to a wrong result, probably a judicial error. An example would be the misidentification of the source of a fingermark. To consider and study all these options while using one model of causality seems rather difficult. The existing propositions will be exemplified next.

5.2.2 Existing models

Existing research on judicial error and forensic science has its limitations, as has been shown by the literature review in Chapter 2. This also has its influence on the models proposed to explain judicial errors. One of them has already been presented in the introduction (Chapter

1.2). To recapitulate, it has been proposed that in critical criminology, judicial errors are perceived as being directly related to the unjust action of the criminal justice system and its representatives towards some parts of society, especially different kinds of vulnerable minorities. The main focus of this school of thought can be defined by five points (see Chapter 2.2), which in short describe the criminal justice system as a tool for maintaining social injustice. The (new) role of (critical) criminology is to react against this perceived unjust (criminal justice) system by protecting human rights. Judicial errors tend to have several characteristics which can be seen to agree to a certain degree with the five decisive points mentioned above. The presence of politically motivated crimes that led to judicial errors, like for instance in various IRA bombings in England in the 1970s and the tendency to find more poor, unemployed, less skilled, "darker-skinned" people convicted, incarcerated or executed (Harmon, 2001) seems to confirm this point of view. Alleged and real injustice triggered and still triggers research topics in this area, as illustrated by the study of Gross et al. (2005) for instance.
By perceiving judicial errors as injustice produced "willingly" by the state and its institutions, to protect the rights and chosen privileges of a minority, the causes and sources of judicial error tend to be attributed to a general malfunctioning of the (criminal justice) system, rather than to other more specific explanations. Though it explains several observations made - for instance the overrepresentation of minorities among the exonerated - it also has one major drawback. It renders changes rather difficult, as they depend on the political motivation to modify something, which according to the critics has been voluntarily introduced by the institutions in charge of the situation. This view can also explain tendencies of existing research, focussing on the social characteristics of the victims of judicial errors and the institutional errors that have been committed. This is especially visible on research in the context of wrongful executions, the worst consequence of judicial errors being those in jurisdictions that carry out the death penalty (Chapter 2.3.1).

A different model is proposed by Nobles and Schiff (1995). They present a systems approach which is based on autopoietic theory[150] and where the focus is "to locate the problematic of miscarriages of justice for the legal system and, in particular, the difficulties faced by the legal system if criminal justice is expected to function as something separate from, and different to, other systems such as politics or the media." (Nobles and Schiff, 1995, p. 300). Summarised briefly, they postulate that the criminal justice system is a "system" which functions within itself and which has developed specific ways of understanding and expressing itself, those being necessary to keep the system going. However, this way of functioning is not necessarily understandable for external actors, who in turn have expectations as to how the "system" should function and interact with them. They explain what and why misunderstandings happen, but do not offer suggestions as to how to solve them.

[150] From a very general point of view, the notion of autopoiesis is often associated with that of self-organization. However, an autopoietic system is autonomous and operationally closed, in the sense that every process within it directly helps to maintain the whole (http://www.reference.com/browse/wiki/Autopoiesis, last visited 2008/09/03).

Yet another approach is presented by Savage (2007), who uses the (criminological) restorative justice model to try to explain the existence and importance (motivational forces) of campaigns (for specific cases such as the Guildford Four) against judicial errors (miscarriages of justice)[151] in Britain. This approach considers the victim (of the judicial error) as the centre of attention and the state as the offender. Campaigns against miscarriages of justice are presented as a form of "popular justice", a last resort to force the offender to accept guilt. If he does not, then the restorative justice model cannot be applied correctly, since only one part of the criteria are fulfilled. Savage (2007) concludes that this last factor is missing in most instances and in consequence his model is only partially adapted to the problem at hand.

Besides these models based on criminological theories, other explanations are based on strategies that identify risk factors which can lead to judicial error, as proposed for instance by Edmond (2002), Harmon (2001), or Ramsey (2003). Continuing in this line, various types of factors have been discussed, for example the type of criminal justice system - adversarial or inquisitorial - which might also influence the way forensic science is used.

Most of these propositions, however, do not really integrate forensic science into the judicial error model. As to errors in forensic science (a sort of subgroup of the more general question of forensic science and judicial error), Risinger et al. (2002) have proposed the influence of observer effects on forensic science expertise (see Chapter 2.3.4). Others have followed (Koppl, 2005), taking up the same reasoning. Further reasons for errors have been attributed to wilful fraud, etc. (see Chapter 2.3.2), or again to different types of cognitive biases, such as tunnel vision (the tendency to focus on one aspect to the detriment of all others, for instance one investigative lead). This model seems therefore adapted to explain certain observations and has furthermore the advantage of allowing experimental testing to increase the understanding of the phenomenon, though this has not been explicitly mentioned by the authors.

Still other suggestions (Chapter 2.3.1) centre directly on recommendations as to how to minimise judicial error[152] due to forensic science. One recurrent proposition is to give independent funding (from law enforcement) to forensic laboratories and to submit them to the same "market forces" as other institutions (Forst, 2003; Koppl, 2005; Saks et al., 2001). It is suggested that to introduce competitiveness between forensic laboratories would increase their incentive to produce good quality work. The second proposal is based on introducing quality management measures (Gough, 1997; Jonakait, 1991). The drawback of both propositions is that they are inspired by mechanisms introduced in similar fields (medical, biological, DNA), without always considering the particularities of the field in which they should be implemented. In short, it has not been tested whether the proposed model (economical) is a good fit for the problem (quality of forensic science).

---

[151] Restorative justice as applied by Savage (2007, p. 196) to miscarriages of justice:
"An '*offence*': the wrongful convictions of the Guildford Four and the Maguire Seven
Victim(s): those wrongfully convicted
'*Offender(s)*': the state and its agencies – the police, the courts and so on
An *encounter* and *dialogue*: the meeting between the Prime Minister and Gerry Conlon
An *apology*: the public apology as an 'outcome' or 'resolution' of the process
An element of *healing* or *repair*: what Conlon referred to as the 'start of the end', the healing effect of the apology itself on victims and their families"

[152] Or simply judicial errors. The focus here will be on forensic science.

One common point of all these models is that they consider the subject of forensic science and judicial error (or in the parts of the topic they are interested in) in the context of already existing theories in different fields. This has the advantage of being able to use already established tools for testing the propositions made, as has been done here for cognitive biases in fingermark individualisation (Chapter 3). One initial assumption has to be made, namely that the chosen explanatory framework applies to the question that is studied. That this is sometimes only partially possible has been illustrated by Savage's (2007) example of restorative justice. What can be done when either the primary assumption of applicable framework or when no existing frameworks seem to apply will be discussed in the next chapter. The focus here will not be on the suggestions, but on the construction of testable models to evaluate them. Even some of the models presented before do not de facto explain how they propose to test their assumptions. This lack of testability is an important drawback to making practical recommendations that will work, as discussed previously.

5.2.3 A possible approach to construct models for the relationship between forensic science and judicial error

Existing models can explain only partially the interaction of forensic science and judicial errors. However, one aim is to find a comprehensive model which covers all factors. Abduction can help to create such a general hypothesis. Illustrated by some examples drawn from relevant similar fields, practical solutions based on network theory are proposed to overcome the existing constraints.

One way of constructing a comprehensive model is to use abductive reasoning, as explained by Peirce. "Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea." (Peirce, 1903a, p. 171-172). The process is described by Peirce as follows.

> "A mass of facts is before us. We go through them. We examine them. We find them a confused snarl, an impenetrable jungle. We are unable to hold them in our minds. We endeavor to set them down upon paper; but they seem so multiplex intricate that we can neither satisfy ourselves that what we have set down represents the facts, nor can we get any clear idea of what it is that we have set down. But suddenly, while we are poring over our digest of the facts and are endeavoring to set them into order, it occurs to us that if we were to assume something to be true that we do not know to be true, these facts would arrange themselves luminously. That is abduction." (Peirce, 1903b, p. 282-283)

Several elements presented by Peirce are of interest here. First, facts are needed. What is named facts by Peirce corresponds to data collected, which is considered relevant and reliable

to describe the subject under consideration. Examples could be studies on error rates in various forensic science fields, the percentage of cases in which forensic science is introduced as evidence, and estimates of judicial errors due to forensic science. Existing data is incomplete (chapter 5.1), sometimes biased, and not always relevant to create the necessary facts. This will also influence the use that can be made of them.

Second, the facts, or data, have to be organised to get a clearer picture. This can be done by suggesting an unverified hypothesis. If it is a "good fit", then the facts will be explained by it. One illustration could be the so-called "gravity standard[153]", the "tendency in severe crimes to lower the standard of identification because it is an important case" (Brendel, 2004, p. 16). It suggests that the more heinous the offence[154], the less features (minutiae, etc.) are thought necessary to agree on an identification of a suspect's mark (fingermark) and the greater the risk of error.

Once a hypothesis is found that "fits the facts" the model can be tested and verified by creating additional data. For the example of the "gravity standard", if tested experimentally, the observer becomes the fingermark examiner and the object observed the fingermark (see Chapter 3.2.3). Thus, the influence of a top-down factor, the gravity of the case for instance, on the observation and evaluation of bottom-up factors (quality of the latent fingermark) can be studied empirically[155]. Thus abduction, as used in this small scale example, allows testing of the reasoning advanced. In more general terms, it also helps to create explanatory models based on data[156], with the particularity that those events can be single occurrences. However, it is necessary to have a reliable and sufficient set of data that establishes the facts needed to propose the model.

Thus abduction describes a general way how a hypothesis can be created, based on the facts observed. However, if the facts are difficult to structure, it could be helpful to use tools that help to visually represent the complex situation, by for instance using nodes (objects/vertices) and edges (directed arrows) to describe factors and show how they are interrelated. Furthermore, this can also be the starting point to focus on elements such as the relationships between different entities and/or factors. Indeed, one of the particularities of the subject is that very different entities are linked together and that it is their combination that can create the final outcome, judicial error due to forensic science.

---

[153] This is the term employed in the report.

[154] A more general proposition of special vulnerability of high-profile cases has been discussed by several authors (Gross et al., 2005; Naughton, 2005).

[155] Dror et al. (2005) tested the influence of three top-down factors - emotional background stories of crimes, explicitly disturbing photographs from crime scenes, and subliminal messages - on a rapid "matching" of latent fingerprints. They observed that bad quality or difficult latent fingerprints had the highest risk of being influenced by these top-down manipulations.

[156] The interest of such experimental research resides in the fact that the real influence of one selected and controlled parameter on fingermark comparison can be studied by a monitored test with a control group. Thus, the impact of the "gravity standard" can be investigated and once the experimental parameters have been optimised, other biases can be studied the same way. Furthermore, in a controlled situation it will also be possible to study cases where errors did not occur despite the presence of risk factors. This might give insight into functioning prevention and discovery mechanisms and help making practical recommendations based on an empirically tested hypothesis. Finally, considering all possible outcomes of initial observational bias such as the gravity standard, will allow for a better knowledge both of the real incidence and the impact on the next stage of the cycle, the presentation of evidence in court.

By employing graphs to organise the data (facts) it is possible to represent abductive hypotheses as networks. They are used to describe the structure, the relationship between objects, but also to measure the relative importance of elements of the (social) network by using different measures such as centrality or betweenness (Newman, 2003). They have been utilised for instance to study how forensic scientists learn to become competent in casework reporting (Doak and Assimakoupolos, 2007), to analyse criminal networks (Morselli and Petit, 2007), or to explore criminal mobility and criminal achievement (Morselli and Royer, 2008). Network analysis has also been used in other fields to build cause-effect or risk models such as environmental (fish catch in Switzerland) (Borsuk et al., 2002) or economics (Neil et al., 2005).

Ultimately this leads to a general representation of the relationship of forensic science and judicial error as one complex and comprehensive graph[157]. Based on the publication by Borsuk et al. (2002) on the decline in fish catch in Switzerland the necessary steps to create a model will be presented. In parallel, it will be considered which of these requirements are also fulfilled for the subject of forensic science and judicial error. After defining the aims of the research, the next steps are:

- To collect and evaluate available data as to what already exists on the subject. The literature review has covered most of this point.
- To identify gaps in existing research and information and propose focused future projects. The research made identified gaps, tried to help fill some of them, and made propositions for further studies.
- To use the findings to identify the incidence and impact of various factors and make practical recommendations for improvement. This has been studied empirically for fingermarks and observer bias, but still needs to be done for other fields and factors. The current research has shown that practical recommendations are difficult to make as long as the factors influencing the model have not been ascertained, one of the points put forward in this research (see 5.1).

Building a (causal) network that takes into account all (important) factors takes considerable time and presumes that sufficient data exist to describe the model. Borsuk et al. (2002) consider that they have access to the necessary information for their domain of observation (decline in fish catch in Switzerland), while for the field of forensic science and judicial error, as mentioned previously, it seems first necessary to elicit the data to find facts and sort them. Furthermore, the reliability of the data available has to be evaluated as well. The Saks and Koehler (2005) paper and the critical re-evaluation by Collins and Jarvis (2008; 2009), show that such basic elements are not yet achieved for forensic science and judicial error. Constructing a complex model on imperfect premises questions the utility of the whole undertaking. This is why efforts have been made in this research to empirically test assumptions, which have been presented as facts. The crucial steps of data gathering and evaluation are not finished yet, no attempt could

[157] Here, it would lead to a graphical network representation of the problem, by for instance formalising illustrations used in this research (see for example Figure 4).

be made to test whether it is indeed possible to use network analysis to the question of the relationship between forensic science and judicial error. However, in some fields of forensic science and of judicial error, networks have been proposed or are utilised for studies. Bayesian networks are being used to describe relations of in/dependences between variables, as illustrated by Baio and Corradi (2007). Social network analysis has been suggested by Zalman (2006) to study the innocence movements and their research agenda by examine the context and spread of reform propositions by these actors.

Borsuk et al. (2002) explain that quantitative data and qualitative knowledge can be introduced into a coherent analytical framework, which is built by "viewing the model as a graph" (p. 109) and can result in a (causal) Bayesian network. The causal chain may be split into more detailed occurrences for a better description of causality, depending on existing data and/or knowledge. This corresponds to the propositions advanced in Chapter 5.1. The network also helps to observe what data is needed, what is already available, and how missing data could be obtained. By considering all this, it is then possible to quantify the respective "weight" of each causal variable proposed, as well as to suggest what modification in the system would be necessary to change the final outcome. In other words, the model can be used to explain past observations and predict future occurrences, including the possible effect of voluntary modifications (e.g. measures introduced), either in a general way or for a specific situation. The model can be fed with a range of different types of data which makes it especially adaptable to multidisciplinary research, as is needed for the subject in the present case.
An example of a model using different types of (more and less reliable and conflicting) data to help take decisions in a military context has been described by Piatti et al. (2006) and is called CREDO. The aim of the project was to use credal network theory[158] to help identify the type of aircraft intruding into a protected airspace, to evaluate the reasons for its intrusion and to decide whether it constitutes a security threat or not[159]. To do so, a comprehensive graphical model considering all possible situations has been constructed. It also includes all available information sources, and explains what values these can take. This is a further example that these networks can be adapted to different kinds of situations, therefore also to the relationship between forensic science and judicial errors.

In short, after a review of what is needed to build a model for a better understanding of the relationship between forensic science and judicial error, it seems that some factors are important for a successful attempt. The model should fulfil the following criteria. First, it should allow for a graphical representation of the problem. Second, it should foresee a possibility of inserting various types of data to measure the different elements of the structure. Third, it should include the possibility to test the effects of changes proposed to minimise judicial errors due to forensic science (or any sub-argument studied). (Causal) Bayesian networks seem to combine

[158] A form of Bayesian networks.
[159] Simply put, the type of aircraft and its flight path determine the potential risks for the protected area.

all of these features and could therefore be built to see whether this theoretical assumption of useful model resists empirical testing.

5.3 Lessons learned and outlook

“But it is the mark of all movements, however well-intentioned, that their pioneers tend, by much lashing of themselves into excitement, to lose sight of the obvious.”
(Sayers, 2005, p. 23-24)

I have shown in the previous chapters (5.1 and 5.2) that the relationship between forensic science and judicial error is only partially researched and understood. Data and explanatory models are barely developed. One of the reasons may be that the focus for now has not been on the relevant questions. Are we really putting the effort in the right place and have we really identified the important issues in the relationship between forensic science and judicial error?
One first step towards the answer is to question and verify the underlying assumptions to be sure to build a model on accurate facts. A second step is to make a correct inventory of the existing situation and a third step to propose adapted measures.

1. Verify on what data the arguments advanced are based
"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."
(Conan Doyle, 1981, p. 322[160]*)*

The quality of a model depends on the underlying data that is available to build and test it. Partial data will lead to a partial understanding and therefore biased model. This can lead to focussing on the wrong questions or looking for unnecessary answers. One example is the discussion on the percentage of forensic science errors within the cases treated by the Innocence Project. Saks and Koehler (2005) find that in 63% of the cases there were forensic science testing errors and in 27% of the cases false/misleading testimony of forensic scientists contributed to the wrongful conviction. Collins and Jarvis (2009) reanalyse the same data set and find that due to several (statistical) reasons the contribution of forensic science has been exaggerated, as less than 11% of the cases are concerned by this factor. While their paper reveals some inherent biases in the Saks and Koehler (2005) publication, it (completely) ignores the best argument to contest the proposition. Nearly all cases of the Innocence Project have used DNA to exonerate. This implies that in all these cases forensic science was an (used or not used) evidentiary element existing at the first trial. Thus Saks and Koehler (2005) can say “within a sample of cases where forensic science was present from the beginning we found 63%

[160] The Adventure of the Copper Beeches

of occurrences of difficulties with the technical evidence." As shown by Collins and Jarvis (2009) this value may still be lower. What the data does NOT say is "considering all cases of wrongful convictions forensic science is one of the major contributors." This information is simply not available. What can be said is "In cases of wrongful convictions discovered by DNA analysis and where forensic science was available at the first trial, forensic science has been identified as problematic in 63% or less of the cases." This is quite a different message. This is why it is important to evaluate the quality and interpretation of the data right from the beginning.

A first step is to see what data is missing. It is for instance not known in how many cases forensic science is used in criminal trials. Nor is it known in how many of these cases there are doubts about its correct use. In the same vein, the experimental tests (Chapter 3) demonstrated a discrepancy between assumed and real error risk in fingermark individualisation, probably to a certain degree due to the context in which those arguments were advanced. The well-intentioned goal of "saving" wrongfully convicted may become more important than scientific or factual truth. People are "caught" in their framework, as described by Nobles and Schiff's (1995) model based on autopoietic principles. This can explain the differences in assumptions and arguments between, for example, forensic scientists, psychologists, and legal scholars.
Sometimes information is claimed to be based on data, though when examined more closely this data does not allow to draw the inferences that are made. This can be due to the fact that the data describes a different phenomenon and allows only a partial inference about the information sought. Proficiency tests in fingermark evaluation would be such an example, as they have their limitations to establish an error rate in the domain. Sometimes something is supposed to be true because it seems obvious, but is proven wrong when tested experimentally. One example is the so-called CSI-effect, which is supposed to influence juries/judges to demand more forensic science because of the CSI TV series. Tested experimentally, this hypothesised influence proved nonexistent (Cole and Dioso-Villa, 2007). Furthermore, a considerable amount of reasoning and discussion in literature on forensic science and judicial error is based on only one set of data (Saks and Koehler, 2005), that has been questioned later (Collins and Jarvis, 2008; Collins and Jarvis, 2009) for its quality.

This is why it is important in a second step to elicit the necessary missing data, both on more general and specific issues. This includes for instance to establish an approximate rate of judicial errors, as it has been done by Killias and Gilliéron (2007) for Switzerland. The study has the same limitations that any study based on the establishment of statistics of official data[161] has. Proceeding this way it might be nevertheless more manageable to obtain structured data rather than to try to estimate the judicial error rate by asking for the opinion of (judicial) professionals in the field. They would necessarily have a lacunary vision of the subject (see the study of Ramsey, 2003). It would be interesting to continue research into error rates in forensic science and into the study of specific error sources such as the observer bias. This should lead

[161] See difficulties linked to the different types of statistics available to estimate the crime rate.

to further experimental research as tested and discussed in Chapter 3. The following questions may be asked:

- Does information on the case negatively bias laboratory or other forensic work, as suggested by Risinger et al. (2002)?
- How do cognitive effects influence fingermark comparison at the different stages of the ACE-V process?
- What effects do these cognitive biases have on crime scene work, on the evaluation of results, on DNA profile analysis, and on other forensic fields?
- How and what decisions are taken at each stage of forensic science work; where can biases have their influence?
- Are there people more robust against certain biases than others?
- Can training/education be adapted to efficiently increase resistance against biases?
- What about other types of errors, what can be done to minimise them?

By trying to answer those questions, automatically recommendations for change are tested as well. What limits them is discussed next.

2. Consider constraints when making recommendations

To build an efficient explanatory model it is important to include most influential factors. The network theories discussed not only establish a general view of the subject, but also help pondering the weight of the different factors considered. Some of them may be easily changed or influenced, other are less variable. These constraints and possibilities form the framework of the model. They include the so-called “human factors”.

One of them is subjectivity, whose impact has been studied experimentally in chapter 3 for fingermark individualisation. It was found to have its influence also in other areas of forensic science, including DNA profile analysis (Krane et al., 2008). To completely eliminate subjectivity seems not an achievable goal. It can be minimised, however. Though subjectivity is often perceived as a negative factor it can have positive features. Thus it is crucial to optimising decision making (see Chapter 2.3.3). The focus should therefore not be to eliminate it as far as possible, but rather to clearly delimitate and document its influence. This allows to assess correctly limitations and possibilities given the circumstances, and thus to optimise methodologies.

A second “human factor” concerns communication. It is known, that the same word can be understood differently, depending on the background of the person using it (see Chapter 2.3.2.4). The related concepts will differ accordingly. This can lead to misunderstandings between practitioners of different fields, since expectations on both sides will not be the same. To handle this difficulty, professionals in all disciplines of interest here (forensic science, law, law enforcement, etc.) should get a short introduction of each others work. This could be

achieved through some joint educational classes for all future criminal justice actors at the beginning of their training/education. This would ideally also lead to a shared vocabulary and a better mutual understanding of each others tasks.

Another constraint is the fact that at any transitional stage information is lost. This can also lead to misunderstandings. Thus initial data available at the crime scene may be only partially represented in court. Each time the information is transferred part of it can get lost or its meaning can be distorted. To minimise this effect, the possibility of introducing a forensic science coordinator is projected in France.
On the other hand, too much information on the background of a case may “contaminate” the forensic science processing stage. It creates a certain bias as non scientific information can lead to tunnel view and thus influence (negatively) the work performed on a crime scene or in the laboratory. We think that this hypothesis needs to be tested experimentally. In the meantime forensic science professionals will agree, based on their (European) experience, that it is nearly impossible to render specimens in a case completely anonymous. The usually heavy media coverage, especially in critical high-profile cases (e.g. extensive crime scene TV coverage), makes it (nearly) impossible to completely render anonymous traces/items/specimens received in the laboratory. Interview partners go even farther and argue that case-relevant information is needed for a more efficient processing and evaluation of traces. We think that the general documentation and specification of what was done and how it was done should be the universal solution. Doing this implies, however, having people trained to document thoroughly their methodologies.

The legal/social context in which recommendations are made is a further constraining factor. Particularities of a criminal justice system, for instance, can lead to the observation of specific phenomena, which will not happen in other circumstances. Solutions might therefore not be adapted to all possible situations encountered (system based error). Quirk (2007), for instance, discusses why *Innocence Projects* can be useful in the USA, but less so in the UK. The proposed independence of forensic science laboratories, as often advocated for the USA situation, might not have the same interest in continental European countries. Other “cultural” differences have been discussed in Chapter 4.5.
In summary, it is necessary to understand the framework of the model before making practical recommendations.

3. Propose feasible measures

Feasible measures proposed in this research should be of interest both to legal academics and forensic science practitioners. They include the following:

- ⇨ Develop experimental and empirical research to obtain structured data on all aspects involved in exploring and understanding the relationship between forensic science and judicial error, based on more accurate and reliable information. The testing can also be applied to verify the efficiency of recommendations made (Killias, 2001).

- ⇨ Build models to explain problems such as error risks that are affected by different factors, for instance observational biases. Various points of views will be integrated automatically by this approach. Underlying data should include information obtained from forensic science practitioners. This helps for a good understanding of the subject and a focus on feasible solutions.

- ⇨ Integrate human factors (such as biases) into any explanatory model defining precisely the role and capabilities of forensic science. They influence any work. These factors should be highlighted both in their positive and negative aspects in protocols or methodologies applied to forensic work. On the other hand, protocols put into place will only be as efficient as the persons implementing them. Thus, to have a reasonable scientific method adhered to might be more honest[162] and efficient than to have control mechanisms in place that cannot reasonably be implemented.

- ⇨ Document the work done. This seems to be a safeguard against a certain amount of different errors (risks). This factor helped minimise the risk of observer bias in tests with forensic science students (chapter 3). It also helps to discover what went wrong if it was not possible to prevent errors (MacCoun, 1998).

In short, the same scientific rigour that is expected from the work performed by forensic scientists should also be employed when studying the relationship between forensic science and judicial error.

[162] Or in other words, science is a process, not a truth, honesty is the key.

6. Conclusions

Does forensic science lead to judicial error? This research tries to give a scientifically structured answer to this question and to offer more insight than existing case studies. To reach this goal three social science tools were used: a comprehensive literature review, empirical tests on fingermark individualisation, and semi-directive interviews with (senior) forensic scientists. This allowed to consider the relationship between forensic science and judicial error from different perspectives including, among others, criminology, psychology and purely technical aspects. Keywords that have been discussed are incidence, impact, source, remedy, and practical recommendations.

The literature review gives an overview of the state of the current research. It also identifies subjects of special interest. One of them is the observer bias. Another is the fact that people tend to be less critical of exculpatory than of inculpatory evidence. The inventory of publications helped also to establish an overview of potential error sources in all stages of forensic science work from the crime scene to the writing of a report. It also showed that most publications combining forensic science and judicial error have been written by non-scientists. (Critical) criminology offers an explanatory framework for this observation by placing the discussion on forensic science and judicial error into the larger context of debates on social and judicial (in)justice. This structured analysis also lays a solid foundation for a more objective discussion. It highlights the inherent bias in available information and the need for more research in several areas, such as the decision making process in forensic science.

