

CONCEPTUALISING BELIEFS AND UNDERSTANDING THEIR  
(MIS)USE IN THE FORENSIC SCIENCES

NADINE M. SMIT

*Doctor of Philosophy*  
*in*  
*Forensic Science*

*University College London*

*April 2019*



## DECLARATION

I, Nadine Smit, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

---

*Dated:*



To my grandparents,

who have shown me the value of moderation, common sense,  
hard work, and the occasional moment of rebelliousness.



## ACKNOWLEDGEMENTS

I would like to express my appreciation to those who have contributed, in so many different ways, to the completion of this thesis.

To my supervisors Professor Ruth Morgan and Professor David Lagnado for their expertise, ideas, and encouragements throughout these years. Thank you both for your patience and the trust you have given me to follow my ideas.

To my examiners Professor Cheryl Thomas and Professor Maurice Aalders for their valuable comments and interesting discussions during my viva.

This research has not been possible without funding from the UK Engineering and Physical Sciences Research Council through the Security Science Doctoral Research Training Centre at UCL. A big thank you goes out to everyone at the Department of Security and Crime Science for the opportunities, memories, and friendships.

I owe a thank you to those who have brought out the best in me throughout these years. Thank you to Esther and Marinel, for the endless moral support. To Evangelos and Alessandra, for encouraging me to start this journey in the first place. To Sherry, for all our discussions and for believing in change. For all the other memories and adventures in London and beyond, Eon, Dorothea, Beth, Nathalie, Helen, Sanaz, Jacob, James, Agathe, Miriam, Aryak, and Rory: thank you.

A final thank you goes out to my parents and brothers, for their love and support through anything and everything.



## ABSTRACT

The growing forensic evidence base is often not fully reflected in the development of theories and the truth-finding process throughout the criminal justice system. This is worrying; it may lead to a misunderstanding of the meaning and value of the evidence by decision-makers, which has been shown to result in wrongful convictions on a significant scale. One of the justifications for the range of issues and concerns expressed over the years regarding the analysis, interpretation, and presentation of evidence includes a limited situational understanding of observations at that time. This thesis addresses the hypothesis that some misinterpretations (i.e. an invalid belief in an hypothesis given the evidence) could have been minimised.

Four related approaches demonstrate that in many cases, better inferences and decisions could have been made. First, the legal, historical, logical, and knowledge-base perspectives on making reasonable and fair arguments from observations are explored. The challenges which need to be overcome between the law and sciences are demonstrated, and a case is made that fundamental theories of analyses and interpretations have not been studied sufficiently in the light of scientific approaches and reasoning processes. Second, a systematic study of Court of Appeal cases identifies that misleading evidence is a prevalent and sometimes avoidable issue in England and Wales. Examples include a misinterpretation and miscommunication of the relevance, probative value, and validity of evidence. Third, to explore the misuse of evidence more in-depth, a novel conceptual framework incorporating key components of interpretations from trace scripts is developed, allowing for systematic approaches to evaluating interpretations. Fourth, this framework is applied to the interpretation of geo-forensic evidence. It is demonstrated that the approaches taken to study uncertainties and the effects of these approaches on the expression of beliefs and decision-making processes are not always sufficiently taken or transparently presented.



## IMPACT STATEMENT

The findings of this thesis, which provide a better understanding of the nature and extent of misinterpretations of forensic evidence and present methodologies to evaluate and avoid these, can and have shown to be of value both within and outside academia. The results have opened up a range of new research possibilities, by laying the previously lacking necessary foundation and systematic methodologies to study misinterpretations beyond the criminal cases, jurisdictions, and evidence types presented in this thesis. Moreover, by filling the knowledge gap by making the misinterpretation of evidence within cases presented at Crown Courts in England and Wales explicit, it can inform policy makers and the public on the current state of forensic science.

The work shown in this thesis has been presented at national and international conferences, which were attended by a range of audiences including practising legal scholars, forensic practitioners, and academics. These disseminations have led to an interest from researchers across domains, and more specifically, have resulted in collaborative projects on financial and medical evidence within England and the United States. More generally, interest has been shown into the methodologies developed in this thesis, and the possibility of extending these to form the foundation of machine learning approaches. Questions have also been raised on the impact these findings have on forensic practice, and how experts can incorporate this knowledge when presenting evidence in court.

In addition to scientific audiences, these findings have shown to be of value to policy makers, where results have been discussed with the Government Office of Science and presented to the House of Lords Science and Technology Committee, where it has informed their recent inquiry into forensic science (House of Lords, 2018). This highlights an acknowledgement that these issues occur and that they should be addressed, and that there is a willingness from policy makers to achieve this. Moreover, this research has been used as a foundation of a response paper questioning the extent to which this applies to legal practice in jurisdictions outside of the United Kingdom (Broeders, 2018).

The findings from this thesis have not only been disseminated through academic conferences and journal articles (Smit, Morgan, and Lagnado, 2017), but have also been discussed in the press and on social media, and thereby reaching members of the general public. Examples include articles both with regards to the specific state of DNA analysis and interpretation (The Conversation, 2017; The Guardian, 2017a), as well as

on the general state of forensic science (The Guardian, 2018a; The Law Gazette, 2018). Moreover, the findings from this thesis have been disseminated to the press by being highlighted in the Elsevier Research Selection. Lastly, letters have been received from imprisoned individuals as a result of the publication of the findings from this thesis, hoping that these findings can shed some light on their alleged wrongful convictions. This shows the opportunities that follow from this research to start and continue the discussion on the misinterpretation of forensic evidence.