Experimental research helped evaluate the potential influence of observer bias – the effect of expectations and suggestion on observations – on the fingermark individualisation process. Forensic science students were divided into two groups, one exposed to a neutral condition and the other to biasing factors. The students analysed, compared and evaluated fingermarks (ACE) under varying conditions. Results show that the stages implying decision making tasks in the process – evaluation – are more vulnerable than those based on observations. Analysis and comparison tend to be robust, if participants are asked to document their work comprehensively. One first observation is that experimental research is feasible. It helps elicit factors that cause errors, but also helps establish successful strategies to minimise the impact of cognitive biases and probably also other factors. In this study only fingermarks were tested. However, other forensic science fields can be studied in the same manner. This should help explore other influential factors and find further practical recommendations. It may also be helpful to increase the understanding of underlying processes of a fundamental nature, such as individualisation. Teaching and training would benefit from this input. It may even lead to improving the selection process for forensic science practitioners. In other words, the impact of forensic science work can be increased and the quality improved if underlying weaknesses are well understood.

Semi-directive interviews were used to gather the opinions and experiences of forensic scientists (head of forensic science units and laboratories) on the interaction of forensic science and judicial error. Results indicate that the forensic scientists interviewed are aware of the possibilities and limitations of their discipline. They discussed real and/or potential error sources, the usefulness of training/education, the use of quality management systems to minimise those errors, and their solutions to improve the quality of forensic science work in order to diminish the risks of error. The opinions on most subjects were very similar for all interviewees despite the international sample including six countries.
The main results are the discrepancies found between information obtained in the interviews and through the literature review. Forensic science practitioners and academic authors disagree on error sources in forensic science work. They also disagree, to a lesser extent, on the necessary measures to improve the situation. Non-forensic scientists tend to focus on technical aspects and underestimate the "human factor". This becomes also visible in their practical recommendations. The human "subjective" element is probably more difficult to observe and quantify than the incorrect application of a predefined protocol. Forensic science practitioners tend to highlight the importance of the abovementioned human factors. Empirical research can therefore give fresh and important insights that have been neglected so far.
To systematically include the point of view of the professionals that will be invited to implement changes will both increase the chance of their cooperation and the feasibility/quality of the recommendations made.

Does forensic science lead to judicial error? The experimental results obtained do not allow a direct or an affirmative answer to this question. One reason is that it is very difficult to find or build an explanatory model that will include the most important factors and resist empirical testing. A second reason is the general lack of empirical data to construct and test any model. This lack of data can have far-reaching consequences by leading to a wrong understanding of the problem. It may for instance result in proposing inadequate measures. Empirical and experimental researches are the way out of this difficulty and the best tools to find useful and practical recommendations. It is only thus that the complex interactions and the more general criminal justice system framework can adequately be considered and mutual understanding and communication improved.

"The only ethical principle which has made science possible is that the truth shall be told all the time. If we do not penalize false statements made in error, we open up the way for false statements by intention. And a false statement of fact, made deliberately, is the most serious crime a scientist can commit." (Sayers, 1995a, p. 373)

7. Acknowledgments

In the same way as it seems to me a difficult task to know where to stop after a long (interrupted) time writing this research it seems an even more impossible task to give a comprehensive overview of all persons who contributed in one way or another to making this research possible and help giving it the shape it has. But let us not be deterred by this last stumbling block!

Those that have read faithfully through the pages so far know that people and organisations can be categorised in groups. What you did perhaps not know is that you certainly fall at least into one of the following groups.

At a practical/organisational level the SNSF grant which partially funded this research was a great help to have enough time available for the task at hand. The research also would have been impossible without the collaboration of the forensic science students marking fingermarks in the empirical tests. I would thank all students who willingly participated in the experimental tests. Without them, the observer bias would not have been studied empirically. The cooperation of forensic science practitioners when answering my questions during the interviews was crucial as well. I would also like to thank them for their kindness in sharing their experiences and opinions with me.

At a more academic level I am indebted to Prof. Martin Killias for having proposed the subject which has given me a considerable amount of enjoyment along the way despite some thorny passages. I am also sensitive of the trust put into my research qualities by leaving me a considerable scope in developing the subject the way I did. Prof. Christophe Champod has guided me all along this work, helping me with the necessary counsel and critical comments at the crucial moments. I keep for example in very good memory many talks starting with practical issues and ending with philosophical discussions. It was then his task to gently bring me back to Earth after these metaphysical excursions and remind me of the practical matters. I am also indebted to both the PhD research co-directors and the School of Criminal Justice (ESC) for having accepted and supported on a very short notice my nearly year-long absence from both University tasks and research responsibilities to a temporary assignment at UNIIIC (United Nations International Independent Investigation Commission) in Beirut, Lebanon.
Dr. Itiel Dror has considerably encouraged me on this journey in adding valuable comments to several drafts of this research as well as giving invaluable insight into the complex subject of observational biases in fingermark individualisation. I am further indebted to Dr. Bob Bramley, especially for his comments on my PhD progress report and the general interest shown for this research. To Dr. Arlette Mottaz Baran go special thanks for helping to develop the semi-structured interviews and keeping a critical eye on them. I very much enjoyed the collaboration with Dr. Nathalie Dongois on the SNSF project, allowing me especially to reconsider my subject

from a more juridical (non forensic scientist) point of view, and thus helping me understand by my own experience the communicational issues discussed in this research.
Various other persons contributed with an incredible amount of fructiferous discussions on a variety of questions linked directly to the research, helping clarifying complex issues and reminding me to come back to Earth leaving more philosophical questions for a next PhD research project. They have sacrificed coffee breaks, lunch breaks, and other impromptu moments to help me clarify the questions that seemed imperative at the moment.
Many people have been also been kindly invited (or sometimes urged or forced) to read through smaller or larger sections of this work and have done so graciously, helping thus considerably increase its readability, digestibility, and comprehension. If this should not be the case it is certainly not the responsibility of the following (alphabetical order) proofreaders (that have not been mentioned before in an other function): Đurđica Hazard, Véronique Jaquier, Heidi Newitt, Urs Nyffeler, Joëlle Vuille, and Dr. Céline Weyermann.

At a personal level thanks go to family, friends, and colleagues all contributing by their presence and helpful support in different ways to my general well-being and thus to this research. It is difficult to think of everybody that has crossed my path these last five years. Most still do so regularly, other less so, or not any more and many new persons have become important in a short time. But is much easier to say how they contributed to this work.

- Some were involved in supporting the initial idea of doing a PhD research and seeing me safely off on this journey. This includes especially my close family.
- Others were in charge of encouraging me all along the way with cheers, chocolate, and other necessary bodily and mental nourishment. Family, friends, and colleagues had their share in this.
- Yet others very necessarily helped me forget my project from time to time, whenever a break was needed. This has been especially the case for the judokas of the Reighikan Dojo, Lausanne.

To all you my sincere thanks for having been or being there for me.

8. References

Aebi, M. F., (2003). "Thinking about Wrongful Convictions", First International Workshop on Wrongful Convictions, August 24-26, Breil/Brigels, Switzerland.

Anonymous, (2003). "Cases in Brief", Archbold News, September 19(8).

Anonymous, (2004). "Police Fingerprints", The Boston Globe, 2004/01/2007, http://www.boston.com/news/globe/editorial_opinion/editorials/articles/2004/01/27/police_fingerprints/; last visited 2008/09/03.

Aronson, J. D., (2005). "DNA Fingerprinting on Trial: The Dramatic Early History of a New Forensic Technique", *Endeavour*, 29(3), 126-131.

Arrigo, B. A., (2001). "Critical Criminology, Existential Humanism, and Social Justice: Exploring the Contours of Conceptual Integration", *Critical Criminology*, 10(2), 83-95.

Ashbaugh, D. R., (1999). "Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology", Boca Raton, CRC Press LLC.

Baio, G. and Corradi, F., (2007). "Handling manipulated evidence", *Forensic Science International*, 169(2-3), 181-187.

Barry George v. Regina, (2007). Case No. 200703209B5, Supreme Court of Judicature Court of Appeal (Criminal Division), [2007] EWCA Crim. 2722, 15th of November 2007.

Bellemare, D. A. and Finlayson, R., (2004). "Report on the Prevention of Miscarriages of Justice", FPT Heads of Prosecutions Committee Working Group, http://www.justice.gc.ca/eng/dept-min/pub/pmj-pej/pmj-pej.pdf; last visited 2008/09/03.

Bernard, M., (2005). "L'analyse de la décision et ses représentations graphiques en sciences forensiques", Mémoire de thèse, Lausanne.

Beyleveld, D., (1997). "Ethical Issues in the Forensic Applications of DNA Analysis", *Forensic Science International*, 88(1), 3-15.

Bird, S. J., (2001). "Scientific Certainty: Research versus Forensic Perspectives", *Journal of Forensic Sciences*, 46(4), 978-981.

Blanchet, A., and Gotman, A., (2001). L'enquête et ses méthodes: l'entretien, Nathan Université, Paris.

Borsuk, M. E., Burkhardt-Holm, P., and Reichert, P., (2002). "A Bayesian Network for Investigating the Decline in Fish Catch in Switzerland", Manno: IEMS, 2, 108-113, http://www.iemss.org/iemss2002/proceedings/pdf/volume%20due/46_borsuk.pdf; last visited 2008/09/03.

Bowers, C. M., (1999). "A Statement Why Court Opinions on Bitemark Analysis Should be Limited", *Newsletter of the International Organization for Forensic Odontology-Stomatology*, 21, 11-12.

Bramley, R., (2003). "Quality in the Laboratory", *Science & Justice*, 43(2), 104-108.

Brants, C., (2003). "Wrongful Convictions in the Netherlands", First International Workshop on Wrongful Convictions, August, 24-26, Breil/Brigels, Switzerland.

Brants, C., (2004). "The Vulnerability of the Inquisitorial System of Criminal Law to Wrongful Conviction", 4th Conference of the European Society of Criminology, 25-28 August, Amsterdam.
Brants, C., (2008). "The Vulnerability of Dutch Criminal Procedure to Wrongful Conviction", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 157-182.
Brendel, E., (2004). "INTERPOL European Group on Fingerprint Identification II IEEGFI II; Method for Fingerprint Identification; Part 2; Detailing the Method Using Common Terminology and Through the Definition and Application of Shared Principles", INTERPOL, http://www.interpol.int/Public/Forensic/fingerprints/WorkingParties/IEEGFI2/IEEGF2.pdf; last visited 2008/09/03.
Broeders, A. P. A., (2006). "Of Earprints, Fingerprints, Scent Dogs, Cot Deaths and Cognitive Contamination - a Brief Look at the Present State of Play in the Forensic Arena", *Forensic Science International*, 159(2-3), 148-157.
Bromwich, M. R., (1997). "The FBI Laboratory: An Investigation into Laboratory Practices and Alleged Misconduct in Explosives-Related and Other Cases", http://www.usdoj.gov/oig/special/9704a/; last visited 2008/09/03.
Bromwich, M. R., (2007). "Final Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room", http://www.hpdlabinvestigation.org/reports/070613report.pdf; last visited 2008/09/03.
Bryson, J., (1985). "Evil Angels: the Death of Azaria Chamberlain in the Central Australian Desert, and the Events Leading to Judgement", Ringwood, Viking, Penguin Books Australia Ltd.
Busey, T. A. and Dror, I. E., (in press). "Special Abilities and Vulnerabilities in Forensic Expertise", in McRoberts, A., Friction Ridge Skin Sourcebook.
Campbell, K. M., (2003). "Policy Responses to Wrongful Conviction in Canada: The Role of Conviction Review, Public Inquiries and Compensation", First International Workshop on Wrongful Conviction, August, 24-26, Breil/Brigels, Switzerland.
Campbell, K. M. and Denov, M. S., (2004). "The Burden of Innocence: Coping with a Wrongful Imprisonment", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 139-164.
CCRC. (2005). "Criminal Case Review Commission", http://www.ccrc.gov.uk/index.htm; last visited 2008/09/03.
Champod, C., (2008). "Fingerprint Examination: Towards More Transparency", *Law, Probability and Risk*, 7(1), 111-118.
Champod, C., Evett, I.W., and Kuchler, B., (2002). "Earmarks as Evidence: A Critical Review", *Journal of Forensic Sciences*, 46(6), 1275-1284.
Champod, C., Lennard, C., Margot, P., and Stoilovic, M., (2004). "Fingerprints and Other Ridge Skin Impressions", Boca Raton, CRC Press LLC.
Champod, C., Taroni, F., and Margot, P., (1999). "The Dreyfus Case - An Early Debate on Experts' Conclusions", *International Journal of Forensic Document Examiners*, 5, 446-459.
Christie, A., (2001). "The Murder on the Links", New York, Harper Collins Publisher.

Cole, S. A., (1998). "Witnessing Identification: Latent Fingerprinting Evidence and Expert Knowledge", *Social Studies of Science*, 28(5/6), Special Issue on Contested Identities: Science, Law and Forensic Practice, 687-712.

Cole, S. A., (2005). "More Than Zero: Accounting for Error in Latent Fingerprint Identification", *Journal of Law and Criminology*, 95(3), 985-1078.

Cole, S. A., (2006a). "Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse", *Law and Policy*, 28(1), 109-135.

Cole, S. A., (2006b). "The Prevalence and Potential Causes of Wrongful Conviction by Fingerprint Evidence", *Golden Gate University Law Review*, 37(1), 39-105.

Cole, S. A., (2008a). "Comment on 'Scientific Validation of Fingerprint Evidence under Daubert'", *Law, Probability and Risk*, 7(1), 119-126.

Cole, S. A., (2008b). "The 'Opinionization' of Fingerprint Evidence", *BioSocieties*, 3, 105-113.

Cole, S. A. and Dioso-Villa, R., (2007). "CSI and its Effects: Media, Juries, and the Burden of Proof", *New England Law Review*, 41, 435-470.

Cole, S. A. and Loftus, E. F., (2004). "Contaminated Evidence", Science, 304(5673), 959.

Collins, J. M. and Jarvis, J., (2009). "The Wrongful Conviction of Forensic Science", *Forensic Science Policy & Management: An International Journal*, 1(1), 17-31.

Collins, J. and Jarvis, J., (2008). "The Wrongful Conviction of Forensic Science", (2008/08/02), http://www.crimelabreport.com/library/pdf/wrongful_conviction.pdf; last visited 2008/09/03).

Conan Doyle, A., (1981). "The Penguin Complete Sherlock Holmes", London, Penguin Books.

Connors, E., Lundregan, T., Miller, N., and McEwen, T. (1996). "Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial", *NCJ 161258*, National Institute of Justice.

Cook, R., Evett, I. W., Jackson, G., Jones, P. and Lambert, J., (1998). "A Model for Case Assessment and Interpretation", *Science & Justice*, 38(3), 151-156.

Cooley, C. M., (2003a). "Forensic Science and the Death Penalty: Why Reform is Needed in the Forensic Science Community to Ensure that the Innocent Are Not Erroneously Sentenced to Death", *http://www.law-forensic.com/fs_and_dp.htm*, (last visited 2003/11/01; on file with the author).

Cooley, C. M., (2003b). "Psychological Influences & the State Employed Forensic Examiner: How to Elicit Evidence Concerning Observer Effect Errors through Cross-Examination and Discovery", *Illinois Association of Criminal Defense Lawyers Newsletter*.

Cooley, C. M., (2004). "Symposium: Capital Concerns: The Death Penalty in America: Reforming the Forensic Science Community to Avert the Ultimate Injustice", *Stanford Law & Policy Review*, 15, 381-446.

Cooper, J., and Neuhaus, I. M., (2000). "The "Hired Gun" Effect: Assessing the Effect of Pay, Frequency of Testifying, Credentials on the Perception of Expert Testimony", *Law and Human Behavior*, 24(2), 149-171.

Coquoz, R. and Taroni, F., (2006). "Preuve par l'ADN - la génétique au service de la justice", Lausanne, Presses Polytechniques et Universitaires Romandes.

Crispino, F., (2006). "Le principe de Locard est-il scientifique? Ou analyse de la scientificité des principes fondamentaux de la criminalistique", thèse de doctorat, école des sciences criminelles, Université de Lausanne.
Cullen, F. T. and Agnew, R., (2006). "Part VIII Critical Criminology: Power, Peace, and Crime", in Cullen, F. T. and Agnew, R., Criminological Theory Past to Present Essential Readings, Los Angeles, Roxbury Publishing Company, 294-303.
Daley, B. (2004). "Foolproof Forensics? Even Science May Not Make Verdicts Infallible", Boston Globe, 2004/06/08.
Denov, M. S. and Campbell, K. M., (2004). "Introduction", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 101-106.
Devenport, J. L. and Cutler, B. L., (2004). "Impact of Defense-Only and Opposing Eyewitness Experts on Juror Judgments", *Law and Human Behavior*, 28(5), 569-576.
Doak, S. and Assimakopoulos, D., (2007). "How Do Forensic Scientists Learn to Become Competent in Casework Reporting in Practice: A Theoretical and Empirical Approach", *Forensic Science International*, 167(2-3), 201-206.
Dongois, N., (2006). "Les formes de Bargain en Europe consacrent-elles une Américanisation des droits continentaux?" in Dongois, N. and Killias, M., L'américanisation des droits Suisse et continentaux, Genève, Schulthess Médias Juridiques SA.
Dongois, N., (2008). "Wrongful Convictions in France: the Limits of "Pourvoi en révision"", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 249-261.
Dongois and Schiffer (personal communication). L'article 6 CEDH, facteur d'éviction des erreurs judiciaires?
Donzé, V., (2005). "La morte rentre de vacances..." Le Matin, 15.05.2005.
Douglas, K. S., Lyon, D. R., and Ogloff, J. R. P., (1997). "The Impact of Graphic Photographic Evidence on Mock Jurors' Decisions in a Murder Trial: Probative Or Prejudicial?", *Law and Human Behavior*, 21(5), 485-501.
Dror, I. E. and Charlton, D., (2006). "Why Experts Make Errors", *Journal of Forensic Identification*, 56(4), 600-616.
Dror, I. E. and Rosenthal, R., (2008). "Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts", *Journal of Forensic Sciences*, 53(4), 900-903.
Dror, I. E., Charlton, D., and Péron, A., (2006). "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications", *Forensic Science International*, 156(1), 74-78.
Dror, I. E., Péron, A., Hind, S.-L., and Charlton, D., (2005). "When Emotions Get the Better of Us: The Effect of Contextual Top-down Processing on Matching Fingerprints", *Applied Cognitive Psychology*, 19(6), 799-809.
Eckert, W. G., (1997). "Introduction to Forensic Sciences", New York, CRC Press, Inc.
Edmond, G. (2000). "Judicial Representations of Scientific Evidence", *The Modern Law Review*, 63(2), 216-251.

Edmond, G. (2002). "Constructing Miscarriages of Justice: Misunderstanding Scientific Evidence in High Profile Criminal Appeals", *Oxford Journal of Legal Studies*, 22(1), 53-89.

Ehrlinger, J., Gilovich, T., and Ross, L., (2005). "Peering into the Bias Blind Spot: People's Assessments of Bias in Themselves and Others", *Personality and Social Psychology Bulletin*, 31(5), 680-692.

Enescu, R., (2008). "Influence de l'administration des preuves sur l'activité mentale menant à la formation du verdict au cours d'un procès pénal", thèse de doctorat, école des sciences criminelles, Université de Lausanne, Lausanne.

Epstein, R., (2002). "Fingerprints Meet Daubert: The Myth of Fingerprint "Science" Is Revealed", *Southern California Law Review*, 75, 605-658.

Evett, I. W., Gill, P., Jackson, G., Whitaker, J., and Champod, C., (2002). "Interpreting Small Quantities of DNA: The Hierarchy of Propositions and the Use of Bayesian Networks", *Journal of Forensic Sciences*, 47(3), 520-530.

Evett, I. W., Jackson, G., Lambert, J., and McCrossan, S., (2000). "The Impact of the Principles of Evidence Interpretation on the Structure and Content of Statements", *Science & Justice*, 40(4), 233-239.

Evett, I. W., and Williams, R. L. (1995). "A Review of the Sixteen Point Fingerprint Standard in England and Wales", *Fingerprint Whorld*, 21(82), 125-143.

Fariña, F., Arce, R., and Novo, M., (2003). "Anchoring in Judicial Decision-making", *Psychology in Spain*, 7(1), 56-65.

Feigenson, N., and Dunn, M. A. (2003). "New Visual Technologies in Court: Directions for Research", *Law and Human Behavior*, 27(1), 109-126.

Fereday, M. and Kopp, I., (2002). "European Network of Forensic Science Institutes (ENFSI) and its Quality and Competence Assurance Efforts", *Science & Justice*, 43(2), 99-103.

Fink, A., (1998). "Conducting Research Literature Reviews: From Paper to the Internet", Thousand Oaks, California, SAGE Publications, Inc.

Forst, B., (2003). "If Wrongful Convictions Were a Disease, What Would Be the Cure?", First International Workshop on Wrongful Convictions, August 24-26, Breil/Brigels, Switzerland.

Found, B. and Rogers, D., (2008). "The probative character of Forensic Handwriting Examiners' identification and elimination opinions on questioned signatures", *Forensic Science International*, 178(1), 54-60.

Frankel, M. S., (1989). "Ethics and the Forensic Science: Professional Autonomy in the Criminal Justice System." *Journal of Forensic Sciences*, 34(3), 763-771.

Freckelton, I. (1989). "DNA Profiling: Forensic Science under the Microscope." *DNA and Criminal Justice: Proceedings of a Conference 30-31 October*, Canberra, 1-47.

Freckelton, I. and Selby, H., (2002). "Scientific Evidence as Circumstantial Evidence", Expert Evidence: Law, Practice, Procedure and Advocacy, 1-2, http://www.aic.gov.au/publications/proceedings/02/freckelton.html; last visited 2008/09/03.

Fresco, A., (2004). "Earprint Evidence a 'Grotesque Injustice'", *Fingerprint Whorld*, 30(1), 57.

Gans, J., and Urbas, G., (2002). "DNA Identification in the Criminal Justice System", *Trends & Issues in Crime and Criminal Justice,* 226, 1-6.

Geller, B., Almog, J., Margot, P., and Springer, E., (1999). "A Chronological Review of Fingerprint Forgery", *Journal of Forensic Sciences*, 44(5), 963-968.

Geller, B., Almog, J., and Margot, P., (2001). "Fingerprint Forgery - a Survey", *Journal of Forensic Sciences*, 46(3), 731-733.

Gervais, C., (2004). "Until You Are Dead: Steven Truscott's Long Ride into History", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 209-212.

Giannelli, P. C., (2001). "Impact of Post-Conviction DNA Testing on Forensic Science", *New England Law Review*, 35(3), 627-638.

Giannelli, P. C., (2002) "Scientific Evidence & Miscarriages of Justice", *16th International Conference of the International Society for the Reform of Criminal Law*, Charleston, SC, USA, 1-14.

Giannelli, P. C., (2007). "Wrongful Convictions and Forensic Science: The Need to Regulate Crime Labs", *North Carolina Law Review*, 86(1), 163-236.

Gill, P., (2002). "Role of Short Tandem Repeat DNA in Forensic Casework in the UK – Past, Present, and Future Perspectives", *BioTechniques*, 32, 366-385.

Gill, P., Whitaker, J. P., Flaxman, C., Brown, N., and Buckleton, J., (2000). "An Investigation of the Rigor of Interpretation Rules for STRs Derived from Less than 100pg of DNA", *Forensic Science International*, 112(1), 17-40.

Goldsmith, (2004). "Attorney General Announces Results of Shaken Baby Syndrome Cases Review, and New Guidance for Expert Witnesses", Attorney General's Office, http://www.attorneygeneral.gov.uk/sub_news_press.htm; last visited 2008/09/03.

Goodstein, D., (2000). "How Science Works", in US Federal Judiciary Reference Manual on Evidence, 66-72.

Gough, T. A., (1997). "Quality Assurance in Forensic Science: the UK Situation", *Accreditation and Quality Assurance*, 2(5), 216-223.

Grometstein, R., (2008). "Wrongful Conviction and Moral Panic National and International Perspectives on Organized Child Sexual Abuse", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 11-32.

Gross, S. R., (1998). "Lost Lives: Miscarriages of Justice in Capital Cases", *Law and Contemporary Problems*, 61(4), 125-154.

Gross, S. R., Jacoby, K., Matheson, D., Montgomery, N. and Patel, S., (2005). "Exonerations in the United States 1989 through 2003", *Journal of Criminal Law and Criminology*, 95(2), 1-32.

Gualco, B., (2003). "Guilty of Innocence: Judicial Error, the Review Procedure and Awarding Damages", First International Workshop on Wrongful Convictions, August, 24-26, Breil/Brigels, Switzerland.

Gudjonsson, G. H., (2002). "Unreliable Confession and Miscarriages of Justice in Britain", *International Journal of Police Science and Management*, 4(4), 332-343.