## CONTENTS

1	INTRODUCTION	1
2	FIVE PERSPECTIVES ON MAKING JUSTIFIABLE ARGUMENTS FROM FORENSIC EVIDENCE	5
2.1	Introduction	5
2.2	Legal perspective	6
2.2.1	Legal systems	6
2.2.2	Evidence admissibility standards	7
2.2.3	Normative benchmarks of legal decisions	8
2.2.4	The law of evidence	9
2.2.5	Judicial decision-making research	11
2.2.6	Science and the law: negotiating a partnership	13
2.3	Historical perspective	14
2.3.1	Pragmatism	14
2.3.2	The post-DNA era	15
2.3.3	The rise of governmental concerns	16
2.3.4	An increased understanding of uncertainties	18
2.4	Reasoning perspective	21
2.4.1	Truth, beliefs, and reasoning	21
2.4.2	Probabilistic-based reasoning	22
2.4.3	Bayesian networks	25
2.4.4	Decision-making throughout the forensic chain	26
2.5	Knowledge-base perspective	27
2.5.1	Knowledge sources and mechanisms	29
2.5.2	Structuring information from empirical studies	30
2.5.3	Making tacit expert reasoning explicit	33
2.6	Thesis' perspective	34
3	MISLEADING EVIDENCE IN UNSAFE RULINGS IN ENGLAND AND WALES	35
3.1	Introduction	35
3.1.1	Unreliable expert evidence and uncertainties	36
3.1.2	Misleading evidence and unsafe rulings	37
3.1.3	Present study: a structural approach	39
3.2	Materials and methods	42
3.2.1	Case selection	42
3.2.2	Coding categories and considerations	43

3.2.3	Case attributes . . . . .	43
3.2.4	Inter-coder reliability . . . . .	44
3.3	Results and discussion . . . . .	45
3.3.1	Significance: number of cases . . . . .	45
3.3.2	Further significance: time between rulings . . . . .	46
3.3.3	Coding method reliability . . . . .	47
3.3.4	Overview of the results in each coding category . . . . .	48
3.3.5	Issues with the nature of the evidence . . . . .	50
3.3.5.1	Relevance of the evidence . . . . .	51
3.3.5.2	Probative value of the evidence . . . . .	52
3.3.5.3	Validity of the evidence . . . . .	53
3.3.5.4	Standard of proof . . . . .	54
3.3.6	General issues and the decisions by the jury . . . . .	54
3.4	Conclusion . . . . .	55
4	<b>NORMATIVE APPROACHES OF EVALUATING TRACE INTERPRETATIONS</b>	57
4.1	Introduction . . . . .	57
4.2	Trace scripts . . . . .	59
4.3	Differentiating between reference samples based on observed features . . . . .	61
4.3.1	Diagnosticity of analysis and abundance of property features . . . . .	61
4.3.2	Independence between features . . . . .	63
4.3.3	Validity of analytical methods . . . . .	63
4.3.4	Data exploration and reduction . . . . .	64
4.4	Classifying unknown traces . . . . .	64
4.4.1	Hypotheses and the (strength of) evidence . . . . .	65
4.4.2	Classifications, exclusions, and posterior likelihoods . . . . .	66
4.5	Summary of primary components . . . . .	67
4.6	Conclusion . . . . .	68
5	<b>EVALUATING CURRENT INTERPRETATIONS OF SOIL EVIDENCE</b>	71
5.1	Introduction . . . . .	71
5.2	Methods and materials . . . . .	72
5.3	Results . . . . .	73
5.4	Comparability of features . . . . .	73
5.5	Validity of analytical method . . . . .	76
5.6	Diagnosticity of observed features . . . . .	76
5.7	Interpretation of observed features . . . . .	78
5.7.1	Clustering and classifications . . . . .	78
5.7.2	Hypotheses . . . . .	79
5.7.3	Conclusions and decisions . . . . .	80
5.8	Discussion and conclusion . . . . .	81

6	DISCUSSION	83
6.1	The misinterpretation of forensic evidence . . . . .	84
6.2	The development of normative frameworks to evaluate interpretations . .	86
6.3	Requirements of making valid forensic interpretations . . . . .	87
6.4	Innovative mindsets driving the future of forensic interpretations . . . . .	89
7	CONCLUSION	91
7.1	Drawbacks and opportunities for valid interpretations within the wider framework . . . . .	91
7.2	Understanding misleading evidence from appeal court judgments . . . . .	92
7.3	Developing normative frameworks for the evaluation of interpretations . .	93
7.4	The state of interpreting geo-forensic data . . . . .	93
7.5	Implications of this thesis . . . . .	94
	PUBLICATIONS	95
	BIBLIOGRAPHY	97



## LIST OF FIGURES

Figure 1.1	Schematic overview of the general processes in play from evidence creation until legal judgments . . . . .	3
Figure 2.1	Schematic overview of the perspectives discussed throughout Chapter 2 . . . . .	6
Figure 2.2	Summary of inductive and abductive reasoning processes . . . . .	22
Figure 2.3	Simplified overview of knowledge types within the forensic sciences	30
Figure 2.4	Overview of the terminology used in the 145 cases where textile fibres were presented as evidence in court . . . . .	32
Figure 2.5	Sankey diagram highlighting the relation between the ten test studies in terms of parameters they vary and characteristics they measure. . . . .	33
Figure 3.1	Diagram highlighting the study set obtained through studying overturned unsafe rulings. . . . .	39
Figure 3.2	The coding scheme developed for this study to classify the reasons for rendering the trial ruling unsafe by the Court of Appeal judge(s). . . . .	44
Figure 3.3	An overview of the evidence type groups and the reasons given why the evidence was misleading. . . . .	49
Figure 4.1	Graphical view of the sections, aims, and methods in Chapter 4 and 5. . . . .	59
Figure 4.2	Generic trace script for the behaviour of a sample either developed as a trace or collected as reference sample . . . . .	60
Figure 4.3	Graphical representation of the effect of combining analytical methods and using different populations in the process of clustering samples . . . . .	62
Figure 6.1	Schematic overview of the four themes which follow from the thesis	84