Gysin, R., (2005). "Der Fingerabdruck steht unter Verdacht", *Plädoyer*, 6, 11-13.
Haack, S., (2003). "Inquiry and Advocacy, Fallibilism and Finality: Culture and Inference in Science and the Law", *Law, Probability and Risk*, 2(3), 205-214.
Haber, L. and Haber, R., (2004). "Error Rates for Human Latent Fingerprint Examiners", in Ratha, N. and Bolle, R., Advances in Automatic Fingerprint Recognition, New York, Springer, 339-360.
Haber, L. and Haber, R. N., (2008a). "Scientific Validation of Fingerprint Evidence under Daubert", *Law, Probability and Risk*, 7(1), 87-109.
Haber, L. and Haber, R. N., (2008b). "Experiential or Scientific Expertise", *Law, Probability and Risk*, 7(1), 143-150.
Harmon, R., Budowle, B., Langenburg, G. M., Houck, M. M., Kelly, J. S., Saks, M. J., and Koehler, J. J., (2006). "Questions about Forensic Science", *Science*, 311, 607-610.
Harmon, T. R., (2001). "Predictors of Miscarriages of Justice in Capital Cases", *Justice Quarterly*, 18(4), 949-968.
Hauck, J. G., (1996). "Zur quantitativen Beurteilung des Beweiswertes von Spurenmaterial", *Kriminalistik und forensische Wissenschaften* (85), 1-13.
Henderson, J. P., (2002). "The Use of DNA Statistics in Criminal Trials", *Forensic Science International*, (128), 183-186.
Hickman, T. A., (2004). "Wrongful Convictions and Commissions of Inquiry: A Commentary", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 183-188.
Hil, R. and Robertson, R., (2003). "What Sort of Future for Critical Criminology?", *Crime, Law & Social Change*, 39, 91-115.
Hoel, A., (2008). "Compensation for Wrongful Conviction", *Trends & Issues in Crime and Criminal Justice*, 356, 1-6.
Hopkins Burke, R., (2005). "An Introduction to Criminological Theory", Cullumpton, Willan Press.
House of Commons, (2005). "Forensic Science on Trial Seventh Report of Session 2004-05", House of Commons Science and Technology Committee.
Howard, R. A., (2002). "Decision Analysis and Law", *Cardozo Law Review*, 22, 1595-1603.
Huff, R., (2002). "Wrongful Convictions and Public Policy: the American Society of Criminology 2001 Presidential Address", *Criminology*, 40(1), 1-18.
Huff, R., (2003). "Wrongful Convictions: The American Experience", First International Workshop on Wrongful Convictions, August 24-26, Breil/Brigels, Switzerland.
Huff, R., (2004). "Wrongful Convictions: The American Experience", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 107-120.
Huff, R., (2008). "Wrongful Convictions in the United States", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 59-70.
Huff, R. and Killias, M., (2002). "Wrongful Conviction: Public Policy and Cross-National Issues", European Society of Criminology Annual Meeting, 4-7 September, Toledo, Spain.

Huff, R. and Killias, M., (2008). "Introduction", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 3-10.

Huff, R., Rattner, A., and Sagarin, E., (1996). "Convicted but Innocent: Wrongful Conviction and Public Policy", Thousand Oaks, CA, Sage Publications.

Hy, R. J., Feig, D. G. and Regoli, R. M., (1983). "Research Methods and Statistics: A Primer for Criminal Justice & Related Sciences", Cincinnati, Ohio, Anderson Publishing Co.

IAI, (1973). "Report of the Standardization Committee of the International Association for Identification - August 1, 1973", *Identification News*, 13-14.

IAI, (1980). "Resolution VII Amended", *Identification News*, August, 3.

Inman, K. and Rudin, N., (2000). "Principles and Practices of Criminalistics: The Profession of Forensic Science", Boca Raton, CRC Press LLC.

INRWC, (2003). "First International Workshop on Wrongful Convictions", August 24-26, Breil/Brigels, Switzerland.

Jackson, G., (2000). "The Scientist and the Scales of Justice", *Science & Justice*, 40(2), 81-85.

Jackson, G., Jones, S., Booth, G., Champod, C., and Evett, I. W., (2006). "The Nature of Forensic Science Opinion - a Possible Framework to Guide Thinking and Practice in Investigations and in Court Proceedings", *Science & Justice*, 46(1), 33-44.

Jakobs, L. E., and Sprangers, W. J., (2000). "A European View on Forensic Expertise and Counter-expertise", *Criminal Law Forum*, 11, 375-392.

Jehle, J.-M., (2003). "Wrongful Convictions from the German Criminal Justice Perspective", First International Workshop on Wrongful Convictions, August, 24-26, Breil/Brigels, Switzerland.

Jobling, M. A. and Gill, P., (2004). "Encoded Evidence: DNA in Forensic Analysis", *Nature*, 5(10), 739-752.

Johnson, C., (2003). "Cleared Murder Accused Victim of DNA Blunder", http://icLiverpool.co.uk, 2004/03/10, http://icliverpool.icnetwork.co.uk/0100news/0100regionalnews/tm_objectid=12718961&method=full&siteid=50061&page=2&headline=cleared-murder-accused-victim-of-dna-blunder-name_page.html; last visited 2008/09/03.

Johnson, P., and Williams, R., (2004). "Post-conviction DNA Testing: the UK's First "Exoneration" Case?", *Science & Justice*, 44(2), 77-82.

Johnson-Laird, P. N., (1999). "Deductive Reasoning", *Annual Review of Psychology*, 50, 109-135.

Jonakait, R. N., (1991). "Forensic Science: The Need for Regulation", *Harvard Journal of Law & Technology, Harvard Law School*, 4(1), 109-191.

Jones, S., (2006). "Criminology", Oxford, Oxford University Press.

Kaufman, F., (1998). "The Commission on Proceedings Involving Guy Paul Morin", Ontario Ministry of the Attorney General, http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/morin/; last visited 2008/09/03.

Kaufman, H., (2001). "The Expert Witness: Neither Frye nor Daubert Solved the Problem: What Can Be Done?", *Science & Justice*, 41(1), 7-20.

Kaufmann, J.-C., (2001). *L'entretien compréhensif*, Nathan Université, Paris.

Kaye, D. H. and Koehler, J. J., (1991). "Can Jurors Understand Probabilistic Evidence?", *Journal of the Royal Statistical Association*, 154(1), 75-81.

Kennedy, J. P., (2004). "Writing the Wrongs: The Role of Defense Counsel in Wrongful Convictions - A Commentary", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 197-208.

Kershaw, S., and Lichtblau, E., (2004). "Spain Questioned FBI's Print Match", International Herald Tribune, 2004/05/27.

Kerstholt, J. H., and Jackson, J. L., (1998). "Judicial Decision Making: Order of Evidence Presentation and Availability of Background Information", *Applied Cognitive Psychology*, 12, 445-454.

Kerstholt, J. H., Paashuis, R., and Sjerps, M., (2007). "Shoe Print Examinations: Effects of Expectation, Complexity and Experience", *Forensic Science International*, 165(1), 30-34.

Kessler, I., (2008). "A Comparative Analysis of Prosecution in Germany and the United Kingdom Searching for Truth or Getting a Conviction?", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 213-247.

Killias, M., (2001). "Précis de criminologie", Berne, Staempfli Editions SA.

Killias, M., (2003). "Wrongful Conviction: Switzerland As a Country With a Continental System", First International Workshop on Wrongful Convictions, August 24-26, Breil/Brigels, Switzerland.

Killias, M., (2006). "Les erreurs judiciaires: un problème américain?", in Dongois, N. and Killias, M., L'américanisation des droits Suisse et continentaux, Genève, Schulthess Médias Juridiques SA, 69-80.

Killias, M., (2008). "Wrongful Convictions in Switzerland the Experience of a Continental Law Country", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 139-155.

Killias, M., Champod, C., and Bär, W. (2004). "Wrongful Convictions in Switzerland in a Comparative Perspective", SNSF, (100012-105817/1), Lausanne.

Killias, M., Gilliéron, G., and Dongois, N., (2007). "A Survey of Exonerations in Switzerland over Ten Years (Report to the Swiss National Science Foundation, 100012-105817/1)", University of Lausanne and University of Zurich, http://www.rwi.uzh.ch/lehreforschung/alphabetisch/killias/forschung/Erreurs_judiciaires_rapport_FNS.pdf; last visited 2008/09/03.

Killias, M., Kuhn, A., Dongois, N., and Aebi, M. F., (2008). "Précis de droit pénal général", Berne, Staempfli Editions SA.

Kirk, P. L., (1963). "The Ontogeny of Criminalistics", *The Journal of Criminal Law, Criminology, and Police Science*, 54(2), 235-238.

Koehler, J. J., (2001). "When Are People Persuaded by DNA Match Statistics?", *Law and Human Behavior*, 25(5), 493-513.
Koehler, J. J., (2008a). "A Welcome Exchange on the Scientific Status of Fingerprinting", *Law, Probability and Risk*, 7(1), 85-86.
Koehler, J. J., (2008b). "Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter", *Hastings Law Journal*, 59, 1077-1098.
Koehler, J. J. and Macchi, L., (2004). "Thinking about Low-Probability Events; an Exemplar-Cuing Theory", *Psychological Science*, 15(8), 540-546.
Koppl, R., (2005). "How to Improve Forensic Science", *European Journal of Law and Economics*, 20, 255-286.
Krane, D. E., Ford, S., Gilder, J. R., Inman, K., Jamieson, A., Koppl, R., Kornfield, I. L., Risinger, M. D., Rudin, N., Taylor, M. S., and Thompson, W. C., (2008). "Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation", *Journal of Forensic Sciences*, 53(4), 1006-1007.
Kuhn, A. and Enescu, R., (2007). "Sentencing: effets d'ordre et paradoxe de la condamnation", University of Lausanne, http://www.unil.ch/webdav/site/esc/shared/kuhn/FNS-Sentencing-2007.pdf; last visited 2008/09/03.
Lander, E. S. and Budowle, B., (1994). "DNA Fingerprinting Dispute Laid to Rest", *Nature*, 371, 735-738.
Langenburg, G. M. (2004). "Pilot Study: A Statistical Analysis of the ACE-V Methodology - Analysis Stage", *Journal of Forensic Identification*, 54(1), 64-79.
Langenburg, G. M., (personal communication). "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".
Leahy, P., (2001). "Serenity Now or Insanity Later? The Impact of Post-Conviction DNA Testing on the Criminal Justice System Introduction", *New England Law Review*, 35(3), 605-607.
Lempert, R., (1997). "After the DNA Wars: Skirmishing with NRC II", *Jurimetrics Journal of Law, Science and Technology*, 37(1), 439-468.
Leo, R. A., (2005). "Rethinking the Study of Miscarriages of Justice: Developing a Criminology of Wrongful Conviction", *Journal of Contemporary Criminal Justice*, 21(3), 201-223.
Lewis, R. (2001). "Scientists in the Dock: Lawyers, Scientists and the Prosecution of Offenders - A Personal View", Deakin University, unpublished article, Geelong, 1-21; last visited 2004/05/18; paper on file with the author).
Locard, E., (1920). "L'enquête criminelle et les méthodes scientifiques", Paris, Flammarion.
Loftus, E. F., and Cole, S. A., (2004). "Contaminated Evidence", *Science*, 304(5673), 959.
Lowe, A., Murray, C., Whitaker, J. P., Tully, G., and Gill, P., (2002). "The Propensity of Individuals to Deposit DNA and Secondary Transfer of Low Level DNA from Individuals to Inert Surfaces", *Forensic Science International*, 129(1), 25-34.
Lucas, D. M., (1989). "The Ethical Responsibilities of the Forensic Scientist: Exploring the Limits", *Journal of Forensic Sciences*, 34(3), 719-729.

Lynch, M. (2003). "God's Signature: DNA Profiling, the New Gold Standard in Forensic Science", *Endeavour*, 27(2), 93-97.
Lynch, M., and McNally, R., (2003). ""Science," "Common Sense," and DNA Evidence: a Legal Controversy About the Public Understanding of Science", *Public Understanding of Science*, 12(1), 83-103.
MacCoun, R. J., (1998). "Biases in the Interpretation and Use of Research Results", *Annual Review of Psychology*, 49, 259-287.
MacFarlane, B. A. (2003). "Convicting the Innocent - A Triple Failure of the Justice System", http://www.canadiancriminallaw.com/articles/articles%20pdf/Convicting%20the%20Innocent.pdf; last visited 2008/09/03.
Maguire, A., (1993). "Miscarriage of Justice: An Irish Family's Story of Wrongful Conviction as IRA Terrorists", http://www.innocent.org.uk/cases/maguire7/moj.pdf; last visited 2008/09/03.
Malkoc, E. and Neuteboom, W., (2007). "The Current Status of Forensic Science Laboratory Accreditation in Europe", *Forensic Science International*, 167(1), 121-126.
Malpass, R. S., and Lindsay, R. C. L., (1999). "Measuring Line-up Fairness", *Applied Cognitive Psychology*, 13, S1-S7.
Martin, J.-C. (2002). Investigation de scène de crime; fixation de l'état des lieux et traitement des traces d'objets, Lausanne, Presses polytechniques et universitaires romandes.
Mathieson, A.-R., and Gross, S. R., (2003). "Review for Error", *Law, Probability and Risk*, 2(4), 259-268.
McCall, G. J., (1984). "Systematic Field Observation", *Annual Review of Sociology*, 10(1), 263-282.
McClure, D., (2007). "Focus Group on Scientific and Forensic Evidence in the Courtroom", National Institute of Justice, report No.: 220692, http://www.ncjrs.gov/pdffiles1/nij/grants/220692.pdf.
McDougall, L., (2004). "Expert Warns Fingerprinting System is 'Riddled with Flaws'", Sunday Herald, 2004/05/30.
Mehlis, D., (2005). "Report of the International Independent Investigation Commission established pursuant to Security Council Resolution 1595 (2005)", UNIIIC, http://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/Lebanon%20S2005662.pdf; last visited 2008/09/03.
Meili, M., (2005). "Tatort Fingerabdruck", NZZ am Sonntag, 2005/10/02.
Meissner, C. A., and Kassin, S. M., (2002). ""He's Guilty!" Investigator Bias in Judgments of Truth and Deception", *Law and Human Behavior*, 26(5), 469-480.
Mellers, B. A., Schwartz, A., and Cooke, A. D. J., (1998). "Judgment and Decision Making", *Annual Review of Psychology*, 49, 447-477.
Mnookin, J. L., (2001). "Fingerprint Evidence in an Age of DNA Profiling", *Brooklyn Law Review*, 67(1), 13-71.
Mnookin, J. L., (2003). "Fingerprints: Not a Gold Standard", *Issues in Science and Technology Online*, http://www.issues.org/20.1/mnookin.html; last visited 2008/09/03.

Mnookin, J. L., (2004). "A Blow to the Credibility of Fingerprint Evidence", The Boston Globe, 02/02/2004, http://www.boston.com/news/globe/editorial_opinion/oped/articles/2004/02/02/a_blow_to_the_credibility_of_fingerprint_evidence/; last visited 2008/09/03.

Mnookin, J. L., (2008). "The Validity of Latent Fingerprint Identification: Confessions of a Fingerprint Moderate", *Law, Probability and Risk*, 7(1), 127-141.

Moenssens, A. A., (2000). "A Mistaken DNA Identification? What Does it Mean?", http://www.forensic-evidence.com/site/EVID/EL_DNAerror.html; last visited 2008/09/03.

Morling, T. R., (1987). "Royal Commission of Inquiry into Chamberlain Convictions", Report of the Commissioner", Darwin, Northern Territory.

Morselli, C. and Petit, K., (2007). "Law-Enforcement Disruption of a Drug Importation Network", *Global Crime*, 8(2), 109-130.

Morselli, C. and Royer, M.-N., (2008). "Criminal Mobility and Criminal Achievement", *Journal of Research in Crime and Delinquency*, 45(1), 4-21.

Naughton, M., (2003). "How Big is the 'Iceberg'? - A Zemiological Approach to Quantifying Miscarriages of Justice", *Radical Statistics,* 81, 5-17.

Naughton, M., (2005). "Redefining Miscarriages of Justice: A Revived Human-Rights Approach to Unearth Subjugated Discourses of Wrongful Criminal Conviction", *British Journal of Criminology*, 45(2), 165-182.

Neil, M., Fenton, N. and Tailor, M., (2005). "Using Bayesian Networks to Model Expected and Unexpected Operational Losses", *Risk Analysis*, 25(4), 1-10.

Neubauer, D. W., (1999). "America's Courts and the Criminal Justice System", Belmont, West/Wadsworth.

Neumann, C., Champod, C., Puch-Solis, R., Egli, N. M., Anthonioz, A., and Bromage-Griffiths, A., (2007). "Computation of Likelihood Ratios in Fingerprint Identification for Configurations of Any Number of Minutiæ", *Journal of Forensic Sciences*, 52(1), 54-64.

Newman, M. E. J., (2003). "The Structure and Function of Complex Networks", SIAM Review, 45(2), 167-256, http://www-personal.umich.edu/~mejn/courses/2004/cscs535/review.pdf; last visited 2008/09/03.

Nobles, R. and Schiff, D., (1995). "Miscarriages of Justice: A Systems Approach", *The Modern Law Review*, 58(3), 299-320.

Nobles, R., and Schiff, D., (2001). "The Criminal Case Review Commission: Reporting Success?" *The Modern Law Review*, 64(2), 280-299.

Nobles, R. and Schiff, D., (2004). "A Story of Miscarriage: Law in the Media", *Journal of Law and Society*, 31(2), 221-244.

Obuchowski, N. A. (2003). "Special Topics III: Bias", *Radiology*, 617-621.

OIG (Office of the Inspector General), (2006). "A Review of the FBI's Handling of the Brandon Mayfield Case", http://www.nlada.org/Defender/forensics/for_lib/Documents/1145382069.31/OIGfinal.pdf; last visited 2008/09/03.

Otto, H.-D., (2004). "Das Lexikon der Justizirrtümer: skandalöse Fälle, unschuldige Opfer, hartnäckige Ermittler", Mönchengladbach, Ullstein Verlag.

Parks, B. C., (2000). "Board of Inquiry into the Rampart Area Corruption Incident Executive Summary", Los Angeles Police Department, http://www.lapdonline.org/assets/pdf/boi_pub.pdf; last visited 2008/09/03.

Parmley, M. C., (2006). "The Effects of the Confirmation Bias on Diagnostic Decision Making", Philadelphia, Drexel University.

Peirce, C. S., (1903a). "Harvard Lectures on Pragmatism", in Hartshorne, C. and Weiss, P., Collected Papers of Charles Peirce, Cambridge, Mass., Harvard University Press, 171-172.

Peirce, C. S., (1903b). "Harvard Lectures on Pragmatism, a Deleted Passage", in Turrisi, P. A., Pragmatism as a Principle and Method of Right Thinking. The 1903 Harvard Lectures on Pragmatism, Albany, State University of New York Press, 282-283.

Penacino, G., Sala, A., and Corach, D., (2003). "Are DNA Tests Infallible?", *International Congress Series*, 1239, 873-877.

Peterson, J. L., Lin, G., Ho, M., Chen, Y., and Gaensslen, R. E., (2003a). "The Feasibility of External Blind DNA Proficiency Testing. I. Background and Findings", *Journal of Forensic Sciences*, 48(1), 21-31.

Peterson, J. L., Lin, G., Ho, M., Chen, Y., and Gaensslen, R. E., (2003b). "The Feasibility of External Blind DNA Proficiency Testing. II. Experience with Actual Blind Tests", *Journal of Forensic Sciences*, 48(1), 32-40.

Piatti, A.. Zaffalon, M., and Antonucci, A. (2006). Auswahl und provisorische Spezifizierung der kredalen Struktur Project CREDO, Vetrag Nr.4500317924. Meilenstein 1 Aramis Nr. R-3210/044-04.

Poveda, T. G., (2001). "Estimating Wrongful Convictions", *Justice Quarterly*, 18(3), 689-708.

Pretty, I. A. and Sweet, D., (2001). "The Scientific Basis for Human Bitemark Analyses - A Critical Review", *Science & Justice*, 41(2), 85-92.

Pretty, I. A. and Turnbull, M. D., (2001). "Lack of Dental Uniqueness between Two Bite Mark Suspects", *Journal of Forensic Sciences*, 46(6), 1487-1491.

Quirk, H., (2007). "Identifying Miscarriages of Justice: Why Innocence in the UK is Not the Answer", *The Modern Law Review*, 70(5), 759-777.

R v. Ward, (1993). Court of Appeal (Criminal Division), [1993] 1 WLR 619, 96 Cr App Rep 1, [1993] 2 All ER 577, http://www.innocent.org.uk/cases/judithward/ward.pdf; last visited 2008/09/03.

Ramsey, R. J., (2003). "False Positives in the Criminal Justice Process - An Analysis of Factors Associated With Wrongful Conviction of the Innocent," PhD Thesis, University of Cincinnati, Cincinnati.

Randerson, J., and Coghlan, A., (2004). "Forensic Evidence Stands Accused", *Fingerprint Whorld*, 30(117), 105-108.

Rattner, A., (1995). "Social Science v. the Judicial System: The Impact of the Accumulated Knowledge of Eyewitness Identification on Criminal Procedures", *International Journal of the Sociology of Law*, 23, 97-106.

Rattner, A., (2003). "The Sanctity of the Criminal Law: Thoughts and Reflections on Wrongful Conviction in Israel", First International Workshop on Wrongful Convictions, August, 24-26, Breil/Brigels, Switzerland.

Rawley, A. and Caddy, B., (2007). "Damilola Taylor: An Independent Review of Forensic Examination of Evidence by the Forensic Science Service", Home Office, http://www.homeoffice.gov.uk/documents/damilola-taylor-review-2007?view=Binary; last visited 2008/09/03.

Risinger, M. D., (2007). "Innocents Convicted: An Empirically Justified Conviction Rate", *The Journal of Criminal Law & Criminology*, 97(3), 761-806.

Risinger, M. D., Saks, M. J., Thompson, W. C., and Rosenthal, R., (2002). "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion", *California Law Review*, 90(1), 1-56.

Rivola, M., (2004). "L'échange de connaissances en police scientifique en Suisse," Postgraduate Research, Lausanne University, Lausanne.

Roberts, P. C., (2003a). "The Causes of Wrongful Conviction", *The Independence Review*, 7(4), 567-574.

Roberts, S., (2003b). ""Unsafe" Convictions: Defining and Compensating Miscarriages of Justice", *The Modern Law Review*, 66(3), 441-451.

Roberts, P., and Willmore, C., (1993). "The Royal Commission on Criminal Justice - The Role of Forensic Science Evidence in Criminal Proceedings", Research Study No 11, London, HSMO.

Roux, C., Beavis, A., Benson, S., Braybon, E., Dawson, M., Doble, P., Flynn, K., Payne, G., Lennard, C., Maynard, P., Reedy, B., Robertson, J., Tahtouh, M., Wallace-Kunkel, C., and Walsh, S., (2007). "Forensic Science in the 21st Century - Will Trace Evidence Ever Reach the Next Level?", http://projects.nfstc.org/trace/docs/final/Roux_21st_century.pdf; last visited 2008/09/03.

Rudram, D. A., (1996). "Interpretation of Scientific Evidence", *Science & Justice*, 36(3), 133-138.

Ryan, G. H., (2002). "Report of the Governor's Commission on Capital Punishment", Commission on Capital Punishment, http://www.idoc.state.il.us/ccp/ccp/reports/commission_report/complete_report.pdf; last visited 2008/09/03.

Saks, M. J. (1989). "Prevalence and Impact of Ethical Problems in Forensic Science", *Journal of Forensic Sciences*, 34(3), 772-793.

Saks, M. J., (1998). "Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science", *Hastings Law Journal*, 49(4), 1069-1141.

Saks, M. J., Constantine, L., Dolezal, M., Garcia, J., Titus, C., Horton, G., Leavell, T., Muntz, J., Rivera, L., Steward, J., Strumpf, F., and VanderHaar, H., (2001). "Toward a Model Act for the Prevention and Remedy of Erroneous Convictions", *New England Law Review*, 35(3), 669-683.

Saks, M. J. and Koehler, J. J., (1991). "What DNA "Fingerprinting" Can Teach the Law About the Rest of Forensic Science?", *Cardozo Law Review*, 13, 361-372.

Saks, M. J., and Koehler, J. J., (2005). "The Coming Paradigm Shift in Forensic Identification Science", *Science*, 309, 892-895.

Saks, M. J., Risinger, M. D., Rosenthal, R., and Thompson, W., (2003). "Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States", *Science & Justice*, 43(2), 77-90.

Saltzman, J., (2004). "Judge Sets Bail in DNA Case; New Trial Sought in '97 Shooting of Hub Officer", The Boston Globe, http://www.truthinjustice.org/cowans.htm, 22/01/2004; last visited 2008/09/03.

Saltzman, J. and Daniel, M., (2004). "Man Freed in 1997 Shooting of Officer; Judge Gives Ruling After Fingerprint Revelation", The Boston Globe; http://www.truthinjustice.org/cowans2.htm, 24/01/2004; last visited 2008/09/03.

Savage, S. P., (2007). "Restoring Justice Campaigns Against Miscarriages of Justice and the Restorative Justice Process", *European Journal of Criminology*, 4(2), 195-216.

Sayers, D. L., (1995a). "Gaudy Night", New York, Harper Collins Publisher.

Sayers, D. L., (1995b). "Unnatural Death", New York, Harper Collins Publishers.

Sayers, D. L., (2005). "Are Women Human?" Grand Rapids, Michigan, William B. Eerdmans Publishing Company.

SCCRC, (2005). "Scottish Criminal Case Review Commission", http://www.sccrc.org.uk/home.aspx; last visited 2008/09/03.

Scheck, B., Neufeld, P., and Dwyer, J., (2001). *Actual Innocence: When Justice Goes Wrong and How to Make It Right*, New York, SIGNET Penguin.

Schiffer, B., (2005). "Scientific Expertise: Incidence and Impact on Judicial Errors (Work in Progress)", 5th Annual Conference of the European Society of Criminology, 31 August - 3 September, Cracow.

Schiffer, B., (2005). "The Observer Bias in Scientific Expertise: Incidence and Impact on Judicial Errors", PhD Thesis Project Progress Report, University of Lausanne, Lausanne.

Schiffer, B. and Champod, C., (2004). "Judicial Error and Forensic Science: Pondering the Contribution of (DNA) Evidence", 4th Annual Conference of the European Society of Criminology, 25-28 August, Amsterdam, Netherlands.

Schiffer, B. and Champod, C., (2006). "Experimental Research on the Possible (Negative) Impact of Observational Biases in the Analysis Stage of Fingerprint Individualisation", poster presented at the 4th EAFS Conference 14-16 June, Helsinki.

Schiffer, B. and Champod, C., (2007). "The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation", *Forensic Science International*, 167(2-3), 116-120.

Schiffer, B. and Champod, C., (2008). "Judicial Error and Forensic Science Pondering the Contribution of DNA Evidence", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 33-55.

Schiffer, B., Champod, C. and Killias, M., (2006). "Miscarriages of Justice: The Comparative Point of View of Senior Forensic Scientists or Managers", 6th Annual Conference of the European Society of Criminology, 26-29 August, Tübingen.

Schiffer, B. and Stauffer, E., (in press). "Organizational Accreditation (in Criminalistics)", *Encyclopaedia of Forensic Science*.

Schwartz, A., (1997). "A "Dogma of Empiricism" Revisited: Daubert v. Merrell Dow Pharmaceuticals, Inc. and the Need to Resurrect the Philosophical Insight of Frye v. United States", *Harvard Journal of Law & Technology, Harvard Law School*, 10(2), 149-237.

Schwartz, A. (2005). "A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification", *The Columbia Science and Technology Law Review*, VI, 1-42.

Schweitzer, N. J. and Saks, M. J., (2007). "The CSI Effect: Popular Fiction about Forensic Science Affects the Public's Expectations about Real Forensic Science", *Jurimetrics Journal of Law, Science and Technology*, 47, 357-364.

Scullion, K., (2004). "Wrongful Convictions and the Criminal Conviction Review Process Pursuant to Section 696.1 of the Criminal Code of Canada", *Canadian Journal of Criminology & Criminal Justice*, 46(2), 189-196.

Seale, C., (2004). "Social Research Methods: A Reader", London, Routledge.

Spackman, M. P., Belcher, J. C., Calapp, J. W., and Taylor, A., (2002). "An Analysis of the Effects of Subjective and Objective Instruction Forms on Mock-Juries' Murder/Manslaughter Distinctions", *Law and Human Behavior*, 26(6), 605-623.

Starrs, J. E., (1984). "Miscue in Fingerprint Identification - Causes and Concerns", *Journal of Police Science and Administration*, 12(3), 287-296.

Starrs, J. E., (2005). "The C.S.I. Effect", *Academy News*, 35(2), 32-34.

State of Maryland v. Bryan Rose, (2007). Memorandum Decision, Circuit Court for Baltimore County, Case No.: K06-0545; 19 October 2007, http://www.nlada.org/Defender/forensics/for_lib/Documents/1193671274.47/State%20v%20Bryan%20Rose%20-%20Circuit%20Court%20for%20Baltimore%20County%20(2007)%20(ACE-V%20inadmissible)%20-%20signed.pdf.

State of Maryland v. Bryan Rose, (2008). Memorandum Decision Re Reconsideration, Circuit Court Baltimore County, Case No.: K06-0545; 21 February 2008, http://www.theiai.org/current_affairs/bryand_rose_20080221.pdf; last visited 2008/09/03

Stauffer, E. and Schiffer, B., (2007). "La certification de personnel forensique: Définition, développement et regard critique sur l'expérience américaine", *Revue Internationale de Criminologie et Police Technique et Scientifique (RICPTS)*, LX, 461-478.

Stauffer, E. and Schiffer, B., (in press). "Training and Certification in Criminalistics", *Encyclopaedia of Forensic Science*.

Steele, L. J., (2004). "The Defense Challenge to Fingerprints", Criminal Law Bulletin, 40(3), 213-240.

Stevens, D. J., (2008). "Forensic Science, Wrongful Convictions, and American Prosecutor Discretion", *The Howard Journal*, 47(1), 31-51.

Stewart, A. E., (2005). "Attributions of Responsibility for Motor Vehicle Crashes", *Accident Analysis and Prevention*, 37(4), 681-688.

Sutherland, E. H. and Cressey, D. R., (1970). "Criminology", New York, J.B. Lippincott Company.

Swift, B. and Rutty, G. N., (2003). "Human Ear: Its Role in Forensic Practice", *Journal of Forensic Sciences*, 48(3), 153-160.

Taroni, F. and Aitken, C., (1997). "Forensic Science at Trial", *Jurimetrics Journal of Law, Science and Technology*, 37(3), 327-337.

Thompson, W. C., (1993). "Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the "DNA War"", *The Journal of Criminal Law & Criminology*, 84(1), 22-104.

Thompson, W. C., (1995). "Subjective Interpretation, Laboratory Error and the Value of Forensic DNA Evidence: Three Case Studies", *Genetica*, 96(1-2), 153-168.

Thompson, W. C., and Cole, S. A., (2005). "Forensics: Lessons from the Brandon Mayfield Case." *Champion Magazine*, 4(1), 42.

Thompson, W. C. and Schumann, E. L., (1987). "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy", *Law and Human Behavior*, 11(3), 167-187.

Thompson, W., Taroni, F., and Aitken, C. G., (2003). "How the Probability of a False Positive Affects the Value of DNA Evidence", *Journal of Forensic Sciences*, 48(1), 47-54.

Triplett, M., and Cooney, L., (2006). "The Etiology of ACE-V and its Proper Use: An Exploration of the Relationship between ACE-V and the Scientific Method of Hypothesis Testing", *Journal of Forensic Identification*, 56(3), 345-355.

Tversky, A., and Kahneman, D., (1974). "Judgment under Uncertainty: Heuristics and Biases", *Science*, 185, 1124-1131.

van der Lugt, C., (2001). "Earprint Identification", Gravenhage, Elsevier Bedrijfsinformatie.

van Koppen, P. J., (2008). "Blundering Justice: The Schiedam Park Murder", in Kocsis, R. N., Serial Murder and the Psychology of Violent Crimes, Totowa, New Jersey, Humana, 207-228.

van Koppen, P. J. and Crombag, H. F. M., (2000). "Oren, lippen en vingers: De waarde van oud en nieuw identificatiebewijs", *Nederlands Juristenblad*, 75, 6-12.

Walker, C., (2003). "Miscarriages of Justice in England & Wales", First International Workshop on Wrongful Convictions, 24-26 August, Breil/Brigels, Switzerland.

Wax, S. T. and Schatz, C. J., (2004). "A Multitude of Errors: The Brandon Mayfield Case", Champion Magazine, 28(8), 6.

Wertheim, P. A., (1996). "The Ability Equation", *Journal of Forensic Identification*, 46(2), 149–159.

Wertheim, K., Langenburg, G. M., and Moenssens, A. A., (2006). "A Report of Latent Print Examiner Accuracy during Comparison Training Exercises", *Journal of Forensic Identification*, 56(1), 55-93.

Whiting, P., Rutjes, A. W. S., B, R. J., Glas, A. S., Bossuyt, P. M. M., and Kleijnen, J., (2004). "Sources of Variation and Bias in Studies of Diagnostic Accuracy - A Systematic Review", *Annals of Internal Medicine*, 140(3), 189-202.

Wickenheiser, R. A., (2002). "Trace DNA: A Review, Discussion of Theory, and Application of the Transfer of Trace Quantities of DNA through Skin Contact", *Journal of Forensic Sciences*, 47(3), 442-450.

Zabell, S. L. (2005). "Fingerprint Evidence", *Journal of Law and Policy*, 13(1), 143-179.

Zaitz, L., (2004). "FBI Case Against Oregon Lawyer Built on Blurry Fingerprint, Logic", Oregonlive.com, 2004/05/30, http://www.oregonlive.com/special/terror/index.ssf?/special/oregonian/terror/may040530.html; last visited 2008/09/03.

Zalman, M., (2008). "The Adversary System and Wrongful Conviction", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 71-91.

Zalman, M., "Criminal Justice System Reform and Wrongful Conviction", *Criminal Justice Policy Review*, 17(4), 468-492.

Zellick, G., (2004). "Annual Report 2003-2004", Criminal Case Review Commission, http://www.ccrc.gov.uk/CCRC_Uploads/2003%20-%202004_AnnualReport.pdf; last visited 2008/09/03.

9. Appendices

9.1 Papers published or written

9.1.1 Judicial error and forensic science: pondering the contribution of DNA evidence

Schiffer, B. and Champod, C., (2008). "Judicial Error and Forensic Science: Pondering the Contribution of DNA Evidence", in Huff, R. and Killias, M., Wrongful Conviction International Perspectives on Miscarriages of Justice, Philadelphia, Temple University Press, 33-55.

Abstract

Judicial error and forensic science: pondering the contribution of (DNA) evidence

Beatrice Schiffer and Christophe Champod

Both judiciary and forensic scientists would agree on the contribution of forensic science in sometimes promoting wrongful convictions, but more and more in exonerating the innocents. It is not our intention to question the fundamental role of forensic evidence as a tool for exclusion, but as raised by Edmond (2002), we may not tend to be as critical towards exculpatory evidence as we are towards incriminating evidence. This remark is especially valid in the context of DNA, which the general public, lawyers and scientists alike see as the “new gold standard” (Lynch, 2003). DNA evidence has been used lately in many cases as exculpatory evidence. In this paper we propose to assess the strength of DNA evidence in combination with other forensic evidence in two cases. The first case is that of Michael Shirley, arrested in 1987 for the rape and murder of Linda Cooke. The evidence collected included a blood group test (of semen), a footwear mark, and bruises and scratches on the suspect. The second case is that of Mark Dallagher, arrested and charged for the burglary and murder of 94-year old Dorothy Wood. Based on incriminating evidence involving ear prints, Dallagher was sentenced to life imprisonment in 1998. In both cases sensitive DNA tests, called LCN (Low Copy Number), excluding the defendants played a critical role in the decisions that led to free both of them in 2003. Both analyses show that the strength of the DNA evidence, although extremely useful to the fact finder, should not be considered sufficient in itself to exonerate a defendant.

Introduction

Judicial error or miscarriage of justice (as defined by the International Network of Researchers on Wrongful Conviction, Breil/Brigels 2003) will “invariably identify at least some element of an earlier conviction as a mistake: whether evidential, procedural or material irregularity” (Edmond, 2002, p. 56). In this paper, we will restrict ourselves to examining the contribution of forensic science as a catalyst of either enabling or detecting miscarriage of justice.

New pieces of evidence/facts provide the evidence needed for successful appeal. With the advent of new techniques forensic science can fill this gap. It allows for demonstrating errors without questioning the integrity of the legal system, being an extraneous factor to it. This is because, in recent decades a number of criminal convictions have been reversed on appeal

(Edmond, 2002), partially on the basis of problems associated with the use of scientific evidence adduced by the prosecution during the trial.

We propose to explore the contribution of forensic science in this context using two case examples which we believe cover most of the issues and will serve as a basis for a more general discussion. They both involve an appreciable contribution of forensic science as inculpatory and exculpatory evidence.

Regina v. Dallagher – the case

In May of 1996, a delicate deaf elderly woman was smothered with a pillow at her home in Huddersfield by a burglar who had forced his entry through a transom window. When the scene was examined, ear prints were found on the window pane immediately below the transom window, which had been forced. The inquiry revealed that the windows had been cleaned three or four weeks earlier. The investigators believed that the burglar had listened for the presence of persons inside by putting his ears to the victim's window. As there were no suspects, the police investigated into the movements of well-known burglars. During this process Mark Dallagher, who offered no valid alibi and had supposedly confessed the murder to a jailhouse snitch, was arrested and his ear prints taken. Two ear experts, Mr Cornelis Van der Lugt and Professor Peter Vanezis (Forensic Medicine and Science at the University of Glasgow), concluded to a match, the former qualifying it "unique", the latter corroborating the match in slightly less categorical terms ("very likely"). Based on this evidence, Mark Dallagher was jailed for life in 1998 after his conviction by the Leeds Crown Court. The case received a lot of public attention, the verdict being qualified as a great step forward for forensic science (http://news.bbc.co.uk/1/hi/uk/235721.stm). On appeal, experts for the defence questioned the scientific validity and strength of ear prints as identification evidence (http://www.forensic-evidence.com). The Court of Appeal (*R v. Dallagher* [2003] EWCA Crim 1903, July 25) allowed the admission of earprint evidence. But based on additional information that might have affected the decision of the jury, the conviction was quashed and a new trial was ordered (Anonymous, 2004).

Subsequent DNA tests were carried out on the lifted ear marks, the analysis showed a partial non-matching DNA profile, which was immediately interpreted as excluding Dallagher. The prosecution dropped the case. That causal relationship was trumpeted in the press and among scientific commentators (Fresco, 2004). Mark Dallagher was freed in January 2004 after serving 7 years in prison.

Regina v. Michael Shirley – the case (adapted from Johnson and Williams (2004))

Michael Shirley, an 18-year-old Royal Navy sailor, was arrested in 1987 for the brutal rape and murder of Linda Cooke, killed by her assailant by stamping violently on her head and neck after raping her. One important aspect of the original trial evidence against Shirley was the result of analysis carried out on semen recovered from Linda Cooke's body. As part of common techniques available at that time, blood group analysis was undertaken. Evidence presented to

the court identified a blood group match between semen and Michael Shirley; the same blood group, it was asserted at the time, was shared by 23, 3% of the British adult male population. Other physical evidence consisted in footwear mark with a logo imprinted on the victim's body, which corresponded to the outsole of the shoes found in his possession. In addition, the prosecution mentioned cuts and scratches on his face and body and bloodstains on his trousers.

DNA profiling could not be undertaken at that time, but in 1999, new DNA profiling techniques were carried out. Two years later, in 2001, reference samples from Cooke and Shirley were obtained, “When Cooke’s DNA bands were subtracted from those exhibited by the mixed profile ‘there remained an array of ‘foreign’ DNA bands which did not match either the victim or the appellant’". The Appeal court interpreted these ‘foreign bands’ to provide significant grounds for Shirley’s appeal against sentence and the court quashed his conviction in July 2003.

Points of discussion

From these two cases, we will discuss three main themes:

- The issues associated with assigning a proper weight to identification evidence as exemplified in Dallagher and its bearing on the question of admissibility in court.
- The fact that new technology, such as DNA evidence, is gaining an international status of gold standard especially when the evidence provides support for the defence.
- The complexity of combining various items of scientific evidence when individually they may point in different directions.

We are fully conscious of the limitation of these themes in the grand scheme of intertwined issues between wrongful conviction and forensic science. We acknowledge the significant contribution of laboratory errors, sloppy work, mixed up “probes” and deliberate falsification of results. Our purpose here is elsewhere: How does well-processed forensic science impact on wrongful conviction - as a cause or as a remedy?

These three main topics will be examined while bearing in mind Edmond’s following statement (2002): “Previously safe and reliable convictions are destabilized by the identification of supposedly flawed scientific evidence and bad scientists. In stark contrast, the new evidence tends to be described, in largely idealized terms, as good or reliable science.” And “However, if a critical approach to scientific evidence in the criminal justice system is so important, it might appear curious that so little attention has been systematically dedicated to examining the scientific evidence sustaining acquittals.” We agree that forensic science is largely criticised if having led to a judicial error and is generally not at all “evaluated” if used to establish the innocence of a convict. The Michael Shirley and Mark Dallagher cases enable the illustration of these topics and help a more general discussion.

The strength of the identification evidence

Two factors brought radical changes in the way we approach identification evidence today. First, the rapid and extraordinary development of DNA evidence which, after initial important vigorous and scientific debate in court and scientific literature, has established itself as a mature technique. The strength of the potential association between the DNA profile from a crime stain and the DNA profile from a known person is routinely expressed through the use of a statistical argument related to the selectivity of the DNA in a relevant population. It is fair to say that DNA evidence is becoming rightly or wrongly the new gold standard in forensic science (Lynch, 2003). The second factor is the admissibility of evidence. Although it is more US specific, it impacts all forensic disciplines around the globe. The path making Supreme Court cases interpreting Federal Rule of Evidence (FRE) 702 on admissibility are the Frye case, superseded by the *Daubert* case (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)) and the *Kumho Tire* case (*Kumho Tire Co. V. Carmichael*, 526 U.S. 137 (1999)). These cases pinpointed the twin concerns of "reliability" and "helpfulness". Five factors represent the reliability prong of *Daubert* and can be used by the judge, acting as a gatekeeper, to guide its decision on admissibility. These are, as summarized by the Advisory Committee on FRE 702:

1. Whether the expert's technique or theory can be or has been tested–that is, whether the expert's theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability;
2. Whether the technique or theory has been subject to peer review publication;
3. The known or potential rate of error of the technique or theory when applied;
4. The existence and maintenance of standards and controls; and
5. Whether the technique or theory has been generally accepted in the scientific community.

This combined effect of DNA and *Daubert* led to a regain of interest in the foundations of well-known identification evidence such as handwriting, toolmarks, firearms and, fingerprints (Saks and Koehler, 1991; Saks, 1998; Epstein, 2002; Steele, 2004).

To our knowledge, earprint evidence has not been submitted to any *Daubert* hearing in the United States. However in State v. Kunze (*State v. D.W. Kunze*, Court of Appeals of Washington, Division 2. 97 Wash. App. 832, 988 P.2d 977 (1999)), the Court heard some twenty experts in identification evidence and came to the conclusion that earmark identification was not a field that had gained general acceptance among peers. The Court ruled that earmark evidence could not be accepted as scientific evidence under the *Frye* test.

Critical reviews have recently been published (Champod et al., 2001, Swift and Rutty, 2003; van Koppen and Crombag, 2000). Compared to established identification fields, such as fingerprints

or handwriting comparison, the body of literature pertaining to earmark identification is limited. Scientific research has been done, but it has mainly been devoted to the study of the variability of ear morphology based on the examination of ear photographs. The relevance of this body of knowledge to the case at hand – involving ear impressions found on a window pane – is questionable.

The dangerous tenet of uniqueness

Considerable confusion exists among laymen, and amongst forensic scientists, about the use of a word such as *unique*. The phrase “all ears are unique” is no more than a statement of the obvious – every entity is unique. As the Court of Appeal stated in Kunze in a generalised manner:

> “We agree with and adopt the statements of a commentator who, after noting two generally held tenets–“that no two snowflakes are exactly the same,” and “that no two fingerprints have ever been found to have the same ridge positioning”--states as follows: "In some quarters, these tenets have been scooped up and extended into a single, all- encompassing, generalized principle of uniqueness, which states that “Nature never repeats itself."
>
> This principle is probably true, although it would not seem susceptible of rigorous proof. But the general principle cannot be substituted for a systematic and thorough investigation of a physical evidence category. One may posit that no two snowflakes are alike, but it does not immediately follow that no two shoe soles are alike, since snowflakes are made in clouds and shoes are not. If no two shoe soles are alike, the basis for this uniqueness must rest on other grounds, and those grounds must be identified and enunciated.”

Making parallels between earprints, the unicity of the ear and fingerprint is certainly the best way to give more weight to the association that it really deserves.

Indeed what matters here is how objects can be distinguished. This distinguishability or variability of earmarks depends crucially on the examination method but also on the intrinsic qualities of the marks to display selective features (extensiveness, clarity, etc.). One striking feature in this field is the constant confusion between the variability of objects (e.g. ears, shoesoles) and of the marks left by these objects.

All external ears are different but a high variability between ears does not necessarily imply that a high variability is expressed in marks left by different people. This “clarity bridge” from a complete three-dimensional malleable organ to a two-dimensional mark revealed on a surface, marred by distortions and artefacts, needs to be investigated in much more detail. The same position was recently reached for bitemark examinations. Following a review, the authors stressed its lack of fully established scientific basis (Pretty and Sweet, 2001; Bowers, 1999). Recent observations of close neighbours (Pretty and Turnbull, 2001) (close agreement between

bite marks originating from different sources) have led the authors to call for a renewed interest in researching the replication of dental features on human skin and to raise a cautionary note on claiming individuality in terms of the suspect's teeth rather than the pattern they make on the bitten substrate. The Dallagher case shows analogies with the case against Krone in Arizona where a crucial piece of evidence was a bitemark left on the victim. Krone was freed in April 2002 after serving 10 years in prison. In the case, the Arizona Supreme Court's *en banc (State v. Krone*, 182 Ariz. 319, 897 P.2d 621 (*en banc*, 1995)) decision said that the physical evidence could neither exclude nor include Krone as the perpetrator, and without the bite mark evidence the State had no case (http://www.forensic-evidence.com/site/ID/bitemark_ID.html).

The lack of transparency in the decision process

The protocol used by practitioners to compare earmark(s) and earprints is comparable to the process of well accepted identification fields (e.g. fingerprints). It can be summarised by the following steps (van der Lugt, 2001):

–The earmarks and the earprints are respectively and independently evaluated in order to assess which parts/features are visible and constitute pressure points.

–Pressure of application and rotation of the head will cause differences between the successive prints from the same individual. Hence, an examination of a series of known earprints from one donor taken under various conditions allows the creation of a model of the expected variations caused by pressure and distortion. This analysis will set the tolerances that will be applied during the comparison process.

–The earmark is compared with the earprints using overlays. The examiners looked at agreement in pressure points and measurements. The more stable features being: the crus of the helix; the tragus; and the anti-tragus. They act as anchoring points for the overlay.

Differences in the comparison process are evaluated by the examiners in the light of the tolerances defined by the known effect of pressure and distortion. A decision is made as to whether any difference is significant (hence leading to an exclusion) or can be accounted for (hence leading to a "match").

From the quality and extensiveness of the overlay, a judgement is made as to whether or not the earmark and the earprints share common origin.

The demonstration of the identification is provided either by transparency overlays and using montages made of cut out photographs (mark and print) or using video overlays.

The fundamental weakness is the evaluation process and the value of a match (the concept of "match" between an earmark and an earprint is not precisely defined). It might be surprising that the identification process is described as a matching process – an assessment of the adequacy of superimposition between the mark and the prints – but that the crucial question of the value to be given to a match is never addressed and is left to the examiner's judgement. This fundamental weakness is however not acknowledged in literature. When a match is declared, the assessment of the rarity of the shared features relies solely on the examiner's experience,

whereas in other fields such as DNA it will be based on published scientific data and probabilities.

Confirmation bias

Both prosecution experts referred to the use of ear image and earprint databases to check for the presence of matching candidates, when comparing the earmarks from the crime scene and the earprints obtained from Dallagher. These comparisons have not been carried out before but after the comparison relevant to that case (between the crime scene earmarks and the earprints from Dallagher) had been made. As the examination relies heavily on a subjective assessment of correspondence, it is imperative for any study of that kind that the examiner's judgement is not influenced by initial mind-sets. Most published accounts for experiments of that type suffer from the same strong methodological bias in that examiners knew beforehand that they were comparing prints from different people.

Subjective bias or confirmation bias (Risinger et al., 2002) in identification techniques are a real risk that needs to be mitigated by appropriate working procedures. It is our view that the Dallagher case suffers from such unconscious bias. The recent case of wrong identification by means of fingerprints may suffer from the very same issue (Cole and Loftus, 2004).

What's next?

Earmark to earprint comparison relies at the moment on individual experience and judgement rather than on a structured body of research undertaken following strict scientific guidelines. Such subjective judgements should be monitored in a structured and disciplined environment (i.e. a large-scale program of collaborative studies and proficiency tests). The expertise of the forensic scientist would be demonstrated not by casework experience only, but by a portfolio recording his/her proficiency in a long series of independently conducted proficiency tests.

The need for scientific research in the field of earprint identification has been recognised and a research consortium has been initiated under the umbrella of the European community. The reasons underpinning the research project are detailed (www.fearid.com) as follows: "The need to start this research follows the development of this type of evidence in several European countries and the United States of America. Ear-print evidence in its current state suffers from a lack of consistent scientific research and as a result is vulnerable to challenge in court. The level of challenge is much higher than for fingerprints and DNA, because earprints are less well known and perception has become more sceptical of expert's opinions. Thus there is a need for a clear scientific basis. This basis demands a systematic study of ear-prints and the development of automated support tools for matching." Any fundamental and systematic study of that nature should precede and not follow the use of earmark as "conclusive" evidence in court. Such a cautious approach should apply to all identification fields that have not met the basic criteria under *Daubert*.

However, in Dallagher (*R v. Dallagher* [2003] EWCA Crim 1903, July 25), the Court of Appeal did not share this view nor did the Court embrace a full assessment *à la Daubert*. The earprint evidence was held admissible, leaving the duty of highlighting its limits to the adversarial system itself through a proper *voir dire* or at trial (that decision was confirmed in another recent Appeal case *R. v. Mark J. Kempster* [2003] EWCA Crim 3555).

The risk is that both parties may not have the same access to expertise and technology to make this adversarial battle a fair and informed assessment. Dallagher, no doubt, suffered from such an imbalance in his first trial.

Investigator/Evaluator

Are we saying that such forensic identification techniques have no place within the criminal justice system? Not entirely, but there is a need to distinguish between investigative and probative use. A valuable model that helps scientists focus on their role is called the '*Investigator/Evaluator*' dichotomy. In reality, scientists operate in both *investigator* and *evaluator* modes in many of the cases they examine. Providing opinion in these two different modes requires different mind-sets. In the *investigator* mode, it is the scientist's role to form a reasonable hypothesis from the observations. While attending a crime scene and recovering earmarks, the police may put forward the following investigative questions:

–How many people are involved?

–What potential set of actions may give rise to this (these) mark(s)?

–What is the range of height of the person at the source of that(these) mark(s)?

–Using reference collections or databases, could you suggest one or more potential names to the investigation?

The scientist will form and communicate what may explain the observations based on his knowledge, experience or through the use of databases. Generally, scientists operate in this mode before a suspect is arrested and charged with an offence. Opinions provide directions and options to the investigation and it is accepted that misleading directions may be offered. It should remain obvious that the fact of searching for someone in a database AND finding a match does not automatically incriminate the person. It only represents an *investigative tool* (.e. forensic intelligence) and other evidence will be needed for conviction. Thus, it is feasible to exploit fragmentary and/or contaminated (DNA) traces/marks which themselves could (never/not) be presented as proof in court.

The problem arises when this data is not further scrutinised and used as confirmed evidence in court. In the *evaluator* mode, the scientist's role is to form a view on the weight of evidence to be assigned to scientific findings. This is the primary role of the scientist in what may be called *post-charge* cases; i.e. cases in which a suspect has been arrested and charged. In this role, the concept of weight of evidence associated with the findings should be approached more carefully.

For example, a partial mixed DNA profile (from more than two donors) may be searched against the National DNA database and provide tentative donors for further verification. Obviously, not all of them (and perhaps none) could have left the trace, as they cannot all be responsible for the offence. Only additional police investigation can reveal other evidence in order to narrow the enquiry. In itself, the DNA information can be crucial investigative information. However, considering the current state of knowledge regarding the assessment of complex DNA mixtures at the moment, we may not be in a position to guide the court with robustness as to the weight of evidence to assign for a given retained potential donor.