## LIST OF TABLES

Table 3.1	An overview of studies into larger datasets and significant cases where issues with the forensic evidence were observed. . . . .	41
Table 3.2	Summary of case details which were included in the study set . .	46
Table 3.3	Overview of the results of each coding category from Figure 3.2 .	49
Table 3.4	Overview of the of the number of cases of misleading evidence and the type of hypothesis that they addressed. . . . .	51
Table 3.5	Overview of the number of cases of misleading evidence and the detailed relationship between the different levels of hypotheses and the nature of the misleading evidence. . . . .	52
Table 4.1	Key components which contribute to the normative approach of interpreting trace evidence . . . . .	68
Table 5.1	Overview of the different type of aims of a study depending on their samples used . . . . .	73
Table 5.2	Overview of observations for each component group from the used case studies. . . . .	74
Table 6.1	Comparison between methods of evaluating interpretations within court judgments and published literature . . . . .	87



## CHAPTER I

### INTRODUCTION

**Jus-tice** / ' dzʌstɪs / noun (..)

**1.1** the quality of being fair and reasonable

- *Oxford English Dictionary, 2016*

Justice should be an idealistic driving force for any aspect of society, and the process of attributing guilt following presumed criminal actions against society is no exception. An inevitable question is, then, how *fairness* and *reasonability* can be achieved within this legal domain. The key mechanism is a system of rules outlining criminal acts, whereby a Trier-of-fact decides whether (they believe that) the requirements and standards of certainty of guilt are met (Horowitz, 1997). A major assumption underlying this process is the idea that one, to a certain extent, *can be* fair and reasonable when making decisions while facing uncertainties. This idea contains two parts: an ability to express a belief in an hypothesis as close to the truth as possible, and a mechanism to make informed and justified decisions following this belief. Where the former requires a valid and sufficient knowledge base and sound inference methods, the latter is based upon established thresholds and acceptable information gains. These processes form the basis for evaluations and decisions from observations at a variety of stages within the criminal justice system, from initial crime scene examinations until conviction and sentencing decisions in court.

The term *belief* is used throughout this thesis as an expression of the probability that an event will occur (i.e. that an hypothesis is true). This term is commonly used to highlight the difference between 'objective' frequencies at which an event occurs and 'subjective' degrees of belief that an event will occur (Pearl, 1988). This notion of subjectiveness does not necessarily suggest that the expression is non-scientific, but merely that it is dependent upon which (scientific) premises are considered. The challenge with expressing a belief in an hypotheses is that these can not simply be expressed as true or false, as the inevitable presence of uncertainties makes knowing the complete truth impossible. A lack of acknowledging, understanding, incorporating, and communicating these uncertainties throughout the criminal justice system has shown to result in wrongful interpretations and subsequent wrongful convictions on a large scale in the UK and

beyond. Despite a range of questions raised about the scientific validity of forensic arguments by scholars and others (Garbolino, 2013), it has not resulted in sufficient rigorous reforms in forensic practice. The major consequences affecting the justice system as a whole and the concerns raised by the forensic community and governmental agencies will continue to emerge if the processes of identifying fallacies and improving practice is solely based on top-down approaches. This thesis therefore supports and contributes to the premise that the forensic field needs to be improved from the bottom-up: using reliable and accurate fundamental principles of the sciences more generally, to allow for a transparent understanding and communication of observations, limitations, and effects of uncertainties.

The process of how these uncertainties arise is represented in Figure 1.1, of which the details are discussed and substantiated throughout this thesis. The process starts with factors affecting the state of evidence once it is 'created' as a consequence of a criminal act, summarised as evidence dynamics. Such evidence is subsequently collected and (partially) submitted for (various) analyses. Within these processes, factors such as confirmation bias due to formed hypotheses and the effect of the case context play an important role when determining the possible relevance and probative value of these items of evidence. As will follow from Chapter 4 and 5, the analysis stage may greatly depend on the experience and access to methods of the analyst. Depending on the objectivity of the analytical method, information on the context in which the evidence was found has shown to influence results. Further interpretation of these results (i.e. expressing a belief in hypotheses) depends on the way these hypotheses are defined. Here, the factors in earlier stages that can cause additional uncertainties must be taken into account, as well as the comprehensiveness and accuracy of databases used to determine population frequencies and thereby the evidential value. The subsequent communication of these results will be one of the main topics throughout this thesis, as it has shown to play a crucial part in how beliefs are updated in the light of new evidence and how fair judgments are eventually made.

Although making less than ideal decisions when reasoning under uncertainties is inevitable, it does not mean that such mistakes should be accepted and simply taken for granted. This is especially the case in reconstructing criminal events, as the stakes in the pursuit of justice are high. Moreover, the field of forensic science operates in the wider scheme of the criminal justice system where often, there is no time to 'wait for the science' to improve arguments, and beliefs are required which, although not perfect in a scientific mindset, might be helpful in a court of law (Mnookin et al., 2011). The general aim of this thesis is therefore to study the extent to which it is possible to express more justified beliefs and subsequently make better decisions in forensic reconstructions. Although what drives science and fascinates scientists is the possibility of developing theories to address phenomena that are currently unknown, it is equally important to