Applying the reasoning to earprints, we think that they represent a very important investigative tool that should not be ignored, but when it comes to an evaluative use for court purpose such as in Dallagher, we need to be very careful as to the way the evidence is presented to the jury. There is no scientific basis for the categorical conclusions expressed in this case and we would even be very careful at any vague probabilistic expression such as "consistent with" or "probable". Because of the constant (although fallacious) comparison between earprints and fingerprints, even leaving a statement concluding to a match between the mark and the print without further qualifying its significance could be (and has been) potentially misleading. Adequate presentation of such evidence in court should be made using a very strong statistical and objective underpinning. That line of argument has recently been taken by the UK Court of Appeal dealing with facial mapping evidence (*R. v. Paul E. Gray* [2003] EWCA Crim 1001) and we would recommend treating all new forms of identification evidence in the same way, namely:

> "We do not however wish to pass from this appeal without making general observations about the use of facial imaging and mapping expert evidence of a reliable kind. Mr Harrow, like some other facial imaging and mapping experts, said that comparison of the facial characteristics provided "strong support for the identification of the robber as the appellant". No evidence was led of the number of occasions on which any of the six facial characteristics identified by him as "the more unusual and thus individual" were present in the general population, nor as to the frequency of the occurrence in the general population, of combinations of these or any other facial characteristics. Mr Harrow did not suggest that there was any national database of facial characteristics or any accepted mathematical formula, as in the case of fingerprint comparison, from which conclusions as to the probability of occurrence of particular facial characteristics or combinations of facial characteristics could safely be drawn. This court is not aware of the existence of any such database or agreed formula. In their absence any estimate of probabilities and any expression of the degree of support provided by particular facial characteristics or combinations of facial characteristics must be only the subjective opinion of the facial imaging or mapping witness. There is no means of determining objectively whether or not such an opinion is justified. Consequently, unless and until a national database or agreed formula or some other such objective measure is

established, this court doubts whether such opinions should ever be expressed by facial imaging or mapping witnesses. The evidence of such witnesses, including opinion evidence, is of course both admissible and frequently of value to demonstrate to a jury with, if necessary, enhancement techniques afforded by specialist equipment, particular facial characteristics or combinations of such characteristics so as to permit the jury to reach its own conclusion -- see Attorney General's Reference No 2 of 2002 [2002] EWCA Crim 2373; but on the state of the evidence in this case, and if this court's understanding of the current position is correct in other cases too, such evidence should stop there."

DNA evidence – the new gold standard for the defence

As mentioned earlier, more than any other forensic science disciplines DNA and its use for evaluative and intelligence purposes, has been submitted to vigorous legal and scientific scrutiny before its acceptance in court. Nonetheless, its strength does not amount to factual certainty. The process is essentially probabilistic and by its nature, it requires a well-thought and balanced communication both at the investigative and evaluative stage. The following case illustrates the dangers of overweighing DNA evidence during investigation. In 1999 a man was wrongly associated with a crime through a SGM DNA profile match (based on 6 loci). British Raymond Easton, suffering from Parkinsons, was accused of having committed a burglary 300 km away from his home, as six STR (Short Tandem Repeat) matched with a bloodstain recovered by a broken window at the crime scene. The “cold” hit produced a match probability of 1 in 37 million (http://www.crimlaw.org/defbrief123.html). The result was obtained by comparing the crime scene sample to the 700000 DNA profiles of the database, one of them being that of Raymond Easton. Despite his protestations of innocence and alibi evidence, he was arrested and spent several months in jail. It was only on the suspect’s solicitor’s demand that additional array of loci were tested subsequently, leading to the discovery of an exclusion on four loci.

How could this case have gone so far? As we have discussed before, scientific evidence can be used in investigator mode, e.g. search in a DNA database or evaluator mode, proof for the legal instances. These two notions were mixed up and the “DNA cold hit” was rightly used as an investigatory tool, and then regarded by prosecution in isolation and regardless of the other circumstantial elements, as evidence with a disproportionate strength (probably amounting to a moral certainty). Contemporary press releases speak of:

–“computer mismatched DNA evidence”, “DNA mistake.” (http://www.exn.ca/Stories/2000/02/18/04.asp)

–“Parkinson's disease sufferer Mr Easton was arrested on the basis of DNA evidence which later proved false.” (http://www.thisiswiltshire.co.uk/wiltshire/archive/2000/08/15/swindon_news10ZM.html).

These statements are not correct, as the match was genuine, only the interpretation of its meaning was poorly understood and that constitutes the biggest danger associated with DNA

evidence. Today, according to the Association of Chief Police Officers (ACPO) good manual practice (UK), in order to convict/arrest a person based on a hit in the DNA-database it is necessary to have other evidence. The DNA hit in itself is considered insufficient. Indeed, most people would be seduced by the argument that "when a DNA "match" is reported with odds of one in 37 million, we will encounter a like match in the DNA pattern only once in 37 million people." (http://www.forensic-evidence.com/site/EVID/EL_DNAerror.html). Dr. Bramley, Chief Scientist of the Forensic Science Service put it right in offering the following statement: "The number of CJ (suspect/offender) sample profiles on our National DNA Database at the time was about 700,000 and the number of scene of crime sample profiles about 70,000. If none of these were related to one another and we were checking for a match using the 6 SGM loci we would be carrying out something like 50 billion pairwise comparisons and would expect several hundred matches just by chance." (http://www.crimlaw.org/defbrief123.html)

The correct understanding of the value to attribute to DNA evidence is crucial. Most often when misunderstandings arise, as illustrated by the Easton case, the problem can be traced up to an incomplete comprehension or communication of the statistics involved. In addition, DNA also involves other risks, such as the use of non pertinent DNA samples, the (in)correct handling of items of physical evidence, and the neglecting of other scientific/circumstantial evidence to thoroughly investigate a case (Coquoz and Taroni, 2006). Therefore, as pointed out by Penacino et al. (2003): "it is important to underscore that DNA testing should be considered one more piece of evidence within the context of a criminal or forensic investigation".

DNA does free people…

Although these situations have no direct bearing, either on the Dallagher or on the Shirley case, they do help to illustrate our reservations about DNA as the new gold standard and people's high expectations. On the other hand, by raising some critical points, we do by no means attempt to diminish the importance of DNA as exoneration evidence. Indeed, the turning point in both Dallagher's and Shirley's case has been the results from DNA analysis showing in both cases a DNA profile not matching those of the respective defendants. This is one of several ways (a) judicial error can be discovered. When forensic science is employed to reveal judicial error, DNA is involved most of the time.

Recent research data gathered by Gross *et al.* (2005) illustrates this fact by listing the number of exonerations in the USA from 1989 to 2003, the former being the year of the first DNA exoneration in this country. Of 328 liberated persons 145 were freed by DNA, which is over 44%. The authors argue that their data is necessarily biased in several ways and mention the following most important reasons:

- Not all possible cases of wrongful convictions have been retained, as it is difficult to compile data of this type (e.g. pending cases).
- In the 145 DNA exonerations, only offences "producing" physical evidence that may be tested for DNA are included, thus leading to an overrepresentation of rape cases, and less so murder cases (a total of 97% of all cases examined by the authors). 88% of the

rape case exonerations were based on DNA evidence against 20% for the murder exonerations (also including rape charges). Thus, robberies, for instance, do not appear in their compilation, as this offence seldom involves DNA testing.

- The initial samples have to be available for retesting at the moment they are needed by the defence
- Only people convicted for long prison sentences (and it may thus be deduced heavy crimes) will "profit" from exoneration, as the others will be released before the sentence is quashed, the average time from conviction to quashing being over eleven years.

Rather unluckily, Gross et al. (2005) do not mention how many wrongful convictions were based on (faulty) scientific evidence, though they mention that in 24 cases perjury by governmental forensic scientists could be detected. Although Saks et al. (2001) point to erroneous forensic science as one of the causes for wrongful conviction in 53 of the 81 cases, they do not specify precisely what this means (nor which forensic science was involved).This is no surprise to us, as forensic science (to convict and to exonerate) is underrepresented and often wrongly understood in research concerning wrongful convictions.

… but beware of false inferences

Returning to both our cases, Dallagher and Shirley, it may be stated that the DNA results have been portrayed by some commentators (Fresco, 2004, Johnson and Williams, 2004) and the UK national press as demonstrating without any doubt the innocence of the two men. This status of gold standard for DNA evidence is more dominant when the evidence favours the defence case. For instance Saks et al. (2001) conclude that "When a person has been wrongly convicted and imprisoned is later excluded as the perpetrator by DNA typing, that person is thereby exonerated more convincingly than had ever been true of cases of exoneration in the past". They pronounce DNA to be "the first method to expose those errors with virtual certainty". We would like to explain some reservations to such blind reliance on scientific evidence by referring to our case examples. It is very tempting (if not fully logical for any member of the jury) to progress the arguments in the following way in the Shirley case:

- the DNA profile shows a contribution of an unknown male;
- the DNA profile originates from the sperm recovered in intimate samples taken from the victim;
- the semen has been left by the rapist;
- hence Michael Shirley cannot be that man.

This deductive progression is, however, fallacious and we will show that, although the DNA evidence lends some support for the defence position (in both cases), it is not powerful enough to end the debate in such a categorical manner. This is due to the probabilistic nature of the progression of arguments between the demonstration of common (or different) sources and the guidance as to the activities involved at the heart of the case.

The need for transparent communication on the hierarchy of issues

One of the key principles for interpreting scientific evidence is that interpretation of evidence is not possible unless one considers the scientific findings in the light of at least two competing propositions. In most cases in the adversarial system, these will represent respectively the prosecution (*P*p) and defence positions (*P*d). It places an emphasis on a balanced and impartial view of the evidence (Jackson, 2000; Evett et al., 2000). An appropriate assessment of the weight of the evidence is obtained by assessing the relative likelihood of the scientific findings given each proposition.

Managing the elicitation of propositions is without doubt the most difficult aspect of case interpretation. To assist the scientist a 'hierarchy of propositions' (Cook et al., 1998) has been suggested. In summary, the hierarchy has three levels: *source*, *activity* and *offence*. The *offence* level question can be seen as the ultimate matter to be proven. The *activity* or *source* level questions could represent intermediate issues further away from the ultimate issue. The forensic findings can provide support for propositions at any level within that hierarchy. A clear understanding of their position within the hierarchy is a first step towards understanding their potential impact on the ultimate issue.

In the context of the cases against Mark Dallagher (earmark) and Michael Shirley (shoemark and blood group test), the following sets of propositions can be put forward (with a lot of similarities) covering both the mark evidence and biological evidence:

Mark Dallagher – Hierarchy of propositions

Offence level	*P*p: Mark Dallagher killed Dorothy Wood. *P*d: Some other man killed Dorothy Wood	
Activity level	*P*p: Mr Dallagher listened at the window *P*d: Some other man listened at the window	
Source level	*P*p: The collected residue came from Mark Dallagher. *P*d: The collected residue came from some other person.	*P*p: The earmarks originated from the ears of Mark Dallagher. *P*d: The earmarks originated from another unknown person.
Sub-source level	*P*p: The DNA profile obtained is from Mark Dallagher *P*d: The DNA profile obtained is from an unknown person.	

Michael Shirley – Hierarchy of propositions

Offence level	*P*p: Michael Shirley raped and killed Linda Cooke. *P*d: Some other man raped and killed Linda Cooke	
Activity	*P*p: Michael Shirley had the last sexual	*P*p: Michael Shirley trod on the body of

level	intercourse with Linda Cooke. *P*d: Michael Shirley is not the man who had the last sexual intercourse with Linda Cooke.	Linda Cooke. *P*d: Some other man trod on the body of Linda Cooke.
Source level	*P*p: The sperm came from Michael Shirley. *P*d: The sperm came from some other man.	*P*p: The footwear mark originated from the right shoe seized from Michael Shirley. *P*d: The footwear mark originated from another unknown shoe.
Sub-source level	*P*p: The DNA profile obtained is a mixture originating from Michael Shirley and Linda Cooke. *P*d: The DNA profile obtained is a mixture originating from an unknown male and Linda Cooke.	

To progress from *sub-source* to *source* level, we need to consider uncertainties surrounding the relationship between the DNA profile obtained and the underlying body fluid assumed. This will depend on the type of presumptive tests conducted in an attempt to identify the body fluid, their specificity and sensitivity, the microscopic examination carried out, the efficiency of obtaining DNA profiles from targeted fluids, etc. It is our understanding that in both cases, the DNA analysis has been carried out using LCN techniques. The LCN (Low Copy Number) technique, by increasing the number of PCR cycles and changing analytical conditions (reagents), allows to gain in sensitivity and enables profiles to be obtained from only a few cells of starting DNA, as is the case with epidermal contact traces (Gill, 2002). This sensitivity is a new powerful investigative tool allowing the re-investigation of old cases. But when it comes to interpreting its results for court purpose (*evaluative* role), the inference may be obvious at *sub-source* level: the DNA does not originate from the designated persons, but its significance in the context of the alleged actions (*activity* level) need further consideration and guidance. We certainly observe a shift from questions such as ‘whose DNA is this’ to ‘how does that DNA happen to occur in this item’.

This progression is not straightforward and requires expert knowledge. Although we could reasonably argue that the progression from source level issues to activity level issues for mark evidence (such as fingermarks, earmarks or footwear marks) can be left to the court in a process mitigated by advocacy, when it comes to DNA evidence in such limited quantity, the scientist ought to guide the court because of the specialised knowledge he can bring to the debate. Indeed, the interpretation of LCN DNA profiles at *activity* level requires consideration of issues such as the transfer of DNA, the persistence of DNA under different circumstances (Lowe et al., 2002; Wickenheiser, 2002), the possibilities of contamination and so-called drop-in of spurious alleles, and also the determination of the reproducibility of profiles (Gill et al., 2000).

This special knowledge is clearly the domain of expertise of the forensic scientist, and no other. Considering technical aspects, Thompson et al. (2003) argue that, “ignoring or underestimating the potential for a false positive can lead to serious errors of interpretation, particularly when the suspect is identified through a “DNA dragnet” or database search” and “even a small positive probability can, in some circumstances, be highly significant, and therefore […] having accurate estimates [of] the false positive probabilities can be crucial for assessing the value of DNA evidence”. Though they discuss “normal DNA” techniques, this argument, by analogy, equally applies to LCN DNA methods. The scientist will have then to articulate the key elements impacting on the move from *sub-source* to *activity* level. The closer to *offence* level the expert is able to assess the evidence, the more fit-for-purpose the interpretation will be, providing greater added value to the court. In essence, the process is probabilistic. However, it must be recognised that the amount of background information required from the framework of circumstances increases as the expert moves up the hierarchy from *sub-source* to *activity* level. The final progression from *activity* to *offence* level should remain within the remit of the court.

When DNA contradicts other scientific evidence

In the Dallagher case, regardless of the discussion on the weight of the association provided through the examination of the earmarks, the case boils down to impression evidence providing support for the prosecution and DNA evidence pointing in the other direction. Similar results were obtained in the Krone case, when the DNA profile obtained from the swabbed bitemark revealed a DNA profile different from the DNA profile of the defendant. In these cases, the combined effect depends critically upon two factors: (1) the probability of a DNA profile originating from sebaceous secretions left by the donor of the earmarks and (2) the probability of detecting an extraneous DNA profile (present for other "innocent" reasons) from such surfaces. In Dallagher, the first probability is not equal to one (a certainty). Indeed, due to the age of the marks, the limited quantity of DNA left following such contacts, and the detection technique applied to reveal the marks, there is an appreciable chance that the DNA profile from the person who left the marks will not be detected. The second probability is not equal to zero (an impossibility), simply due to the extreme sensitivity of the LCN techniques and the uncertainty surrounding the exposure and preservation of the window. Taking into account both factors, although it is established that the DNA detected is not originating from Mark Dallagher, we cannot simply deduce that he is not the donor of the earmarks. The DNA evidence undoubtedly provides support for the defence proposition at activity level but not to the point to demonstrate that the earmarks have been left by another person.

In the case against Michael Shirley, the fact that a DNA profile has been obtained which does not correspond to the appellant does not mean *de facto* that the sperm heads observed under the microscope did not originate from him. Due to the time elapsed between the preparation of the slides and DNA analysis, no DNA profile was detected from the heads, and the observed profile obtained came from another DNA source. To evaluate these findings further within a framework of propositions at *activity* level, the scientist will need to take into account additional

information on background levels of DNA and issues surrounding the transfer, persistence and recovery of DNA. Addressing propositions at activity level forces more detailed consideration of the nature of the relationships between the propositions and the result. We will have then to take into account the following:

(1)the possibility that intercourse took place with an unknown man some time previous to the offence;

(2)whether or not ejaculation occurred during these acts of intercourse and also the nature of the seminal fluid produced;

(3)the expectations of obtaining respectively the observation of sperm heads, an enzymatic serological result and a DNA partial profile from the recovered fluids;

(4)other sources of extraneous DNA profile either from the DNA contamination dating back to the preparation of the samples or due to today's laboratory process.

A difference here is that the interpretation of the biological results at activity level requires both expert knowledge ((3) and (4)), and circumstantial information ((1) and (2)) that is outside the remit of the scientist. Interpreting such results at a fit-for-purpose level is then a communication challenge as well as a delicate task of finding the appropriate positioning for the scientist. Without any guidance (and likewise for the defence under different circumstances), prosecution tends to question a non-match and to advance alternative explanations for these findings. This happened in the Shirley case, as well as other cases (http://www.prisonerlife.com/articles/articleID=49.cfm). If the forensic expert presents a scientific and objective account of his/her findings, the question will have to be discussed where it properly belongs: a balanced and informed advocacy.

Among the elegant tools to be prepared for such an eventuality, are the Bayesian networks. They combine relevant probabilities according to the laws of probability (Evett et al., 2002) and thus allow the avoidance of the pitfalls of intuition when it comes to probabilistic reasoning.

The combination of conflicting pieces of scientific evidence

The combination of various pieces of scientific evidence can be very challenging. Bayesian Networks are promising tools that enable to handle complex situations. One important observation is that the combination of various pieces of scientific evidence, such as in these cases, cannot occur at *source* level, but has to be done higher in the hierarchy of propositions. If the various items of evidence help, through their respective strength, to address the same set of propositions, then the combination amounts to a multiplication of their respective values (their respective likelihood ratios in the jargon of forensic scientists) taking into account potential dependencies. That requirement for a correspondence in the propositions considered is met in the Dallagher case (at *activity* level) and an overall weight of evidence could be offered. Remember that in this scenario, the knowledge and experience needed to progress from source to activity issues fell within the strict remit of the forensic scientist. Hence, the scientist should be encouraged to present evidence at this level: the court would certainly benefit from a

combined figure instead of two elements pointing in opposite directions. Because of the dominance taken by the technical expertise in this combination, it does not usurp the jury's responsibility and we do not believe that it contradicts the *Adams* decision rejecting Bayesian procedures [*R. v. Dennis J. Adams* (1996) 2 Cr App Rep 467]. Following the interpretation of Lynch and McNally (2003), 'the Court in *Adams* did not discount the appropriateness of the Bayesian method for assisting judgments about *scientific* evidence [...]'.

In the case against Michael Shirley, the matter is more complex. As we have seen before, the assessment of the biological results alone requires technical knowledge and assessments of the circumstances of the case. In addition, the combination of the biological results with the footwear mark result can only be undertaken if the inference is progressed higher in the hierarchy of propositions, towards the *offence* level:
P_pMichael Shirley carried out the offence as alleged.
P_dMichael Shirley is not that man, another man is the offender.

This combination has then to take into account:
–the strength of the relationship between Michael Shirley and the shoe under examination;
–the relevancy to the offence of respectively the last intercourse and the deposition of the mark on the victim's body;
–whether or not the two activities (raping and treading) have been undertaken by one or two offenders.

As noticed before, these elements are related to an assessment of the circumstances of the case and are outside the remit of the scientist. In the absence of other evidence we doubt that the scientist can assist with any additional competence (or added value) to such combination, and think that this task should remain the duty of the fact finder.

Conclusion

We have explored two cases where scientific evidence played a critical role, both in leading to the first conviction, and also to the exoneration of the defendants. The first case (*R. v. Mark Dallagher*) allowed the discussion on the strength of identification evidence, especially when the means of identification rests on new possibilities or new technologies. The analysis of the earmark evidence shows that there is always a risk associated with features presented as unique to convey more strength than they actually deserved. In our opinion this risk is increased:

- when the opinion is an informed subjective opinion;
- when the descriptive and inferential models lack transparency;
- when the risks for confirmation bias have not been mitigated;
- when the field is not underpinned by a strong experimental and research background.

We recommend that these conditions (which are highly correlated with the *Daubert* factors) should be fulfilled before the findings are used as evidence in court, but acknowledge that there is room for investigative use of such techniques. In this case the Court of Appeal (*R v. Dallagher* [2003] EWCA Crim 1903, July 25) adopted a more open gate to the admissibility of the evidence putting responsibility on the adversarial system. And to our surprise, the same Court has been more restrictive regarding facial mapping techniques (*R. v. Paul E. Gray* [2003] EWCA Crim 1001) and obviously treated differently identification evidence that we, as scientists, would have handled in the same fashion.

When DNA evidence comes into play, what mainly makes the difference is the point of view of the observer. Edmond (2002) discusses the “conceptual disparity (or asymmetry) between assessment of scientific evidence used to convict and the scientific evidence used to acquit". This was well illustrated by the perceived contribution of non-matching DNA profiles in the case against Dallagher. We observe a blind reliance on the “new” technique. Although DNA analysis *per se* is widely accepted, it might not be so clear cut in certain situations. For instance, the LCN method is so sensitive that it will become more and more difficult to distinguish DNA evidence from “normal” background DNA present on a crime scene.

The risk of overstating the value of DNA was also present in the Michael Shirley case. Our analysis attempted to expose the complexity of interpreting such a framework of scientific evidence at the appropriate activity level. This task requires expert knowledge as well as specific assessment on key elements arising from the understanding and assessment of the circumstances of the case, to the point that only the framework of circumstances allows attributing the “proper” weight to the non-matching DNA results.

Readers may have noticed that we did not explicitly discuss the scientific evidence leading to the first conviction of Michael Shirley, namely the shoemark and the blood group test. Indeed arguments advanced on the subject of earmarks are, by analogy, applicable to both these items. However, we feel that considering “risk factors” possibly leading to judicial error will help better understand how wrongful convictions could happen in both cases (but especially Michael Shirley’s). Three features are common to both cases, namely the atrocity/severity of the crime committed, the initial lack of a suspect and very poor circumstantial evidence. In our opinion, the need for (any) (scientific) proof to convict outweighed the limitations of the latter. In conclusion, we will cite Edmond (2002, p. 86): “Just as police and forensic scientists are portrayed as routinely oriented toward producing cases sufficiently robust (or tainted) to warrant and produce successful prosecutions, so too, many legal commentators appear equally confident in the unreliability of prosecution forensic science and the reliability of the scientific evidence designed to acquit in the high profile miscarriage of justice cases.”

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

Schiffer, B. and Champod, C., (2007). "The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation", *Forensic Science International*, 167 (2-3), 116-120.

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 fingermarks, 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 twelve fingermarks 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 fingermarks 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 twelve fingerprints before and after having followed specific training in fingermark individualisation. The aim was to see how training/education impacts the analysis of fingermarks. For Test II, were participants given eleven fingermarks 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 things, 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.

Keywords: fingermarks, observational biases, errors, misidentification, experimental study

1. Introduction

Traditional forensic identification evidence - especially fingermark individualisation - though accepted in court for over hundred years, has been challenged lately. Reasons are, among

others, the highly publicised erroneous identification by the FBI in the Mayfield affair (OIG, 2006) or legislative needs such as renewed attention on admissibility criteria (Giannelli, 2002). The misleading influence of observational biases in forensic science (Risinger et al., 2002) is mentioned as a possible explanation for errors in the Mayfield misidentification (OIG, 2006). It has been theorised that the lower the quality of the fingermark, the more demanding and subjective the analysis process will be, and accordingly the more vulnerable to stimuli potentially inducing observational biases (Dror et al., 2005). Those stimuli can be “circular reasoning” (looking for features found in the comparison print on the fingermark), disregard of the “one discrepancy rule” (explaining away discrepancies) and incorrect verification of the results when questioned (OIG, 2006). All of these problems have their origin in the inaccurate following of the fingermark individualisation process ACE-V (Analysis, Comparison, Evaluation and Verification) (Ashbaugh, 1999), especially jeopardising the independence of the four stages. Within this process basically three elements exist. They are:

- the fingermark - the object to be observed
- tools and methods used - the way the observation is made
- the fingermark 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 (2004) showed that significant differences exist already in the beginning of the process. Indeed, the amount of minutiae found by fingermark examiners in contrast to lay persons increased. Enlargement of the fingermark 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 fingermark individualisation have shown that the quality of the fingermark (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 (Dror et al., 2005). Furthermore, a study with five experienced fingermark examiners showed that, when submitted twice the same fingermark in a highly different (and emotionally charged) context three of them changed their initial opinion (Dror et al., 2006). Kerstholt et al. (2006) studied observational biases in shoe print comparison. 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 twelve shoe print examiners. Contrary to their expectations they did not find that any of these influences had 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 fingermark and the protocol to be followed for observation. However, in addition to these, 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 fingermark 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 fingermark, the more demanding and subjective the analysis process will be. According to Tversky and Kahneman (1974) 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, fingermark 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 fingermark 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 fingermark individualisation as advanced by some authors (Risinger et al., 2002; Cole, 2006b).
Several factors were studied. The first was the effect of training on the analysis stage of fingermarks (Test I). Forensic science students were tested before having acquired thorough knowledge in fingermark 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 fingermark and per individual) would increase with training and that the overall variation between individuals for the minutiae counted for a given fingermark would diminish. It was also predicted that the amount of fingermarks considered exploitable (useful for comparison purposes) and identifiable would increase as well with training. Langenburg (2004) showed already the higher efficiency of trained professionals compared to lay persons. We felt it 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 the form of a matching and a non-matching comparison fingerprint were 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 fingermark individualisation.