### Biases influencing interpretations

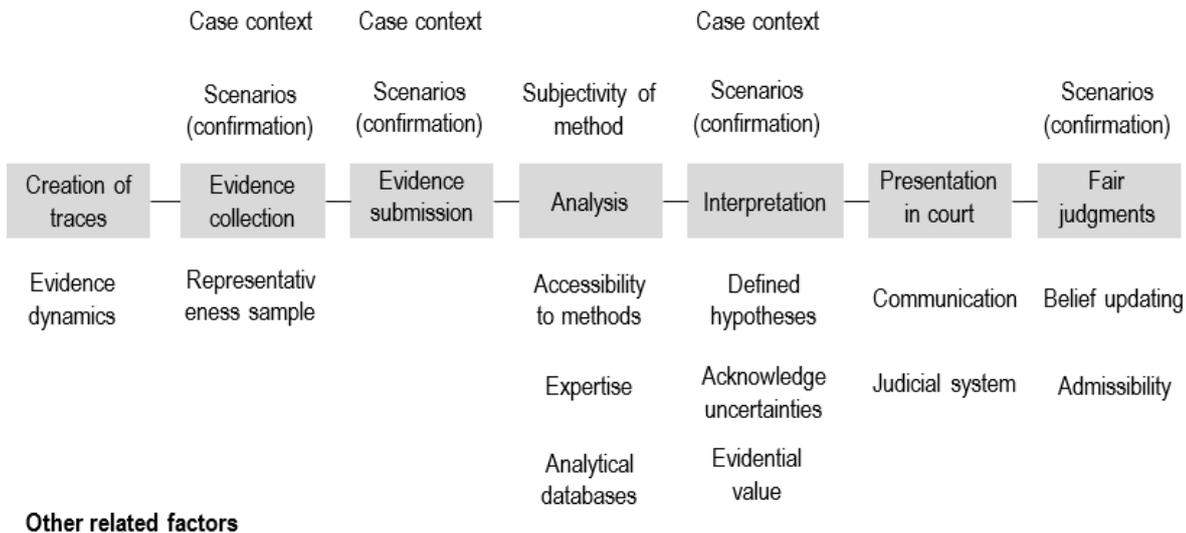


Figure 1.1: Schematic overview of the general processes in play from evidence creation until legal judgments. Each step (grey box), faces both factors originating from biases (top) and due to other factors such as the knowledge base or proposed hypotheses (bottom).

think critically towards earlier findings and evaluate whether one can rightfully build upon that knowledge. Rather than focussing on the decisions and thresholds for reaching various decision-making outcomes, this thesis will therefore focus on the initial step of the conceptualisation of information and reasoning towards beliefs. The challenge lies with acknowledging the presence of hindsight bias – where it is easy to argue how arguments and decisions should have followed from the knowledge base – better, while ignoring what knowledge gain has been achieved in the meantime that has contributed to those ‘better’ decisions. However, understanding where such discrepancies are can support anticipation of the growing knowledge base in the future. The following four chapters aim to address the extent to which these evaluations have been helpful within the criminal justice system by taking a theoretical, legal, and scientific approach.

First, Chapter 2 explores the legal, historical, logical, and knowledge perspectives on making reasonable and fair arguments from forensic evidence. This chapter shows the challenges which need to be overcome between the law and sciences, and that many of the fundamental ideas of analyses and interpretations have not been studied sufficiently in the light of scientific approaches and logical reasoning processes. Therefore, although there are distinct differences in the nature between the sciences and the law, and there is still much to discover regarding the interpretation of forensic trace evidence, there are many things which are currently known which should be better acknowledged and incorporated in reasoning.

Following a theoretical approach in Chapter 2, Chapter 3 aims to understand the extent and nature of misleading evidence currently present in criminal cases in England and Wales, following successful appeals in the Criminal Court of Appeals. The findings from a systematic empirical study to address this question found that the majority of successful appeals were based upon the same materials available in the original trial, rather than the presentation of new relevant information. The validity of witnesses, probative value of forensic evidence, and relevance of character evidence were the most prevalent combinations of identified issues, with the majority of misleading evidence types relating to their interpretation at activity level. The findings suggest that many of these misleading aspects could have been understood at an earlier stage by providing more transparency in the relationship between evidence and hypotheses.

Understanding these issues from a legal perspective highlights specific points which should be addressed by forensic scientists. However, the question remains whether it is currently possible to address these given the current capabilities of the sciences more generally and situational capabilities of individual scientists. Therefore, Chapter 4 aims to address the ability to develop a framework encapsulating a normative approach to trace evidence evaluations. It develops and introduces the notion of trace scripts and key components of trace interpretations, which are needed in order to comprehensively and systematically evaluate the ways in which interpretations should and can be done.

Using this normative framework, Chapter 5 takes a descriptive approach to trace evidence evaluations, using the interpretation of geo-forensic evidence as an example. It aims to understand current interpretations and evaluates the extent to which these are (in)valid and (in)sufficient in supporting legal questions and expressing beliefs in established hypotheses. It follows from this study that in the interpretation of geo-forensic evidence in particular, there is a clear lack in understanding the analytical capabilities when assessing data and an ignorance towards understanding the limitation of findings.

The discussion in Chapter 6 introduces four general themes which have emerged throughout this thesis. First, the nature of misinterpretations of evidence and possible explanations for their presence. Second, the variety of normative and descriptive frameworks which can and should be used to study present issues and concerns in the forensic sciences. Third, the need to recognise that distinct knowledge is required from outside the forensic domain to ensure robust interpretations, as the more general reasoning methodologies and scientific endeavours all have a similar nature independent of the scientific domain. The final theme addresses the lessons learned regarding the development of a structured knowledge base and valid reasoning methods, anticipating on the future growth in the forensic and wider scientific domains, and with that, an increased expectation of contributing to the idea that expressed situational beliefs are 'good enough' in the search for justice.