2.2 Materials and design

Fingermarks from practical student case work were given for analysis to three experienced fingermark examiners in order to determine among others the number of minutiae found. Their findings were used as criterion for selecting fingermarks for the tests. A total of sixteen fingermarks relatively close to the Swiss threshold of twelve minutiae and of varying nature were included in both tests (twelve in Test I and eleven in Test II).

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, which 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 (Figure 6). 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 (terrorism case versus attempted petty burglary). 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.

Figure 6 Illustration of the design for Test II on stimuli for potential observational biases

Group	Reference	Availability of known print			Context		
A	Same fingermark	none	Matching print	none	Non-matching print	High-profile	Low-profile
B		Matching print	none	Non-matching print	none	Low-profile	High-profile
Fingermark	1	2	3	4	5	6-8	9-11

The results to be observed were the same for both Tests I and II. For each fingermark, the response consisted of 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).

 ridge ending (A)  bifurcation (B)  point (P)

Participants were encouraged to differentiate the minutiae marked either by letter (A, B, P and U (for unknown)) and/or by colours. Participants were expected to use the standard procedure of analysing a fingermark learned during their theoretical and practical lessons based on Ashbaugh (1999). The illustrations of the types of minutiae were given in order to avoid misgivings about the definition of a minutia.

3. Results

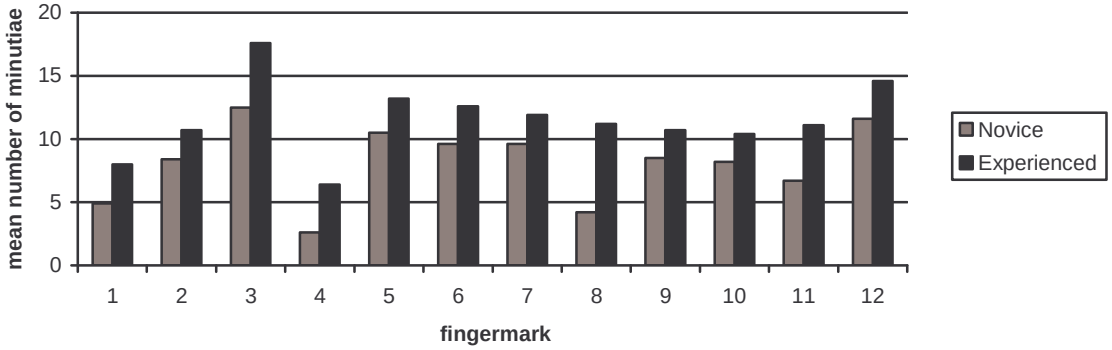
3.1 Test I - Training

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 found 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 experienced. The 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 twelve 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 increases as well for all fingermarks. The same is true for the maximum, excluding however three out of twelve

fingermarks. In general, the range of minutiae found by all participants for the same fingermark decreases after having been trained in fingermark examination (two out of three fingermarks).

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



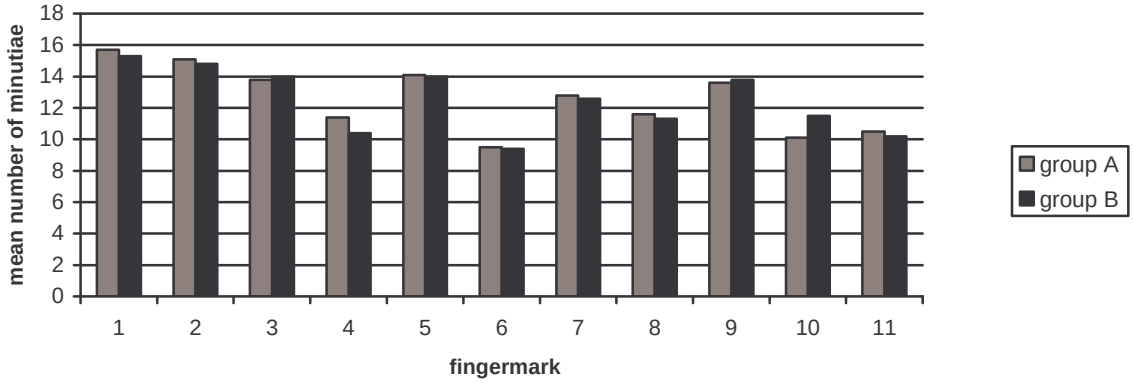
As to the types of minutiae found, it has been observed that over all minutiae marked, the total number of ridge endings passes from 34.2% to 44.1%, a very significant increase. Inversely, the amount of bifurcations decreases from 54.7% to 47.6%. The rest of the change due to points and unknown minutia is negligible in comparison. Concerning the classification of the fingermarks in terms of exploitable and identifiable, it can be observed that the mean number of exploitable nearly doubles from novice to experienced, while it more than doubles 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 following factors: (1) availability of a known comparison print, and (2) differing background case information. For the mean number of minutiae over all fingermarks (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 fingermark 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 ten fingermarks. Considering the presence/absence of a known fingerprint, be it the matching or the non-matching condition, no significant difference can be observed either. The same result applies 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 fingermarks as exploitable or identifiable.

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



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 fingermark 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 fingermark is perceived can be observed as well. However, this effect is limited, as still quite important variations do persist between examiners. This is in accordance with the observations of Evett and Williams (1995). 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 or context information has been observed. This was true for all fingermarks used in the test. These results do, to a certain degree, contradict previous findings or hypotheses, for instance Risinger et al. (2002) and their overview of studies on the detrimental effects of expectation on reasoning and perception. Also the study by Dror et al. (2006), showing with a within-subject design that context information might influence the conclusion drawn from a fingermark 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 (Tversky and Kahneman, 1974). 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. (2006), 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 fingermark examination might be a better way to explain 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.

9.1.3 Is ACE vulnerable to contextual biases?

Abstract
Experimental research on potential error sources in fingermark individualisation tends to show that following a protocol such as ACE-V (Analysis, Comparison, Evaluation, and Verification) can minimise errors (by minimising stimuli inducing observational biases) in the analysis stage (Schiffer and Champod, 2007). Whether this is true for the evaluation stage is not yet clear. The aim was therefore to study the influence of stimuli potentially inducing observational biases, as described in the FBI report on the Mayfield misidentification (OIG, 2006) at the evaluation stage. This includes contextual bias (availability of comparison print) and situational pressure (stress) due to the perceived importance of the case.
The population used for these experiments were students enrolled in the forensic science program of the University of Lausanne (BSc and MSc level). Two factors were evaluated. First, it was tested whether (only) analysing one fingermark (Analysis stage only in a low/no bias context; condition 1) would differ from reanalysing the same fingermark in a ACE process being exposed to considerable pressure created by an assessment situation (ACE and bias due to

situation; condition 2). A possible difference was expected to manifest itself by for example an increase of minutiae from the analysis only to the ACE.

Second, it was tested whether if under considerable pressure students were tempted to "complete" their findings of minutiae on the fingermark by looking for information on the comparison print (circular reasoning). To test this, one half of the student test population (24) were given the Madrid bombing fingermark with the comparison print of Daoud (condition 3), the other half (24) with the comparison print of Mayfield (condition 4). They were expected to follow an ACE process and to pronounce themselves as to the result of their comparison. A possible bias was expected to result in a difference of minutiae observed between the two conditions.

For the first test, results show that an important difference exists between analysing fingermarks as an exercise (condition 1) and doing it during an assessment under pressure and doing the comparison and evaluation as well (condition 2). The quantity of minutiae observed increases considerably. For the second test, it is observed that the amount of minutiae does not vary significantly between the two groups having respectively a matching or a non-matching comparison print (Daoud or Mayfield; condition 3 and 4, respectively). As to the opinions formed, one participant excluded wrongly the matching Daoud print, whereas five participants (out of 24) agreed, tentatively, to a "tends towards individualisation" with the Mayfield comparison print, though no individualisation was pronounced. No verification stage was included in these experiments.

This research project, which includes, among other parts, the experimental tests to study potential observational biases in the analysis, comparison, and evaluation stage of fingerprint individualisation presented here, is financed by the Swiss National Science Foundation (SNSF 100012-105817).

Keywords: fingermarks; observational biases; errors; FBI misidentification; experimental study; Mayfield.

Introduction

As already discussed in previous research (Schiffer and Champod, 2007), traditional forensic identification evidence - especially fingermark individualisation - though accepted in court for over one hundred years, has been challenged lately. Reasons are, among others, legislative changes such as a renewed attention on admissibility criteria (Giannelli, 2002), the highly publicised erroneous identification by the FBI in the Mayfield case (OIG, 2006), or the possibility of wrongful convictions (Cole, 2006b). The misleading influence of observational biases in forensic science (Risinger et al., 2002) is put forward as a possible explanation for errors in the Mayfield misidentification (OIG, 2006). It has been theorised that the lower the quality of the fingermark, the more demanding and subjective the analysis process will be, and accordingly the more vulnerable to external and contextual stimuli potentially inducing observational biases (Dror et al., 2005). Those stimuli can be "circular reasoning" (looking for features found in the comparison print and guess for them on the questioned fingermark) or pressure due to the

nature of the case or the general situation (OIG, 2006). Avoiding a strict adherence to the fingermark individualisation process ACE-V (Analysis, Comparison, Evaluation and Verification) (Ashbaugh, 1999), by not separating the four stages properly for instance, seems to jeopardise the objectivity of the result as well.

Earlier published research on the vulnerability of the ACE-V process to stimuli inducing observational biases tends to show that:

- Experimental tests with lay persons on the influence of bottom-up (trace relevant) and bottom-down (observer relevant) factors on fingermark individualisation have shown that the quality of the fingermark (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 (Dror et al., 2005).
- The analysis stage of the ACE-V process tends to be rather robust against stimuli which may induce observational biases such as a non/matching comparison fingerprint or a high versus a low-profile context (Schiffer and Champod, 2007).
- A study with 5 experienced fingermark examiners showed that, when submitted twice the same fingermark in a highly different (and emotionally charged) context, 4 of 5 participants changed their initial opinion on individualisation (Dror et al., 2006).
- A further study by Dror and Charlton (2006) included 6 fingermark examiners that made each 8 comparisons which they had in the past already examined. Half of them were individualisations, the other exclusions. It was studied whether resubmitting them (without their knowledge) and adding biasing contextual information would change their previous decision. This was the case in 6 out of 48 comparisons, both if contextual biased information was present or not.
- An additional publication (Dror and Rosenthal, 2008), combines the results of the two previously mentioned tests for a further analysis of the data to discuss biasability and consistency.
- A study by Langenburg (personal communication) tested accuracy, precision, reproducibility, repeatability, and biasability of conclusions of 6 fingermark examiners by submitting them to 60 ACE and 60 ACE-V trials. Results tend to show that if the quality of the fingermark decreases, repeatability decreases as well. Furthermore, erroneous exclusions increase if a second examiner verifies the result of a first examiner whereas all erroneous individualisations were detected.

According to Kerstholt et al. (2007) for shoe print comparisons, another area of identification evidence, if a given protocol (having a similar concept than the fingermark ACE-V) is followed expectations (background of the case) and complexity of task (difficulty of the comparison) did not influence the evaluation of simulated case work of twelve shoe print examiners. Experience tended to change the way decisions were justified, but not the results as such.

Furthermore, untrained individuals see fewer minutiae than trained individuals. This is both true when comparing lay persons to experienced examiners (Langenburg, 2004) or when participants were studied before and after having been trained in fingermark comparison

(Schiffer and Champod, 2007). For the last study, the variation of minutiae observed for the same fingermark between individuals decreased when being more experienced, though did not disappear. This concurs with findings of Evett and Williams (1995) for professional examiners.

Known and admitted errors in forensic science – often mediatised in the context of judicial errors – might be (tentatively) classified into three main categories. They will be reviewed quickly in order to clearly position the objective of this research.

First, willing and conscious misconduct (fraud) exists. One example is the case involving Fred Zain (Giannelli, 2001). In fingermark individualisation the forging of a fingermark (Cole, 2005; Geller et al., 1999; Geller et al., 2001) would be an example, as in the murder of Sir Henry Oakes in 1943 (Cole, 2005). However, the focus of the paper is not on this type of error, and they will therefore be excluded from further discussion.

Second, methodological error can happen, which includes the poor choice or performance of a method/technique. In fingermark comparison this would include the incomplete following of the ACE-V protocol or similar established rules (if it is theorised that problems can be mainly solved by training and by following a proper quality assurance program). One example is “circular reasoning” in the Caldwell case as reported by Starrs (1984). The focus of this paper will be mostly on these questions.

Third, interpretational issues can have their impact. This would include supporting an opinion not justifiable by the “analytical” results, for instance overemphasizing the value of observations or “explaining away” differences, as it has been discussed in the Mayfield misidentification in the context of the Madrid bombings, also called the Brandon Mayfield case (OIG, 2006). This is partially studied here.

It can be added that heavy pressure due to the nature of the case - high-profile - and the need for a clear-cut result - a mandatory requirement under the resolution of the IAI (IAI, 1973; IAI, 1980) - or confirmation by forensic science (investigative a priori) of other investigative leads of the police investigation are factors thought to stimulate the occurrence of observational effects/biases at the decision-making stage of fingermark comparison. The Mayfield misidentification seems to be one highly mediatised case where several of these factors seem to be intertwined.

The aim of this research is to study - by recurring, among others, to a reproduction of the fingermark which led to the FBI misidentification - the cumulative effects of the presence of a non/matching comparison fingerprint in the comparison/evaluation stage of fingermark individualisation in a context prone to pressure, namely the concluding end of year assessment of a course devoted to fingerprint examination. Experimental research is used to study more generally potential error sources in fingermark comparison that have been discovered in one case example. Such empirical research should help assess whether or not observational biases can develop into errors compromising the reliability of fingermark individualisation as advanced by some authors (Cole, 2006a; Cole, 2006b; Risinger et al., 2002).

The following factors were studied. For the first test (Test I), it was postulated that context can influence the way a fingermark is observed. As suggested in previous research (Schiffer and Champod, 2007), this influence will show itself in a diverging amount of minutiae observed for the same fingermark submitted under the two conditions. Here, the context factors tested were the presence/absence of a (non/matching) comparison print (and by extension the difference between analysis stage only versus ACE process undertaken in full) and the amount of situational pressure, the variation being the submission during an exercise and an assessment session, respectively. It is expected that when comparing no comparison print and exercise (condition 1) versus (non/matching) comparison print and assessment (condition 2) the amount of minutiae observed will increase.

For the second test (Test II), it was postulated that the ACE process, if correctly implemented by separating the various stages, should minimise the influence of contextual factors. To test this, it was supposed that the presence of a matching (condition 3) or non-matching comparison print (condition 4) (risk of circular reasoning) should not influence the amount of minutiae found. However, it should influence the decision taken as to a correspondence or not, one half tending towards an exclusion, the other half towards an individualisation. The groups submitted to these two conditions were composed of BSc and MSc students, a factor that was not supposed to influence (considerably) the amount of minutiae observed nor the decision taken.

2. Method

2.1 Participants

For Test I, participants were 26 and 29 bachelor students of the School of Criminal Sciences of the University of Lausanne. Condition 1 of the test was submitted during an exercise (t0), condition 2 during an assessment of the course in forensic identification (in the following: assessment) (t1), as described below. The difference (3) in the number of BSc students between the two tests is due to their absence in the first test.

For Test II, participants were 48 forensic science students of the School of Criminal Justice of the University of Lausanne. They were all submitted the test during an assessment. Twenty-nine students were involved in the BSc program with one theoretical course of 28 hours and 56 hours of practicals in fingermark individualisation issues. Nineteen students were registered with the MSc degree with an additional training of 28 hours and practicals in fingermark examination.

2.2 Materials and design

Two fingermarks were selected for this study. Fingermark A was one already used for previous research (Schiffer and Champod, 2007). Fingermark B was the fingermark found in the context of the Madrid bombing incident which then led to a misidentification (OIG, 2006). The comparison prints were those respectively of Mr. Daoud (comparison print BD) and Mr. Mayfield (comparison print BM). Fingermark B was chosen because of the considerable discussion that arose around its individualisation.

For Test I (see Figure 7), a between-group design was chosen. In condition 1 (t0) participants were asked during an un-assessed exercise to mark the minutiae, the type of minutiae, and to describe the quality of the mark in two categories (exploitable: useful for comparison purposes but insufficient for individualisation; and identifiable: sufficient detail for individualisation) (Schiffer and Champod, 2007). This covers the analysis stage according to ACE-V (Ashbaugh, 1999). In condition 2 (t1) participants were submitted the fingermarks during an assessment being unaware that they were participating at the same time in a test. Participants were invited to proceed to a fingermark individualisation. They were expected to follow the ACE-V protocol. They were asked to mark the minutiae and other characteristics found, decide on the quality of the fingermark, and proceed to an evaluation of individualisation or exclusion. Inconclusive is not proposed so as to motivate participants to chose one of the before mentioned categories. Manipulation between the two conditions consisted in the general context exercise versus assessment and availability of the comparison print (yes/no). This design was implemented both for fingermarks A and B.

Figure 7 Summary of the test conditions for Tests I and II

<table>
<tr><td colspan="2">Test I</td></tr>
<tr><td>Condition 1 (at t0) (26 BSc students)</td><td>Condition 2 (at t1) (29 BSc students)</td></tr>
<tr><td>Fingermark A only</td><td>Fingermark A with matching comparison</td></tr>
<tr><td rowspan="2">Fingermark B only</td><td>Fingermark B with matching comparison print (BD)</td></tr>
<tr><td>Fingermark B with non matching comparison print (BM)</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2">Test II</td></tr>
<tr><td colspan="2">Fingermark A with matching comparison used as “internal standard” (for all 48 BSc and MSc students)</td></tr>
<tr><td>Condition 3 (24 BSc and MSc students)</td><td>Condition 4 (24 BSc and MSc students)</td></tr>
<tr><td>Fingermark B with matching comparison print (BD)</td><td>Fingermark B with non matching comparison print (BM)</td></tr>
</table>

t0: Declared collaborative exercise undertaken during the course of the year

t1: Standard assessment of the course on forensic individualisation

The submission during an assessment and without the knowledge of the students (of the purpose of the study) was chosen to emulate pressure in high-profile cases. Materials comprised the four times enhanced fingermarks all printed with a photo-quality Fuji Pictrostrat printer and the same images printed on normal paper for annotation purposes. The results to be observed for each fingermark were the total number of minutiae annotated.

In Test II (see Figure 7), all 48 participants were submitted to a between-group design. Context was manipulated by subjecting one half to condition 3 and the other half to condition 4 for comparing the same fingermark B. Participants with condition 3 had the comparison print BD (Daoud), those with condition 4 had the comparison print BM (Mayfield). The test sheets for both conditions were randomly distributed within the room and participants randomly seated themselves as well. Participants were expected to follow a ACE protocol to justify their decision on the individualisation of fingermark B, by marking the minutiae and concluding on terms used in the course (Bachelor: identification, exploitable, exclusion, based on a 12 minutiae threshold; Master: identification, tends towards identification, inconclusive, tends towards exclusion, exclusion).

Fingermark A was used both in Tests I and II. In Test I it was used under conditions 1 and 2, in Test II it was used as "internal standard" to verify if the groups behaved in the same way, or in other words to see if they were comparable.

2.3 Procedure

For condition 1 of Test I, the fingermarks were submitted during a normal class as an exercise. For condition 2 of Test I and for both conditions 3 and 4 of Test II, the experiment was carried out during a written assessment occurring at the end of the year - 2nd year for BSc (3 years in total) 1st year for MSc (2 years in total). The question included a short description of the fingermarks (detection technique used, substrate, etc.) to allow an analysis of the quality and other relevant issues related to the mark. The task was then described, which was to analyse each given fingermark by annotating each minutiae found, and to assess the potential correspondence fingermark with its comparison print as mentioned before. Participants were expected to use the standard procedure of analysing, comparing, and evaluating marks and prints based on Ashbaugh (1999).

3. Results

3.1 Test I

The aim was to observe no statistically significant difference between the amount of minutiae found under condition 1 and 2. This observation would mean that the difference between both the contextual bias created for condition 2 by the assessment (pressure) and availability of the comparison print has no effect compared to condition 1, the no/low bias situation.

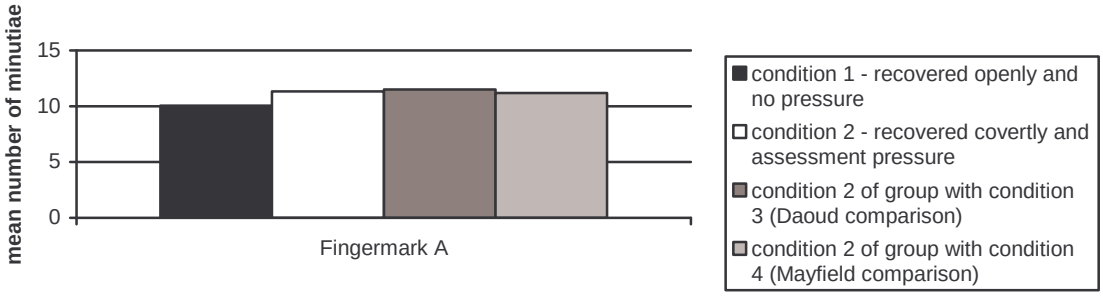
For fingermark A (Figure 8), a total of 26 BSc students found a mean of 10.7 minutiae under condition 1 and a total of 29 BSc students found a mean of minutiae 11.4 under condition 2.

Figure 8 Fingermark A – Values for the minutiae found under condition 1 and 2

Fingermark A	Condition 1 (26)	Condition 2 (29)
mean	10.7	11.4
range	8	5

Minimum of minutiae found	7	9
Maximum of minutiae found	15	14

Graph 1. Variation of the mean number of minutiae annotated by (bachelor) students on fingermark A for condition 1 and condition 2



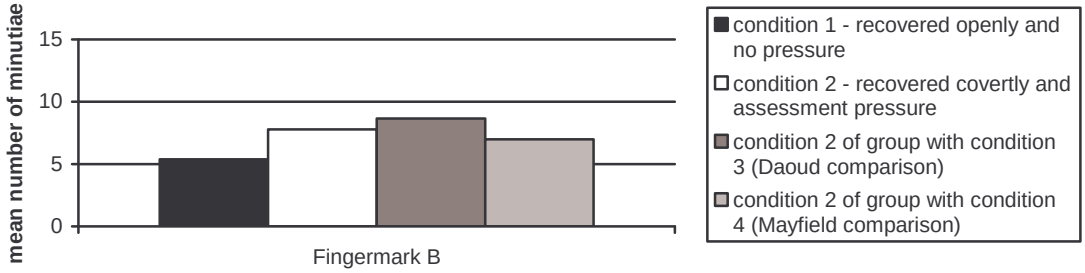
Using a statistical t-test with α=0.05, no difference between the two stages has been found between the means (p-value: 0.174). However, it can be observed that the range of minutiae observed by all participants decreases from 8 to 5 and the minimum of minutiae increases from 7 to 9. Considering all minutiae found by all participants, an increase of 17 minutiae in total or 0.7 minutiae per person in the mean was observed. In short, no (statistically significant) difference between the analysis stage and evaluation stage has been observed, as far as the amount of minutiae observed is concerned. A slight tendency to a greater conformity between all participants can be observed for condition 2, however.

Figure 9 Fingermark B – Values for the minutiae found under condition 1 and 2

Fingermark B	Condition 1 (BSc 26)	Condition 2 (BSc 29)
mean	5.6	7.9
range	7	15
Minimum of minutiae found	2	2
Maximum of minutiae found	9	17

For fingermark B (Figure 9) (the Madrid Bombing fingermark), a total of 26 BSc students found a mean of 5.6 minutiae for condition 1 and 7.9 for condition 2. Using a statistical t-test with α = 0.05, a significant difference between the two stages can be observed (p-value: 0.01). It can furthermore be observed that the range of minutiae annotated by all participants' increases from 7 to 15, the minimum number of minutiae designated stays at two, while the maximum number goes from 9 up to 17. Considering all minutiae found by all participants, an increase of 59 minutiae in total or 2.3 minutiae per person in the mean was observed. Whether participants under condition 2 had a matching or non-matching comparison print did not influence the results.

Graph 2. Variation of the mean number of minutiae annotated by (bachelor) students on fingermark B for condition 1 and condition 2



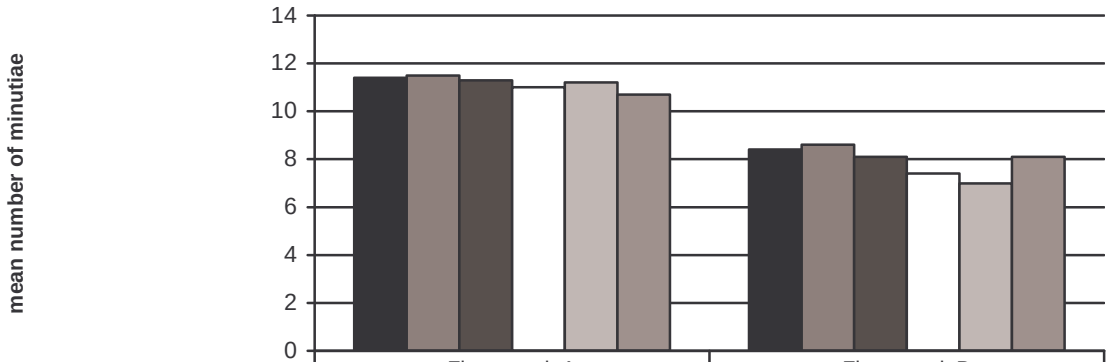
Comparing the results for both fingermarks A and B it can be observed that for both of them an increase of minutiae observed can be seen, though it is only of statistical significance in one of the cases (Madrid Bombing fingermark B). As to the potential influence of the stimuli inducing contextual biases (condition 1 and 2) they seem not to have affected both fingermarks in the same way.

3.2 Test II

Analysis of the fingermarks

If contextual factors have no influence, then the expectation was to observe no differences between the amounts of minutiae found under both conditions 3 and 4. First, the mean number of minutiae found by participants for fingermarks A and B under the conditions of Test II were compared (Graph 3). For fingermark A, the internal standard, participants under condition 3 found an average of 11.4 minutiae, whereas under condition 4 participants found 11.0 minutiae. For fingermark B, participants under condition 3 found an average of 8.4 minutiae, whereas participants under condition 4 found 7.4 minutiae. When applying a t-test at $\alpha = 0.05$ no statistically significant difference between both conditions 3 and 4 can be reported for both the fingermarks studied. The minimal and maximal number of minutiae found for fingermark A and B as well as the range of minutiae found (difference between minimum and maximum) were noted. No difference could be observed for both fingermarks under both abovementioned conditions. Neither does any statistically significant difference exist for the mean number of minutiae observed when the type of education (bachelor versus master) is considered as additional factor (Bachelor and condition 3 versus Bachelor and condition 4; Master and condition 3 versus Master and condition 4; Bachelor and condition 3 versus Master and condition 3; Bachelor and condition 4 versus Master and condition 4.). This has been observed for both fingermarks A and B (Graph 3).

Graph 3. Variation of the mean number of minutiae annotated under condition 3 and 4 on fingermarks A and B

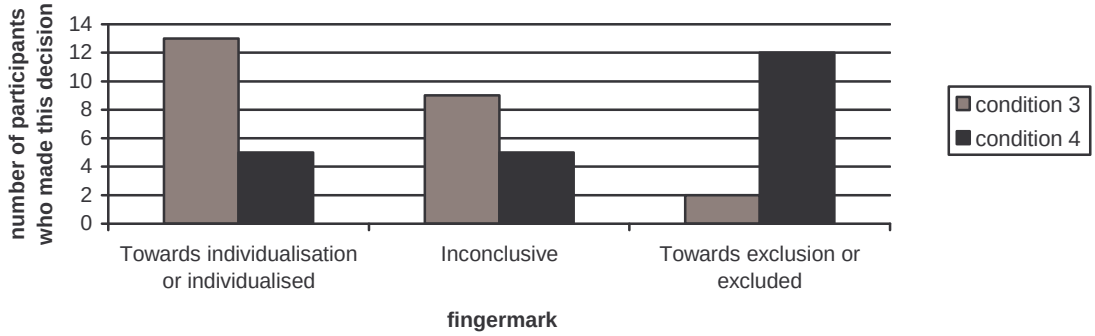


	Fingermark A	Fingermark B
Condition 3 Bsc and MSc	11.4	8.4
Condition 3 BSc	11.5	8.6
Condition 3 MSc	11.3	8.1
Condition 4 Bsc and MSc	11	7.4
Condition 4 BSc	11.2	7
Condition 4 MSc	10.7	8.1

Evaluation stage of fingermark comparison

For fingermark B, several observations can be made (Graph 4). If the results of participants are regrouped in three categories - (towards) individualisation, (towards) exclusion, and inconclusive - for condition 3 there are 14 (towards) individualisation, 1 (towards) exclusion, and 9 inconclusive (including 6 who decided that they could not decide). Under condition 4, 5 (towards) individualisation (very slightly towards), 12 exclusions, and 7 inconclusive (including 3 who decided that they could not decide and 2 participants which did not give a classifiable answer) are found.

Graph 4. Fingermark B: number of participants per condition 3 (match) and 4 (non-match) and their respective choices as to (towards) individualisation, (towards) exclusion, and inconclusive



If the results are reconsidered in terms of wrong and right decisions for both conditions, the following table emerges (Figure 10). Correct result is understood as a statement (towards)

individualisation for the fingermark B, if the reference print was Daoud's (condition 3), and (towards) exclusion, if the reference print was Mayfield's (condition 4). Excluding the Daoud reference print for the Madrid bombing fingermark was also considered incorrect, as well as including Mayfield's comparison for the print.

Figure 10 Correctness of the decision made by the participants according to their condition

	Correct result	Wrong result	Other decision
Condition 3	14	1	9
Condition 4	12	5	7

When testing whether committing an error is related of being part of either group submitted to condition 3 or 4, no statistically significant result can be observed.

Figure 11 Correctness of the decision made by the participants according to their education

Students	Correct result	Wrong result	Other decision
Bachelor	14	1*	14
Master	12	5	2

* Wrong exclusion

When testing whether the type of studies (BSc/MSc) influences the errors made, it can be observed that the exclusion was made by a bachelor student and all wrong towards individualisation by the master students (Figure 11). However, according to statistical correlation tests (Chi2) this difference is not statistically significant, though the result is still of interest. The main difference in training between BSc and MSc students (besides the time) is the way they learn to formulate their conclusions. BSc rely on a numerical threshold whereas MSc are encourage to use a holistic approach as proposed by the IAI (IAI, 1973; IAI, 1980).

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 the stimuli inducing observational biases was observed for fingermark B. However, a less pronounced increase was observed for fingermark A, the result not reaching statistical significance. It might be inferred that the combined presence of pressure (assessment) and availability of the comparison print influences the evaluation stage (reported features in correspondence obtained following the ACE process) of the ACE process, though it does not seem to affect both fingermark comparisons in the same way.

One possible explanation for this observation is that the more difficult the fingermark (bottom-up), the more influence external stimuli (top-down) have on the observation, which would agree with results found by Dror et al. (2005). Fingermark B has been judged difficult by international experts (OIG, 2006), whereas fingermark A presents much less difficulties.
The combination of two stimuli - availability of the comparison print and pressure - does not allow to draw a direct inference as to which of both factors might be more influential. However, results from earlier research (Schiffer and Champod, 2007) with a very similar student population tend to show that the availability of a (non/matching) comparison print in the analysis stage does not influence the mean amount of minutiae observed, nor did the presence of a low-bias context. It might thus be inferred that the situational pressure produced by the assessment could be the more influential factor for increasing the amount of minutiae observed. Results by Langenburg (personal communication) tend to show that if participants are aware of being tested they will be more careful in their evaluation, which will show itself by a more "conservative approach", or an increased tendency to falsely exclude.

According to Edmond (2002) the influence of contextual biases such as external pressure are stronger in high-profile cases. It has been assumed for this research that recovering data covertly while conducting an end of year assessment is comparable to external situational pressure in high-profile cases for both MSc and BSc. Lacking the credits associated with this topic will prevent them from missing any other branch in order to obtain their degrees. Experimental research often collects data openly, a fact that might influence the participants as well as the way they perform. For the research setting used here, the expectation was that the need to do a satisfactory assessment would be stronger than alertness for experimental design features potentially discovered by participants. Under those conditions the influence of the experimental setting can be minimised and real work conditions simulated. Or at least a high-pressure environment was created.
According to these abovementioned criteria, the most vulnerable part in the present test is the analysis stage of data recovery, as this happened openly and on a more voluntary basis, as no assessment took place. Motivation of participants for this test part is more difficult to evaluate and might be inferior to the assessment part. On the other hand, most experimental research relies on voluntary participation and will therefore be closer to the non pressure condition, whereas it is argued here that the assessment condition is rather close to real-life stress. Thus, if conditions of both tests are not judged comparable, they could be to a certain degree accepted as illustrating the difference between experimental and operational situations.

4.2 Test II

For Test II with a total of 48 participants equally divided into two conditions 3 and 4, the mean number of minutiae found was expected not to vary between the two conditions if the stages of ACE were done independently. Results for fingermark B show that though a small difference

can be observed, the variation has no statistical significance (t-test). A difference between condition 3 and 4 can also be observed for fingermark A, though it is even smaller than for B and not significant. This last result was expected, as no difference in condition exists for fingermark A.
These results are in agreement with initial expectations that the availability of the comparison print will not influence the observation of minutiae in the analysis stage (see also results of earlier research; Schiffer and Champod, 2007). It can be inferred that the analysis stage seems to be relatively robust against “circular reasoning” if the proposed ACE process is applied (Kerstholt et al., 2007) in sequence (Krane et al., 2008) and by documenting each step (accountability, MacCoun, 1998). However, according to expectation, being submitted to condition 3 or 4 influences the decision taken as to individualisation/exclusion for fingermark B.

For fingermark B, participants under condition 3 were expected to conclude on a (towards) individualisation with the comparison print, whereas participants under condition 4 were expected to conclude (towards) exclusion for the comparison print as possible source for the fingermark given. Unexpectedly, one participant under condition 3 excluded the Daoud comparison print as potential source for the Madrid bombing fingermark and five participants under condition 4 concluded towards individualisation for this same fingermark and the Mayfield comparison print. In addition, the number of undecided participants which did not give a clear answer is more important under condition 4 than 3. Though more unexpected decisions were taken in condition 4 (which included having the wrong comparison print), statistically speaking no correlation can be observed between the amount of unexpected results in condition 4 rather than condition 3. In other words, no bias could be measured.
One possible explanation for these non-biased results can be attributed to the small number of participants. This argument is, however, somewhat weakened as quite a different constellation of results and/or participants is needed to reach a statistically significant result. Furthermore, most existing experimental research interested in stimuli potentially inducing bias, especially in the forensic science field, is confronted with the same difficulty of limited participant numbers and research and pronounces themselves with even smaller research populations, though the design considers within-subject variation (Dror and Rosenthal, 2008). It would be too easy to dismiss the results because of their (perceived) small amount of participants. To have five participants tending to conclude wrongly - though none of them actually identified the Madrid bombing fingermark with the Mayfield comparison print - is not negligible, however. To have one wrong false exclusion is not enough to say much, but in comparison to the other type of error, false positive, the opinion presented is stronger. This agrees with findings of Langenburg (personal communication), one explanation being that the consequences of the error are perceived differently depending on its type.

Three propositions are made here. First, that if ACE is followed correctly (including extensive annotations and explanations as to the findings made), it can minimise misidentifications as not

very experienced participants were more prudent in their decision than more experienced professionals (OIG, 2006). This agrees with experimental studies in other areas which tend to stipulate that telling people beforehand that they will have to account for their decision diminishes cognitive biases (MacCoun, 1998). This presupposes, however, that the difference between the students and the professionals lies in the way the ACE protocol was followed for the same fingermark individualisation and not in other factors potentially having an influence.

Second, when studying the results more closely, it can be observed that all the five participants that decided on “tends towards individualisation” for condition 4 followed a MSc training. The main difference between MSc and BSc students is the way they learn to conclude when evaluating a fingermark comparison, the first being trained to use a holistic model whereas the second are trained to use a threshold based on 12 minutiae. In this specific context to consider a minimal threshold of minutiae proved to be a safeguard against bias when forming an opinion on a particularly difficult fingermark. It could be worth remembering that the quality of a fingermark is not fully independent of the amount of minutiae found. A “quality threshold” could be considered in difficult cases.

Third, in fingermarks that seem difficult, it could be interesting to have several examiners annotate independently the amount and location of minutiae found. If disagreement is very important, then perhaps the quality of the fingermark is insufficient to reach individualisation on a defendable basis? As advanced by Cole (2006b), without the Daoud individualisation the Mayfield misidentification would perhaps not have been discovered (so soon) nor published. As in all cases (including the Madrid bombing fingermark) the ground truth can not be known, it might be more correct to state that a fingermark is too poor in quality for comparison rather than to arrive to a Shirley McKie situation (Cole, 2006b) where two mutually exclusive hypotheses are maintained. It might be feasible to agree on the quality of a mark by deciding some criteria (e.g. variation between examiners) while to establish the ground truth is very difficult, if not impossible.

The aim here is not to discuss the holistic versus the threshold approach for individualisation, but rather to state that the evaluation stage seems to be the most difficult stage of the ACE process and the most vulnerable to external stimuli, as it is here that differences seem to be most visible. Finally, it must also be mentioned that the verification stage of the process has not been done in the present research and that - if properly done – the amount of unexpected results might have decreased further, at least for the wrong individualisations (Langenburg, personal, communication).

5. Conclusion

In short, whereas the analysis (and comparison) stage seem to be rather robust against external influences (see for instance fingermark A) the comparison and evaluation stage seem to be more sensitive spots in this process. This agrees with previous findings of decisional stages being more vulnerable to external stimuli (top-down) inducing bias than more observational

stages (Dror et al., 2005). The (perceived) difficulty of the fingermark (bottom-up) seems to have a certain influence as well, as observed in Test I.

If these results are, tentatively, translated into practical recommendations, the authors would stipulate that:

To *apply* and *document* ACE-V properly should minimise the influence of confirmatory biases and help checking the work done. This means performing the four stages in their order and independently, documenting at each moment what is done. Krane et al. (2008) call this in the context of forensic DNA interpretation “sequential unmasking”. Documenting also means explain any changes that are made, for example, finding additional minutiae on the mark after consulting the comparison print. Justifying such changes obliges to be very clear why they seem acceptable and thus enhances the credibility of the work done. If no justification can be found, then this could be a sign to be careful and to take appropriate steps, as could be stipulated in a written protocol for difficult cases. Documentation, including justifications and explanations, is easier checked for errors, thus helping to gain transparency and also help to solve problems of inconsistency more quickly. Furthermore, they are basic requirements of any quality management system.

A numerical threshold could be a useful “quality benchmark” against which to test ones non/numerical decisions. If for instance a fingermark is individualised with a given set of observed features including (only) five minutiae while the threshold can be assumed to be at 12 or 16 (according to historically used values), particular care should be given to explain why the features observed seem to allow to make this strong conclusion. This means accepting that a numerical threshold can be used as an empirical “thumbs-rule”, even if no statistical research exists to use it, which may help considering issues of quality and difficulty of a fingermark (perhaps even before comparing it to a print, but just to judge it “aptitude” for the comparison process). A further way to use the counting of minutiae is to find a way of measuring the difficulty of a fingermark. If several examiners independently find a very different amount of minutiae, and probably also define their minutiae differently, then this could be an alarm sign and a reason for not undertaking any comparison, judging the quality “too bad”, and/or the difficulty “too high”.

The authors would also caution against inferring any error rates from this data, as the aim was to study specific phenomena linked to the ACE fingermark comparison process and not to evaluate a general error rate. Haber and Haber (2008a) have given an extensive, and widely discussed (Champod, 2008; Cole, 2008b; Haber and Haber, 2008b; Koehler, 2008a; Mnookin, 2008) overview as to what should be tested how and why. Perhaps the results here partially contribute to this aim, but the focus was on delimitating more clearly risk areas such as difficulty of the fingermark and the two main stimuli studied – non/matching comparison print and non/stressful situation, and to be able to propose practical recommendations.

To sum up, difficulties exist they are present where they may intuitively be expected and they also seem to correspond to the intuitive position of a critical fingermark examiner “they happen,

but not very often". The study of several specific factors allowed to gain a better insight of potential and real risk areas and to propose practical recommendations to continue increasing the quality of fingermark comparisons.

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

9.1.4 Forensic science and judicial errors: comparing the point of view resulting from a literature review with the opinion of interviewed practitioners as to error sources and remedies

Beatrice Schiffer, Christophe Champod

Introduction

For some time now forensic science is a recurrent topic emerging in the discussions about judicial errors. Generally, it is viewed under two completely different roles. It is either presented as a further error source, such as eyewitness testimony (Saks and Koehler, 2005), or as a possible tool for exoneration, best known by the success of DNA analysis in the Innocence Project (http://www.innocenceproject.org/; last visited 2008/08/06). While the underlying forensic science techniques or methodologies are exactly the same, the *use* made of them seems to affect the way they are perceived (Edmond, 2002; Schiffer and Champod, 2008).

One reason for this could be that most publications are written by legal or social science scholars interested in the criminal justice system (in the following: CJS scholars). Publications by forensic scientists, those implementing these techniques are, in comparison, underrepresented. One example of this contradictory view are the publications by Saks and Koehler (2005), Harmon et al. (2006), and Collins and Jarvis (2008). The first advocate that forensic science is an important contributor to judicial error, the second contest this view, and the third state, based on the same data, that this contribution of forensic science has been exaggerated. A lack of information on the practitioners' point of view is also visible when trying to get an overview as to possible error sources in forensic science that may lead to judicial errors. It is argued here, that as forensic scientists are among those potentially involved in creating errors, they also seem to be best placed to minimise or avoid them. Their input therefore seems crucial. Opinions presented by CJS scholars, as defined previously, are certainly interesting, but they might, unwillingly, misrepresent forensic science and its errors. This is why information gathered here by directly interviewing forensic scientists is compared to views found in literature. First, a short overview of literature considered will be given, and then the methodology used for the interviews will be presented briefly.

Subjects covered in literature include, among others, discussions on error sources in forensic science in general (Bryson, 1985; Cooley, 2003a; Cooley, 2003b; Cooley, 2004; Edmond, 2000; Edmond, 2002; Freckelton and Selby, 2002; Gans and Urbas, 2002; Giannelli, 2001; Giannelli, 2002; Jakobs and Sprangers, 2000; Johnson and Williams, 2004; Leahy, 2001; Loftus and Cole, 2004; Maguire, 1993; Risinger et al., 2002; Saks, 1989; Schwartz, 2005), linked for example to fingermarks (Cole, 1998; Cole, 2005; Cole, 2006a; Cole 2006b; Cole, 2008; Dror and Charlton, 2005; Dror et al., 2006; Dror et al., 2005; Epstein, 2002; Mnookin, 2004; Mnookin, 2001; Saks and Koehler, 2005), or remedies (Forst, 2003; Koppl, 2005; Nobles and Schiff, 2001; Saks et al., 2001). They relate to the subject of forensic science and judicial error, as it is assumed that forensic science error, if unchallenged, can lead to a judicial error. Thus forensic science as such – its quality – is already of interest, regardless of the use made of it, for example as evidence in court.

The semi-structured interviews covered topics such as the awareness of forensic scientists as to their field being a source of judicial error, an inventory of “narrow escapes”, reasons for the discovery of those “narrow escapes” and errors, error explanation and avoidance mechanisms, and possible solutions. Participants were a total of eleven forensic scientists, most of them in charge of a forensic laboratory or a forensic science unit, some of them retired. Two of them were approached in the developing stage of the interviews; the others were contacted later. Criteria for their selection were their position/function as a former/current forensic science “manager”, an interest for questions linked to the quality of forensic science, their willingness to be interviewed, and their faculty to express themselves fluently either in (Swiss)German, French or English.
For the Swiss part of the sample (four participants), the selection tried to consider representatives of forensic units of different size, language, and geographical position (Rivola, 2004). For the international part of the sample, European countries with various forensic science systems such as the UK, France, Sweden, and Finland were considered. The Australian participant was included to have an outlook at another, more global level.
Concerning practical issues, the interviews were taped, transcribed, analysed, compared, and evaluated. Questions were framed in a semi-structured way, as proposed by Blanchet and Gotman (2001) and Kaufmann (2001). They covered the abovementioned fields of interest and were initiated by triggering questions in one of the three languages mentioned above. An example would be “if you hear ‘miscarriage of justice’, what comes spontaneously to your mind?”

The most striking difference in opinion between published and orally obtained beliefs can be found in subjects that are of fundamental interest in the debate of forensic science as a source of judicial error. Those that will be discussed here include:

- Where do most errors happen?
- What are the possible solutions?

The interest of these questions will be first introduced shortly. Then, an overview of the opinion prevalent in literature is given. After that, the view of the interviewed forensic science practitioners will be presented and a comparison of both made. Possible explanations for similarities and differences will then be discussed, including also the lessons to learn from it.

Where do most errors happen?

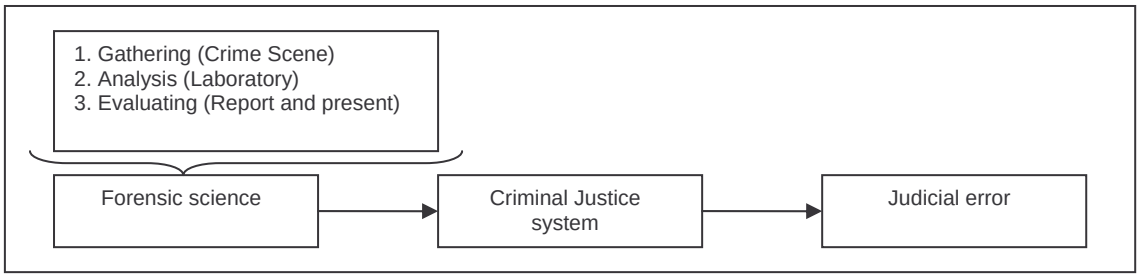
In a simplified model, the forensic science process from crime scene to court can be divided into three main stages: gathering, analysing, and evaluating of traces obtained from this process (see Figure 12). The first stage includes all work done at a crime scene, for instance collecting fingermarks, DNA, shoe wear marks, taking pictures, etc. Securing traces like fibres, glass fragments, gunshot residues on a victim or a suspect is also part of it. The second stage concerns all analysis of traces collected, including traditional laboratory work, for instance enhancing fingermarks or analysing potential biological specimens for DNA profiles. Here all analysis of items will be considered as laboratory work, including fingermark individualisation or other identification evidence, as the focus is on the analysis of specimens and not on the type of method used. The third stage consists in evaluating and presenting the results obtained from work done previously. Examples would be the assessment of matching DNA profiles, the evaluation of the significance of particular evidence in specific scenarios, etc. The distinction of these three stages is usually less clear-cut in everyday work than presented here, but such a broad brushed model is enough for the present discussion.

Error lists as reported in literature (see previously mentioned publications), as far as forensic science is concerned, include two main categories, voluntary errors such as outright fraud or non-voluntary errors such as inadvertently mislabelling specimens. Examples for the first would be Fred Zain or Pamela Fish who willingly produced results matching the preconceptions of their “clients” asking for the examinations (Giannelli, 2007). Examples for the second include overestimating a presumptive result such as in the Maguire affair (Maguire, 1993), the non-finding of a trace in the Damilola Taylor case (Rawley and Caddy, 2007), and the misidentification of a fingermark in the Brandon Mayfield (OIG, 2006) or Stephan Cowans (Cole, 2005) case. The focus here will be on the second category.

Interviewed practitioners name the same kind of errors, naming for instance the abovementioned or similar cases to illustrate their point. What is neither covered in literature nor could be obtained by the interviews is an estimation as to a general (judicial error) or more specific (forensic science) error rate, as attempted by Harmon (2001), Huff (2008), Naughton (2003), Poveda (2001), and Risinger (2007). This topic has also gained some interest in the context of admissibility questions such as raised by Daubert/Kumho, leading to a growing literature on the subject (Haber and Haber, 2008; Risinger et al., 2002; Schwartz, 1997; Schwartz, 2005). While the list of possible types of errors resulted in a similar list for the literature review and practitioners, a marked difference was observed as to the estimated

incidence and impact of such errors. If considering the simplified model of the three stages – crime scene, laboratory, or report writing – where the error risk is perceived to be highest varies between the two groups. According to error sources as represented in literature, the work in the laboratory seems to be most vulnerable (Bromwich, 2007; Saks and Koehler, 2005). In literature, no direct ranking of errors exists, nor is there any quantification. However, when compiling which error sources are mentioned in case reports/reviews, etc., a certain hierarchy emerges. Thus, work in laboratory seems to be most vulnerable to error.

Figure 12 Forensic science and judicial error – a simplified model



Forensic science is gathered, analysed, and evaluated. In this last stage it is then presented at trial and introduced into the legal part of the process. If anything went wrong, either at the forensic science or justice system stage, a judicial error can result in consequence.

The second most critical stage seems to be the evaluation and presentation of (scientific) evidence in court (Saks and Koehler, 2005). Again, based on various sources such as case reviews, etc., there seem to be difficulties in correctly evaluating and presenting the value of scientific evidence (see the aforementioned Maguire case (Maguire, 1993), or more recently Jill Dando (Barry George v. Regina, 2007)). The interest in this area is also visible in a certain amount of publications on how to present evidence, the possibilities of (mis)understanding statistical and probabilistic data presentation, etc. (Taroni and Aitken, 1997; Thompson and Schumann, 1987).

The least represented stage in literature is the initial collection of traces on a crime scene or a person (victim or suspect) even though there are some recorded instances of complete falsification of evidence. Examples are autopsies (Giannelli, 2007) or falsification/creation of fingermarks (Geller et al., 1999; Geller et al., 2001). However, the quality of the crime scene work done is seldom discussed as such in literature. There have been instances where its quality has been questioned, most times together with other parts of an investigation, though this has probably not often been discussed in an academic context. One example is the reanalysis of the Hariri crime scene by other teams than the initial Lebanese forensic team (see the first report of the United Nations International Independent Investigation Commission to the United Nations Security Council (Mehlis, 2005)), one of the reasons probably being the political context of this case.

There are some possible explanations why according to the literature the laboratory is criticised most, followed by the evaluation stage, and then the crime scene work. Feasibility of re-evaluation might be a criterion which may depend on the quality of the documentation and the availability of evidence for retesting/reanalysing (Schiffer and Champod, 2008). Laboratory work is most often documented and, depending on the legislation, specimens may still exist and can be reanalysed. Thus, it is possible to form a second opinion that could contradict the first. The same could apply for the evaluation stage that will, in its turn, also depend on the documentation available and can be questioned in the same way. As to crime scene work, it can usually not be redone (with some exceptions) and is more difficult to document, as every crime scene is different and standard procedures are not easy to implement. Therefore, it might be more difficult to have a second opinion, this factor being a possible explanation as to why it has been less criticised in literature. DNA can be used to illustrate the difference between laboratory and crime scene work. In the first case, a DNA specimen will be analysed in accordance with established procedures. In the second case, to search for and to collect a DNA specimen is not a process that can be easily standardised, though some basic rules can be fixed. One of the reasons is that it is not possible to know what is really available at a crime scene. If nothing is found, it can be either due to initial absence of traces or due to incorrect work performed. Even if retesting a specimen by a second laboratory might find a trace missed by the first one (Rawley and Caddy, 2007), it is difficult – if not impossible – to decide what exactly the reason for the initial non-finding was. Furthermore, there might be a certain tendency to forget that a trace starts to exist the moment it is found. Before that, it is just an object in a given environment.

Contrary to the opinion resulting from the literature review, the interview partners see the most risky area in the crime scene work, followed by the evaluation of the results, and the analysis of the traces in the laboratory. This view is expressed across nationalities and different types of institutions, independently, whether the practitioners interviewed are more involved in laboratory or both in laboratory and crime scene work.