## CHAPTER II

### FIVE PERSPECTIVES ON MAKING JUSTIFIABLE ARGUMENTS FROM FORENSIC EVIDENCE

#### 2.1 INTRODUCTION

The aim of this chapter is to identify and study key aspects of the wider theoretical framework of forensic science, in order to conceptualise and highlight the gaps between the current situation and an idealistic goal of evidence interpretation. Here, an interpretation is seen as a belief in an hypothesis following the combination of case-specific observations in an investigative setting and a general knowledge base (further explained in Section 2.5). Moreover, this chapter aims to identify the obstacles to reaching such a reality and introduces mechanisms to possibly address those.

As part of this analysis, four related viewpoints are discussed, which are shown in Figure 2.1, and are followed by a section discussing the perspective taken in this thesis. First of all, accepting that forensic reconstructions support various aspects of legal reasoning, questions need to be raised on the use and requirements of forensic evidence within criminal law, using a number of legal systems as examples. This is followed by a section which sheds light on the scientific inferences and legal decisions from a historical perspective, in order to learn from past achievements and mistakes. Third, accepting that the reasoning from evidence towards a belief in hypotheses follows a coherent and established approach, the challenges and possibilities of applying such approaches to forensic inferences are studied. The fourth perspective focusses on the knowledge and information needed within these arguments and the extent to which the knowledge base and investigative-specific information can be obtained. The final and fifth perspective formulates the ways in which this thesis contributes to achieving the ultimate aim of making fair and reasonable judgements from forensic evidence in the light of these four perspectives.

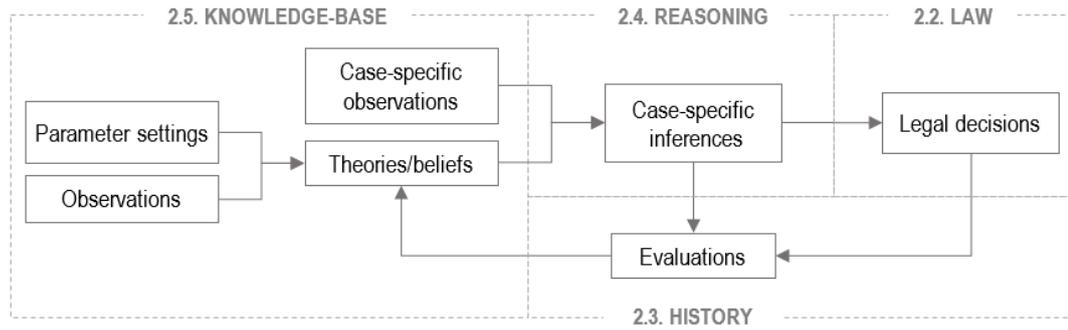


Figure 2.1: Schematic overview highlighting the relationship between the four perspectives discussed in Sections 2.2 - 2.5. Section 2.6 discusses the fifth perspective, which introduces the approach taken in this thesis.

## 2.2 LEGAL PERSPECTIVE

### 2.2.1 *Legal systems*

It is important to explore the aspects and constraints of the relationship between legal systems and scientific domains that are relevant when studying the added value of evidence. Legal systems can generally be distinguished by their means of truth-finding: through a discussion of opposing parties (the adversarial system) or by a single authoritative investigation (the inquisitorial system) (Brants, 2011; Koppen and Penrod, 2012), although most are in practice to certain extents influenced by both (Field, 2009). Moreover, the judge has a more active role throughout and before the trial in the inquisitorial system as compared to mainly ensuring that cases are dealt with justly and following the procedure rules. Additionally, previous case decisions form a more important basis on future decisions in an adversarial system as compared to inquisitorial systems. With respect to the truth-finding process, the defence in inquisitorial-led systems usually does not have an active role in establishing the facts in a case. For example, in both Switzerland (after the Swiss Code of Criminal Procedure replaced the cantonal regulations in 2011 (Gilliéron, 2013; Killias, 2008)) and the Netherlands (Brants, 2011), the prosecutor directs the pre-trial investigation, and handles the examination, charges, and prosecutions. The defence plays a subordinate role and often merely evaluates whether the process is conducted fairly under the respect of the law. As a result, the information is available before the trial and dual proceedings are avoided (Gilliéron, 2013). In more adversarial-led systems such as England and Wales and the United States, the opposing parties both play an active role in the investigation and present their evidence during the trial to the Trier-of-fact (Findley, 2011). Due to the nature of the adversarial system, it has been said to be driven by consensus rather than the inquisitorial views of finding out the truth (Spottswood, 2015). It is, however, argued that the inquisitorial system faces vulnerabilities due to a single authority leading the investigation, such as tunnel vision in confirming the suspect's guilt (Roach, 2010). For example, requesting an additional

expert often requires special circumstances (Roach, 2010), and because they are seen as ‘non-partisan’, their statements are often accepted without being challenged (Brants, 2010). Additionally, in practice, confusion can exist between judges, prosecutors and defence attorneys as to whose role it ultimately is to question the validity of the evidence (Vuille, 2013). Another aspect which may influence the extent to which forensic findings are questioned is that many of the criminal cases are generally resolved before the case even goes to court, through the so-called alternative proceedings in Switzerland or plea bargaining in the US (Gilliéron, 2014). Even though such plea-bargaining options save a considerable amount of time and money, they are often criticized for having a negative effect on the integrity of the criminal justice system (Kulesza, 2011). In such cases, an extensive amount of trust is put into the preliminary results of forensic analyses, where little opportunity has presented itself to question these results at that point in time (Edmond et al., 2015). One of the ways to safeguard against the effects of insufficiently questioning evidence during a trial is by legally regulating their use through evidence admissibility standards.