Reasons given as to why the gathering of traces is seen as the most critical moment includes the following observations mentioned by practitioners. Whatever is forgotten on the crime scene, contaminated or modified can not (or very rarely) be recovered or redone. Crime scene work defines the material that will be available for all further work and thus shapes the initial input. Another reason given is that not that many efforts have been made into putting in place quality management procedures for crime scene work (in contrast to laboratory for instance).

Practitioners place the evaluation and presentation of results next in their judgment of risk areas. Reasons given for this view is the subjective and sensitive nature of the task, becoming visible for instance when trying to evaluate the results obtained. The possibility of over- or underestimating the value of the data is seen as one of the dangers. A further element mentioned that adds to the difficulty is the lack of information on the case, which is perceived as increasing the risk of making mistakes in the interpretation of scientific evidence.

Laboratory work is considered the least critical stage by the forensic scientists interviewed. They explain this with the existing quality management procedures in place, the possibility of having well-qualified people working in the laboratory, and the more scientific nature of this work as compared for instance to crime scene work. Participants tend to say that most efforts in increasing the quality of forensic science work have been made in the laboratory portion, the result being that it is now the "safest" of the three stages. As illustration of this the efforts of regulation made by ENFSI and its subcommittees are mentioned.
A further possible explanation for this view could be that a certain bias exists in the perception of the participants, as most of them are head of a laboratory and therefore are better informed about the efforts made in their field and perhaps also more prone to find the mistakes in other fields they are not concerned with. However, this argument would certainly also apply to the literature review by the CJS scholars.

When comparing the opinions resulting from the literature review and forensic science practitioners' interviews, one startling difference is where they see the highest risk area for errors in forensic science. According to literature it is the laboratory, whereas practitioners name the crime scene. One possible explanation for this divergent view is that their perception is based on the information accessible to them. Thus, legal scholars form their opinion mainly by the analysis of discovered judicial errors (as already mentioned focusing on laboratory errors) and the few studies on error rates which concern mostly laboratory work (this includes, in the definition given here, areas such as the individualisation of fingermarks and other "identification" evidence). For a general discussion of limitations of such an approach see Leo (2005). Both of these sources have their limitations. Case reviews and reports represent only an imperfect sample, as they depend on the fact that the forensic error is detected. Discovery is linked to several preconditions, such as the availability of specimens and the possibility of retesting them, for example, with a newer method. This can be illustrated by the reanalysis of biological traces by (new) DNA profile analysis. Studies on error rates in forensic science have their limitations as well. This includes the fact that they are often based on proficiency tests that are not done blindly, where the difficulty of the test has not been adequately calibrated, where it is not possible to know how serious the test has been taken by participants, and how well it is adapted to test what it is supposed to test. Haber and Haber (2008a) describe these limitations for the ACE-V methodology in fingermark comparison. On the other hand, forensic scientists working mainly in laboratories are the direct "clients" of material gathered at a crime scene and are therefore well placed to be aware of its qualities and imperfections, which will influence the work they will be able to perform.
The difference in perception of error risks at the crime scene and laboratory stage is perhaps also a reflection of the (non) familiarity of CJS scholars with the work performed at the crime scene. Why this is so, besides the tentative explanations advanced, could be an important question to answer in order to find an agreement on possible remedies to propose. In the same way, it is interesting to see why both groups agree on the evaluation stage as the second

dangerous stage. Following the same reasoning, perhaps because they are both involved at this stage, one part as client, the other as provider. The question of whether these differences are also visible when proposing solutions to minimise error risks in forensic science will be discussed next.

What are possible solutions proposed?

When discussing forensic science as a possible source of judicial error, one of the main interests is finding solutions to minimise the occurrence of such cases. In the same way as observed for error sources, differences of opinion are visible between the literature review and the results of the interviews with forensic science practitioners.

In literature, several propositions have been made that can be summarised into some general categories. One suggestion is to separate physically, financially, and institutionally forensic science laboratories from law enforcement (police forces and/or prosecutor offices, depending on the structure of the CJS). The aim is to reinforce independence and minimise undue influence of case data, other contaminating information or institutional pressure (Koppl, 2005; Saks et al., 2001). In this context a strict separation of the analysis and the evaluation/interpretation of the results have been proposed as well (Jonakait, 1991). The division is therefore not only between various entities, for example laboratory and police, but also between various stages of the forensic science work, such as analysis and interpretation of subsequent results.

A second suggestion is to increase the quality of work performed in the laboratories. Several incentives and helps for this are proposed. One is to augment available resources regarding finance and qualified personnel (Forst, 2003). A further possibility is to reinforce quality management measures, for instance by rendering accreditation compulsory. Still another incentive is to create competition between various laboratories, thus using “market forces” to render them fitter and more able to provide quality work (Koppl, 2005).

On a more academic level (or in the context of admissibility questions in the USA), the need for fundamental research with regard to the underlying paradigms in forensic science, as well as a (scientific) discussion of error rates seems to be another lacuna to be filled (Crispino, 2006). For instance, in the domain of fingerprints it is postulated that no research exists to prove that two different fingermarks can be differentiated or that a fingermark can be linked to a fingerprint, since the scarce statistical models consider only part of the question. An example of ongoing discussions can be found in Champod (2008), Cole (2008a), Haber and Haber (2008a; 2008b), Koehler (2008a), and Mnookin (2008).

In contrast to the summary given above, interview partners argue that the more separation exists between the different forensic stages - but to a certain degree also with other partners (e.g. toxicological laboratories, police investigation) – the higher the risk of miscommunication is, and, in consequence, the more the risk of errors increases. They therefore tend to have a critical view of the proposed separation of the different stages. They sometimes even propose a contrary movement. One participant proposed to create the function of a “forensic coordinator”,

who would have the main task of making sure that communication between all partners is assured and takes place in the best available conditions.˙ Depending on the structures in place in each country, these recommendations seem to be more or less adapted or necessary. For instance in Switzerland, due to its relative small size and federal structure, interaction between the different stages and (scientific) partners of the CJS is stronger than in France with its centralised and more hierarchically divided police forces, though there is still scope for improvement.

Concerning the accreditation of laboratories and similar quality management measures, most participants think that on the laboratory level a lot of progress has been made lately and that this is sufficient for now. Furthermore, they do not all agree that such measures taken by them alone will increase the quality of the work in the long run. Some participants feel that efforts should now be put on the crime scene work level to improve it according to existing or possible quality management standards to reach a similar level as in laboratory work. However, it can not be excluded that for this last point a certain bias with regard to their field of work and interest of the interview partners plays a role, since at least half of them are not and have not been directly involved in crime scene work.

Several participants think that a solid and comprehensive education in forensic science, with a strong scientific background is certainly helpful to guarantee work quality. They feel that to increase the educational level and training of forensic scientists is a helpful safeguard against errors at all stages. Some interview partners even feel the need to install specific courses on ethics and related issues, such as error risks and error management.

Some participants feel that additional research in forensic science is needed on a variety of aspects, but it is not the point that is stressed as being critical for diminishing error sources, in contrast to the literature overview.

Participants mention an additional point they feel is critical regarding forensic science, which is not raised by the authors in the criminal justice system. It is the *use* that will be made of forensic science. Generally, the focus is mostly, if not exclusively, on forensic science in its function as an item of evidence presented in court. It is then called evidentiary, as defined by Jackson et al. (2006). But it can have a role even earlier in the process that is as an input on the direction a police investigation can take, therefore being of investigative use. In this last perspective, the focus will be more on the potentially interesting information rather than on the “correctness” of the data. This means that the standard applied will not be the same as if this result was to be presented in court. For instance, even if a fingermark does not present enough information for an individualisation (whatever the standard used), the fact that it might be similar to two or three people in a database could open new investigative leads in a case. In this context, forensic science would be used in a way similar to other unconfirmed information coming in, as for instance information given to the police by (anonymous) callers who feel that they have some information to share. In this perspective it might be necessary to accept “imperfect” forensic science, for instance that the result of a preliminary test might not be confirmed or that there will

not be enough data to progress from an investigative to an evaluative use of the specimen. One risk of introducing the concept of investigative use of forensic science is that it might be confused with the evaluative use where the “standards” are higher to be accepted as evidence in court. On the other hand, this approach allows exploiting and using forensic science that normally would be “thrown away” because it does not “pass the evidentiary threshold”, though it still might be of benefit for the criminal justice system (Jackson et al., 2006).

Publications tend to either discuss in a general way how to minimise judicial error (MacFarlane (2003), or they examine errors in forensic science, for example fingermark comparison (OIG, 2006; Haber and Haber, 2008a; 2008b). How exactly both are linked is seldom analysed. If they are both considered, most times the use of DNA profiles for exoneration is what will be mentioned first.

Again, when comparing the solutions to minimise forensic science error presented in the literature and in the interviews differences, but also similarities are visible. Agreement exists on the positive impact of a functioning quality management system on the quality of work performed. However, while CJS scholars feel that the focus should be on laboratories, forensic science practitioners tend to see the next step in a concentration of efforts at the crime scene stage. Disagreement exists also as to whether or not forensic science should be separated into independent entities so as to avoid possible contamination with case information and undue organisational pressure. While this view is represented in literature, practitioners tend to propose the contrary, to strengthen collaboration between stages and entities, so as to maximise the potential of forensic science. A similar disagreement concerns further propositions for change, where CJS scholars tend to stress the need to improve the scientificity of forensic science while practitioners insist on the need to increase the quality of communication between all entities. As discussed briefly when considering sources of problems in forensic science, the difference in point of view could be due to the access of necessary information to evaluate the question. Other possible, more general explanations of differences will be discussed next.

Possible explanations for differences

Two possible factors will be presented here to explain the points of view resulting from the literature review and the interviews. The first centres on the observation that most publications are written by researchers in the USA while interview partners are, with one exception, from European countries. The second considers the feasibility of the proposed solutions which might depend on the practical knowledge of errors committed.

Cultural differences

Literature on forensic science and judicial errors tends to be written by researchers in the USA and some of their British colleagues. Continental Europeans are rather scarcely represented in literature (Broeders, 2006; Jakobs and Sprangers, 2000) while they are the majority in the interviews. This might explain to a certain degree differences observed, as the CJS (adversarial or inquisitorial), the social and political background from which participants are leading the

discussion will have an influence on the problems perceived and the possible solutions envisaged. Thus, for example, variations on appeal possibilities or self regulatory measures to guarantee the quality of forensic science presented in court (as the introduction of a forensic science regulator in England and Wales; http://police.homeoffice.gov.uk/operational-policing/forensic-science-regulator/reports-publications/; last visited 2008/08/22) can have a direct influence on the kind of problems that are perceived to be more urgent to solve. Following the same reasoning, the propositions made to remedy those difficulties will have to be adapted to the situation and will therefore also vary according to the context. Quirk (2007) exemplifies this by comparing the function of the Innocence Project in the USA to the Criminal Case Review Commission (CCRC) in the UK (CCRC for England, Wales and Northern Ireland; SCCRC for Scotland). She explains why those mechanisms are in place in each country and why transferring them from one CJS to another might not be successful, as the needs and possibilities of the systems vary. However, it is argued here, that if there are subjects where across legal systems the same problems seem to persist this could be an important sign to agree that these difficulties are more inherent to forensic science *per se* than to its application in a given CJS. In consequence, efforts for change should be greater and might also be easier to implement. In some countries measures have already been put into practice, as the introduction of the abovementioned CCRC in the UK. Why this has been done in certain countries and not in others may not only depend on the differences in the CJS, but also on the social and political importance given to the subject, as well as the (perceived) frequency of occurrence of the problem. For instance, France has proposed to put in place a panoply of measures and modify several existing mechanisms in place following the historical Seznec case, but also the more recent Outreau scandal. Whether this will really happen has still to be seen.

Ideal or feasible solutions

The cultural diversity mentioned might not only concern nationalities and the CJS that goes with it, but might include also the difference of CJS scholars giving an opinion on (scientific) errors as opposed to forensic science practitioners doing the same. Thus, it could be said that the contrast might not only be on adversarial versus inquisitorial CJS, but also on law versus science.

If taking one example, in literature, the separation of forensic science laboratories from police investigation is perceived to be important to increase the quality of forensic science work (Forst, 2003). One of the reasons for this proposal has been the discovery of falsified forensic science results by some laboratories or people working in them, such as Fred Zain or Pamela Fish (Giannelli, 2007) in the USA. Part of the occurrence has been explained by the too close collaboration of the laboratory personnel with the police investigators (the forensic practitioners “identify” themselves as being on the side of the police against the “bad guys”). In other countries, for instance the UK, this separation exists more strongly due to the (relative) independence of the forensic science laboratories. Forensic practitioners interviewed claim that

this separation is exactly what diminishes efficiency of forensic science. Some Swiss interview partners assert that cooperation between forensic scientists and police officers might even be supportive, as it helps to early cross-check the results obtained both in the laboratory and in the investigation. A further argument that is advanced is that in the early stages of a case the investigative input of forensic science can help to better orientate the police investigation and thus help minimise for instance tunnel bias.

Thus, what is seen as a sore spot by some CJS scholars is seen as a distinctive advantage by the interview partners. The ideas of the latter might be inspired by double-check systems that exist in DNA profile or toxicological analysis. In the first area, the initial specimen will be divided into two and processed in two independent DNA profile analysis laboratory chains. The final results will then be compared to check they are the same. In toxicological analyses for blood alcohol testing (alcohol = ethanol in this context) for instance, blood and urine samples are tested according to established procedures for the same substance and the results are then compared. By this independent double-check in two body fluids it is possible to find out if results correspond and if not, to find out why. However, this approach is not possible for all forensic traces. They are by their nature very often very small and not always divisible. The question that seems to persist is to know which of both propositions seems to be more correct, as those presented are to a certain degree contradictory. It is argued here that this contradiction results also in a different approach as to how to solve problems.

In literature, the focus is on better control mechanisms and more “structural” solutions (Forst, 2003; Koppl, 2005). Interview partners seem to think that other propositions such as improving the training and education of forensic science practitioners could be useful tools as well. They recommend an increase in the level of necessary education, but also the addition of new components such as courses on ethics and error management. This approach can be perceived to be less transparent and visible than to add some more laboratory regulations. Thus, it might be argued that practitioners tend to propose more *feasible*, “individual” changes whereas CJS scholars focus more on “systemic” modifications, which tend to represent an *ideal* or more abstract solution.

Taking the example of separation versus collaboration between the different stages, it is understandable to want to minimise bias by separating them. This represents the ideal point of view. However, from a feasible point of view, complete separation is difficult to implement and keep, as humans have a tendency to communicate with each other. Furthermore, separation implies also “simplifying” or rendering the tasks performed monotone, which might well influence the motivation of the person working on it and thus indirectly the quality of the work performed.

In addition, even if introducing new regulations, those will only be as effective as people implementing them. The human element still determines the strength of the (quality) chain. In other words, even if it might be tempting and easier to modify the general working context, to encourage people in their professionalism seems to be the more successful option in the long run. However, it is easier to propose the establishment of independent laboratories than to

implement courses on forensic science and ethical issues in forensic science, especially since it is much more difficult to see whether those courses will have an effect or not. One part of this discussion is also taking place within the forensic science community, where the two points of view are, broadly summarised, accreditation of laboratories versus certification of personnel (Schiffer and Stauffer, in press; Stauffer and Schiffer, in press).

Comments

What errors happen where is important because answers to these questions will shape the possible solutions. For example, if a protocol has not been followed correctly (e.g. Mayfield (OIG, 2006)), then the possible solution is to make this protocol more exhaustive or clear, so that it will be followed. If the error, however, has been attributed to the daily mood of the examiner, then any protocol will not help solve this issue, as it cannot guarantee good mood. Some might argue that “events” due to bad mood are not errors, this is not the point of view taken here, any wrong outcome is an error, but it can have different causes.

The problem is even more complex. If an error is suspected, then the usual procedure will be to check if this is the case or not. If it is an error, then the reasons for it have to be found. One way of doing so, or the usual way, is to go over existing documentation. This will allow the catching of problems that happened with following the protocol or that have been written down, the daily mood will hardly be documented in the laboratory notes. Thus, “systemic” errors may have a much higher chance of discovery than “human” errors. It is here that the forensic science practitioners have an advantage over CJS scholars, as they know, by their own experience, that those “human errors” can happen as well. Rawley and Caddy (2007) in the Damilola case observe that a trace has been missed though the usual “systemic” tools, protocols, have been employed more or less correctly. Why this error did still happen can not be explained satisfactorily, as it very probably falls into the category of human errors. This is not a way to justify or excuse them, but to explain that they are difficult to catch and deal with. This has to be considered when studying error sources and remedies. This is why the opinion of forensic science has to be included in the discussions or practical recommendations.

In other words, systemic solutions, for example a protocol to follow, will help to guarantee a certain level of quality. But, by them alone it will not be possible to ensure professional work. To do this, it need well educated and trained professionals, who are dedicated to their work. Ideally, this would lead to a professional conscience, a safeguard against errors right at the source of the problem. External developments in this direction are visible in the creation of the forensic regulator in the UK, as mentioned before, or the establishment of lists of recognised experts. In Switzerland the Swiss Chamber of Technical and Scientific Forensic Experts (http://www.swiss-experts.ch/; last visited 2008/08/22) and in the UK the Council for the Registration of Forensic Practitioners (http://www.crfp.org.uk/; last visited 2008/08/22) are examples for this. However, focus should be on implementing the strategy already with the selection, education and training of future forensic science professionals.

Conclusion

By interviewing forensic science practitioners it was possible to confront for the first time their point of view regarding the sources of and remedies to judicial error due to forensic science with the existing opinion represented in publications authored by CJS scholars. Various similarities, but also differences have been observed. For instance, divergences already exist when considering error sources, practitioners seeing the highest risk in the work done on the crime scene, whereas authors tend to see in laboratories the most critical of the three stages of crime scene, laboratory, and evaluation/presentation of evidence. Concerning possible solutions, those proposed in the literature advocate a strict separation of forensic laboratories from investigative bodies, as well as the implementation of quality management measures such as found in the accreditation process. Interview partners in contrast tend to stress the need for communication between forensic science stages and other judicial bodies, as well as the need for increasing the general quality of crime scene work. A further aspect they underline is the need to focus on the education and training of future forensic science professionals, both on methodological and also ethical issues.

Differences as seen here have been in part explained by the fact that literature tends to be published by authors drawing on their USA experience whereas most forensic scientists interviewed work in a European context. A further factor in explaining the differences seems to be the access to information allowing for the understanding of error sources and the making of propositions. This goes hand in hand with the tendency of CJS scholars to propose systemic (ideal) solutions in contrast to individual (feasible) recommendations by forensic practitioners. However, where agreements exist between both opinions expressed, it is suggested here that the underlying problem or solution could be universal and merits therefore special attention.

9.2 Covering letter used for the experimental tests

UNIL | Université de Lausanne
Institut de police scientifique
bâtiment Batochime bureau 6422
CH-1015 Lausanne

Lausanne, le 11 janvier 2006

Recherche concernant les traces digitales

Bonjour,

Dans le cadre de ma thèse, j'ai besoin de ta collaboration et je te serais très reconnaissante si tu pouvais indiquer pour chaque trace digitale les éléments suivants :

si la trace est identifiable ou exploitable
les minuties que tu observes
leur nombre
leur type: A = arrêt de ligne, B = bifurcation, P = point ou U = inconnu

 arrêt de ligne (A)  bifurcation (B)  point (P)

Le but de l'analyse de toutes ces traces digitales est de voir lesquelles posent le plus de problèmes pour les utiliser par la suite dans une recherche plus large, où leur qualité sera un critère primordial.

Les résultats ne seront valides que si le travail est effectué individuellement.

Merci d'avance pour ta collaboration

9.3 Experimental grid for the interviews

Approche générale Avez-vous déjà entendu parler d'un cas d'erreur judiciaire à l'étranger/en Suisse ou les SF avaient joué un rôle primordial? Pensez-vous que les différences du système pénal adversaire vs. inquisitoire jouent un rôle en ce qui concerne le nombre de cas constatés? Avez-vous d'autres possibles explications pour ces différences? Exemples d'affaires connues (bombes IRA)	- Si je vous dis „erreur judiciaire", qu'est-ce qui vous vient spontanément à l'esprit? (pourriez-vous donner une définition « à vous » d'erreur judiciaire?) (evt. Proposer une propre définition pour le reste de l'interview) - pourriez-vous donner des exemples de cas ? (avoir quelques exemples prêts) (evt. – relancer sur l'étranger, la Suisse) Il existe des débats en Criminologie qui disent qu'il y a plus de cas d'erreur judiciaire dans des pays Anglo-Saxons tel les Etats-Unis ou le Royaume Unie qu'en Europe continentale. Que pensez-vous de cette idée? Qu'est-ce qui pourrait, pour vous, permettre d'expliquer ces différences (possibles)? Résumé: autres explications? Certains proposent que ce soit la différence système accusatoire vs. inquisitoire qui fait la différence, que pensez-vous de cette idée? Certains (juristes, publications) disent que les sciences forensiques sont une source importante d'erreurs judiciaires, comment vous vous positionnez par rapport à ceci ? D'autres disent que les sciences forensiques peuvent être un outil important pour mettre en évidence des erreurs judiciaires ? Que pensez-vous ? Ils stipulent également que les sciences forensiques peuvent prévenir des erreurs judiciaires (par exemple en remplaçant des moyens plus problématiques, tels les témoignages). Quelle est votre opinion à ce sujet ?
Questions si les échantillons sont gardés Et qu'est-ce que vous faites des échantillons une fois qu'un jugement a été rendu? Vous les gardez jusqu'à quel moment? Et puis vous en faites quoi après? à quel	En ce qui concerne la différence de cas d'erreur judiciaires entre le système USA et CH pourrait tout simplement être due au traitement différentiel une fois que l'affaire a été jugée, c'est-à-dire combien de temps les prélèvements et échantillons son gardés. Dans quelle mesure êtes-vous d'accord (avec ceci)? Pourriez vous me donner un exemple type de ce qui se passe avec un échantillon tout le long dès la scène de crime jusqu'à sa fin chez vous? (Combien de temps les gardez-vous au maximum? Une fois qu'un jugement a été rendu? à quelle

stade de la procédure peut-on détruire les échantillons?	stade de la procédure peut-on détruire les échantillons?) Résumé: L'hypothèse/idée si les échantillons ne sont pas gardés il n'est pas possible d'en refaire des analyses, ce qui limiterait le nombre de cas où l'on pourrait constater des erreurs et donc ultérieurement des erreurs judiciaires? Avez-vous déjà regretté une fois d'avoir « détruit » un échantillon? (Peut-être parce que vous auriez aimé re/faire une analyse/comparaison)
Le rôle de la formation dans cette problématique Pensez-vous qu'une telle formation peut aider à prévoir, anticiper, éviter et corriger de telles erreurs ? Pensez-vous qu'une formation plus spécifique est nécessaire pour approcher et sensibiliser les gens à ce type de problématique?	D'autres arguments évoqués dans des publications sont la formation des collaborateurs de police scientifiques. Que pensez-vous de cet argument? Les mêmes gens continuent en disant que si on arriverait à améliorer la formation il y aurait certain problèmes qui disparaîsseraient. Etes-vous d'accord avec ceci? Si on voudrait sensibiliser les collaborateurs scientifiques par rapport à la problématique des erreurs judiciaires, à quel moment de leur formation faudrait-il le faire (formation de base/continue))?
	On a parlé de sources d'erreurs possibles d'une manière générale, dont les SF comme une d'entre eux et j'aimerais bien approfondir le sujet en tirant profit de votre expérience pratique.
Exemples de (presque)-erreurs dans la pratique de l'interlocuteur Avez-vous déjà eu dans votre laboratoire des "presque" erreurs ou qqch aurait pu se passer, mais ceci n'a pas eu lieu Pourriez-vous donner des exemples evt. Proposer des exemples abstraits (découvert à temps, pas utilisé, fausse identification, échantillons confondus) Comment c'était découvert (donner des exemples	Avec votre expérience de terrain, quelles sont les situations à risque? A quelle situation de risque avez-vous été confronté dans votre carrière ? avez-vous rencontré dans votre entourage (collègues, discussions, etc.)? Dans quelle mesure jouent-elles un rôle dans la pratique quotidienne? (Evt. Appeler ceci des situations de « presque- erreurs »?) Quels étaient les éléments qui ont permis de mettre en évidence ces difficultés? (procès normal de contrôle, contrôle par deuxième collaborateur, relecture, problème « logique ») (procès- extraordinaire: investigation policière contredit résultat analytique, trouver explication pour ceci; contre-expertise (ex. ZH ou autre; identification ED au labo ne pas soutenue par l'AFIS)) quelle était votre/la réaction à la découverte de l'erreur ?

possibles) contrôle par 2ème collaborateur réaction à un constat d'erreur? Avez-vous des check-lists, des formulaires/ protocoles internes ? Quelles sont, selon vous, des raisons possibles pour des erreurs telles qu'on vient de parler ? Le poids de ses erreurs ?	Mesures prises; information de qqn; changement de manière de travailler; délégation de travaux difficiles, etc. Avez-vous mis en cause des éventuelles check-lists, formulaires/protocoles internes, les établies car inexistants, etc.; sinon ceci vous semble une mesure utile ou pas ? (si j'ai bien compris des tendances d'accréditation etc. touchent même les laboratoires de police scientifique); utilisées ? S'il y a des erreurs ou presqu'erreurs, quelles sont, selon vous, les raisons possibles? (erreur humain; on n'est pas parfait et ça arrive; manque de formation; de connaissance; surcharge de travail- fatigue- manque de concentration; problème technique) Résumé: ou dans le processus du prélèvement de la trace jusqu'au juge est-ce que ces types d'erreurs vont se faire ressentir; pour dire autrement, le point faible de la chaîne? (lors de l'information communiquée aux policiers afin de faire avancer l'investigation ou lors du jugement de culpabilité par le juge)?
Solutions à proposer ? Ce qui est nécessaire pour résoudre ces problèmes (ces problèmes = erreurs possibles, etc.)	Si j'ai bien compris vous dites qu'il (n)' y a (pas) de problèmes concernant les erreurs judiciaires en police scientifique. S'il faudrait changer quelque chose pour éviter ce type de difficultés, vous commenceriez par quel point ou aspect? La première amélioration possible? Elle permettrait d'accomplir quelle tâche?