### 2.2.2 *Evidence admissibility standards*

In many federal courts in the United States, the requirement of ‘general acceptance’ was introduced in the Frye standard (Frye v. United States, 1923) after discussion that surrounded the admissibility of polygraph test results. However, the belief that the 1975 Federal Rules of Evidence superseded the Frye standard in the Daubert case in 1993, led to the formulation of the Daubert standard, which incorporated requirements regarding the falsifiability, peer-review, error rates, and standards for the analysis of evidence (Daubert v. Merrell Dow Pharmaceuticals, 1993). Some years later, the Kumho-ruling stated that the Daubert standard should also be applied when non-scientific evidence is presented (Kumho Tire Co. v. Carmichael, 1999). Even though these rulings were criticized for their limited impact on admissibility rates in the United States (Groscup and Penrod, 2002), and many states continued to use solely the Frye standard (Cheng and Yoon, 2005), they did require judges to scrutinize the (scientific) evidence in more detail (Cheng and Yoon, 2005). More generally in European countries (including England and Wales, Switzerland, and the Netherlands), the rights and principles of the legal systems were greatly impacted by the European Convention of Human Rights (Keller and Stone Sweet, 2008; Killias, 2008), protecting, for example, the suspect’s right to a fair trial (Art. 6 ECHR). Furthermore, admissibility standards in the UK were widely discussed after the publication of the Law Commission report in 2011, in which a statutory admissibility test was recommended (The Law Commission, 2011). The UK government responded that, “without certainty as to the offsetting savings which might be achieved, when set against current resource constraints it is not feasible to implement the proposals in full at this time” (UK Ministry of Justice, 2013, p.4). They did, however, argue for an amendment of the binding Criminal Procedure Rules providing, for example, the judge with

more information regarding the expert evidence pre-hearing (UK Ministry of Justice, 2018). Although some admissibility standards are not always applied fully, their importance and need are increasingly being acknowledged (Tully, 2018). When inadmissible evidence is nevertheless introduced in court, the judge has the option to instruct the jury to disregard such evidence in their decision-making process. However, Lieberman and Arndt (2000) have reviewed a range of studies suggesting that jurors are often not capable of completely ignoring such information. On top of this, experienced judges are often not able to suppress such subjective reactions either (Landsman and Rakos, 1994; Wistrich, Guthrie, and Rachlinski, 2005). It is therefore crucial that these effects are limited by establishing and following clear admissibility standards.

There is also a downside to having scientific standards set by the courts, especially in common law systems where precedential decisions constitute the basis for the body of law. For example, issues may arise such as those highlighted in the case of *Regina v T*. Here, the validity of using likelihood ratios in the interpretation process was rejected in the UK court (*R v T*, 2010). Whilst it cannot be denied that the expert failed to transparently report his methods, the ruling impacted the use of likelihood ratios as a whole, resulting in a range of critical and concerning responses from the forensic community (Thompson, 2012). Needless to say, the use of evidence to express beliefs in hypotheses requires an extensive safeguarding system in ensuring that it contributes to fair and reasonable decisions. This adds to the point taken in this thesis that a forensic expert should take the major responsibility for providing valid arguments in a transparent and non-partisan manner. This is crucially important as they are part of the basis on which a Trier-of-fact makes their decision. Where the theoretical framework and normative benchmarks of these decisions is discussed in Section 2.2.3, the importance of understanding these arguments will be emphasized more throughout the remainder of this chapter, to lay the foundation for the study addressing the misinterpretation of evidence from appeal court judgments in Chapter 3.

### 2.2.3 *Normative benchmarks of legal decisions*

Accepting that the Trier-of-fact does not know the complete truth and is required to make decisions in the light of a range of uncertainties, the difference between factual guilt and finding someone guilty needs to be addressed. The question which follows is how it can be ensured that the 'best' decision has been made given the complexity and significant consequences of making a 'wrong' decision. This process faces two challenges: how to specify the normative benchmark of an ideal judgement and how to describe and evaluate the behaviour of the Trier-of-fact in the light of this benchmark (Mitchell, 2010).

One of the normative perspectives to judicial decision-making follows from the standards of proof that have been set throughout an investigation to allow, for example, holding a suspect, obtaining a search warrant, or convicting a suspect. In some states in the United States, the standard of proof for the latter decision is that of ‘beyond reasonable doubt’, where the proof must be of “such a convincing character that a reasonable person, after careful consideration, would not hesitate to rely and act upon that proof in life’s most important decisions” (Eighth Circuit Judicial Committee, 2017, p.92). In the Crown Courts in England and Wales, the judge should direct the jury that “they have to be satisfied so that they are sure before they can convict” (Judicial College, 2018). Where admissibility standards (Section 2.2.2) mainly focus on when evidence is allowed in court in the first place, jury directions highlight how to apply current legislation to established beliefs, and standards of proof focus on decision thresholds, less attention has been paid to legally regulating the way in which evidence should be used when acting as a foundation of formulating arguments or expressing beliefs. Those who have addressed this issue argue, for example, that normative judicial decisions could not only be based on optimal trial outcomes or best fits with the current law, but should also consider, among others, statistical and epistemological models (Spottswood, 2013). One of the key normative benchmarks should at least contain the idea that the observations from evidence can merely contribute to developing a belief in an hypothesis, and that decisions from these beliefs require stated assumptions and justifications with regards to decision thresholds.

Facilitating studying the normative-descriptive gap, the Trier-of-fact in, for example, Switzerland and the Netherlands, must state the rationale behind their decision (Gilliéron, 2014), where the obligation of a Swiss jury to present their reasoning and the difficulty of meeting these requirements contributed to a move away from a jury system (Killias, 2008). Were jurors in England and Wales are prohibited to disclose their argumentation (Contempt of Court Act 1981), it does generally not prevent research on juries, although this ruling has caused uncertainties around such possibilities (Thomas, 2008).

#### 2.2.4 *The law of evidence*

Historically, it has been criticised that the evaluation of evidence in a legal setting was often approached as the law of evidence, which concerns itself with the rules of what and how evidence should be presented in court, rather than with the study of the evidence itself (Murphy, 2007). As one of the pioneers in the field of evidence, Bentham even stated that the rules of evidence were merely restrictions upheld by lawyers on what would later be called the science of evidence (Murphy, 2007). Moreover, Bentham criticized the pragmatic way in which the law was systematised and applied (Walton, 2010), preferring the idea of a ‘natural system’ of evidence evaluation, in which the Trier-

of-fact should assess the probative force of the evidence based on the extent to which he was persuaded. He acknowledged that this could very well be less than moral certainty, and stated that in order to establish such probabilities, one needs to both assess the probability of the testimony being true and adjust the original probability in the light of new testimonies, creating a ‘chain of reasoning’ (Twining, 1985).

During the 19th century, various attempts were made to rationalise the field rather than the use of fragmented categories of sources of evidence and fields of criminal law (Twining, 2011). It was not until Wigmore (1863–1943), who was the first to use the term science of evidence in a scholarly setting (Schum, 1994), presented the idea of structured arguments (‘the logic of proof’) at the beginning of the 20th century, that some coherence could be achieved (Twining, 2011). Wigmore extended Bentham’s belief by suggesting a ‘chain of plausible inference’ and developed transparent methods to structure arguments from complex sets of evidence and their relevance, probative force, and credibility (Schum, 2009; Wigmore, 1937), without having them governed by legal rules (Twining, 2011). However, throughout the 20th century, his ideas of developing principles of proof were barely developed, some of the possible reasons being that the development of the forensic sciences took place outside of legal schools, that interdisciplinary work was unpractical, and the belief that ‘just common sense’, rather than formal instructions, was enough for fact-determination (Twining, 2011). Subsequent scholars including Twining and Schum studied the existence of a science of evidence, highlighting, among many things, the difference between the evidence and the actual events reported by the evidence and the relevance of structured arguments. They did not only aim to make inferences from large amounts of information, but also focussed on the teaching aspects of evidence in order to increase the number of rational decisions made by fact-finders (Hastie, 1993; Schum, 1994; Twining, 2003). Moreover, they argued that the complex and multi-disciplinary study of the science of evidence should be recognised as its own domain, and advocated for the involvement of other domains in the process of inferential reasoning, including logic, epistemology, sociology, psychology, and when studied in the legal domain, law and the forensic sciences (Murphy, 2007). Ultimately:

“Evidence never speaks for itself, but has to be interpreted through the filters of models, assumptions and analyses. Generic attributes of evidence include accuracy, credibility, objectivity, relevance, provenance and weight. One item of evidence may corroborate another, or conflict with it, or explain away its apparent message. Items of evidence and hypotheses can form complex interrelated chains or webs, outstripping unaided human comprehension. If there were such a thing as a general theory of evidence, it would have to explicate and analyse such issues.” (Dawid, Twining, and Vasilaki, 2011, p.3)

The ideas of Wigmore (1973) have since been taken on and developed into the argumentative approach to reasoning, where arguments are provided for and against propositions that could have happened (Schum, 2005; Verheij, 2014). Additionally, it has been

suggested that the processes related to quick and intuitive reasoning on the one hand and slower and conscious information processing on the other (Kahneman, 2003) are closely related to two other fact-finding processes often discussed: the probabilistic model and the so-called story model respectively (Spottswood, 2013). A probabilistic approach requires a continuous updating of beliefs in hypotheses when new information becomes available until a final belief is reached. In contrast, the story or explanation-based approach allows for a less restricted approach, by allowing for the heuristic development of stories and determining which explains the evidence best over-all (Hastie, 1993; Pennington and Hastie, 1992; Spottswood, 2013; Winter and Greene, 2007).

### 2.2.5 *Judicial decision-making research*

The three often cited judicial decision-making models including the legal, attitudinal, and rational choice models have been used to study and discuss the impacts of, among others, legal, strategic, and political preferences of judges (Segal and Spaeth, 2002). Where researchers within the attitudinal model have focussed mainly on the predictive aspect of the ideology of judges and, at a later stage, its effect on how they responded to different case facts, others have focused their research on judicial choice with respect to making rational choices and focus on strategic decisions which support their ultimate goal (Baum, 2008; Maveety, 2003). More specifically, it has been shown that within multimember courts, gender, race, and partisan-based effects of collegiate judges may overturn ones individual belief (Quinn, 2012). It must be noted, however, that others have expressed concerns that such specific studies must be interpreted and communicated with caution, and that it must be specifically acknowledged that there is a difference between moral or political reasoning processes which are intrinsic and extrinsic to law (Edwards and Livermore, 2009). The focus of the remainder of this section is not on personal beliefs and preferences impacting case outcomes, but on the way in which jurors and judges process information throughout a court case and how this relates to decisions following the presentation of forensic evidence.

Understanding these reasoning processes is important, especially because it has been shown to contribute to the interpretation process. For example, if evidence was presented first by the prosecution followed by the defence, individuals were less likely to convict in the step-by-step (statistical) approach, with little impact of whether background information was available (Kerstholt and Jackson, 1998). In contrast, following an end-of-sequence (story) approach, a recency effect was observed given the presence of background evidence, while a primacy effect appeared in cases without this information (Kerstholt and Jackson, 1998). Others have also observed that the presence of background information with ambiguous evidence (Smith, Bull, and Holliday, 2011) and the order in which it is presented (Spottswood, 2015) influences the weight that this evidence is ultimately given. Moreover, some findings dispute underlying assumptions of the hu-

man ability to weight evidence independently (Greene et al., 2002) and assumptions that jurors' behaviour does not conform to principles of probability (Winter and Greene, 2007), which is in line with the findings that jurors do often adopt the story model (Tinsley, 2001). Despite these descriptive observations, some have found that simple statistical methods perform better than experts in making judgements (see, for example, clinical examples by (Meehl, 1954)), and can prevent common reasoning mistakes such as the prosecutors fallacy (Fenton, 2011). Moreover, the decision-making process of juries often lacks a systematic structure to understand the evidence, partly due to the confusing manner in which it is presented and their inability to comprehend technical evidence (Tinsley, 2001). Crucially and on top of this, jurors have been shown to give little weight to whether the evidence has been scientifically validated or not (Koehler et al., 2016), where the credentials of an expert have shown to affect the weight given to the evidence (Koehler et al., 2016), and even becomes a key factor when complex evidence is involved (Cooper, Bennett, and Sukel, 1996). On top of the effect of such reasoning processes, research has also focused on more case specific details and their effects on jury decisions in England and Wales. For example, Thomas (2010) has shown that there is no evidence of discrimination in jury verdicts for black and minority-group defendants. Neither has this number disproportionately increased following the increase of these groups in the general population, nor is there a substantial difference of such relative inequality when facing different offence types (Thomas, 2017).

Regarding judges, it has been shown that they face biasing factors in their truth-finding process, suggesting that they tend to favour evidence confirming their hypothesis (Plous, 1993). Although, arguably, this effect could increase down the path of a step-by-step approach, it highlights the need in each belief-updating step to acknowledge and possibly consider alternative hypotheses. Given the understanding discussed in this section that jurors often impose a story-model in their reasoning (Pennington and Hastie, 1992), and that judges apply a lot of other factors indirectly related to the case in question, it must be understood to what extent their decisions are close to an ideal decision given the available evidence. It is believed here that scientific experts and statements must adapt themselves to restrict the ways in which results can be implemented beyond their added value of information in a legal context. On the other hand, judges must stay critical of expert evidence presented to avoid giving it too much weight in their judicial fact-finding process (Bell, 2010). Although the stereotypical understanding of the sciences and the law being cultural contradictions has been opposed (Roberts, 2013), their different nature is often mentioned when articulating the challenges within making the best use of evidence (Chisum and Turvey, 2011; National Research Council, 2009).

### 2.2.6 *Science and the law: negotiating a partnership*

When forensic findings are used in a court setting, it sometimes obliges experts to draw case specific inferences (e.g. individualisations (Cole, 2013)) from a limited knowledge base, a process (even referred to as a ‘leap of faith’ (Aitken and Taroni, 2004)) which is often not supported by scientific theories or gained solely through logical reasoning. It has been argued that the law is unable to sufficiently embrace the changes that take place within the forensic science domain through research and development (Sperling and Cooper, 2013). Additionally, one of the reasons why the limitations of scientific evidence are lost in the communication process is a limited understanding, by both the courts and practitioners, of the meaning of error in forensic analyses (Christensen et al., 2014). This includes the extent to which scientific errors, methodological limitations, general uncertainties, and mistakes, influence research outcomes (Christensen et al., 2014). This also becomes evident when looking at the way results are communicated by the scientists, in a rather vague verbal manner including the use of terms such as “matching” evidence, “similarities”, and an object being “probably the source” (Cole, 2013)(see Section 2.5).

Another aspect which is often argued to play a significant role, especially in jury-systems, is the so-called CSI-effect, which describes the idea that individuals have a beyond-realistic expectation on what can be achieved with forensic evidence. However, many observations related to this effect are anecdotal in nature or follow the beliefs from legal experts observing changes within juror decisions after the start of crime shows (Robbers, 2008). However, systematic studies have not shown a significant difference in convictions pre- and post- CSI era (Cole and Dioso-Villa, 2009) or between viewers and non-viewers (Podlas, 2005), and research has suggested that those familiar with such media are actually more sceptical towards forensic evidence (Schweitzer and Saks, 2007).

Many of the human aspects of legal decision-making will continue to play a role in legal decisions, although examples have shown the use of algorithms in different stages of the criminal justice process, including risk-based sentencing and custodial decisions (Kehl, Guo, and Kessler, 2017). These approaches can perhaps overcome the misuse and biases related to the interpretation of forensic evidence, although many of these algorithms are based upon ‘biased’ previous information, and crucially, a possibility of losing transparency and the ability to justify legal arguments. There is perhaps a better equilibrium to be found. Within the forensic science discipline, however, the use of advanced data analytical methods is growing, including the interpretation of biological and chemical analyses, which are further studied in Chapter 5. The key to applying forensic knowledge into the legal field is first of all to make valid and justifiable arguments, and second to present these in an understandable and transparent manner. Although the legal systems have safeguards in place to prevent the misuse of evidence,

this does not mean that it is their full responsibility. Legal actions such as legal directions or admissibility standards are not in the hands of the sciences but can be guided by it. Ultimately, by understanding each other's questions, aims, and limitations, it should be feasible to "negotiate successful partnerships between science and law, rooted in a shared, institutionalised culture dedicated to the ends of justice" (Roberts, 2013, p.57). In order to achieve this, there is a need to understand how far (forensic) science has come and how it developed itself throughout history to understand how courts can be best equipped in order to make fair and justifiable decisions.

## 2.3 HISTORICAL DEVELOPMENT OF THE FORENSIC SCIENCE DISCIPLINE

### 2.3.1 *Pragmatism*

Much has been written about the history of the forensic sciences and the cases and individuals that shaped its development. This section aims to discuss the extent to which previous observations and shortcomings can help shape the future of forensic science. Many anecdotal examples exist on the early days of 'scientific' observations in criminal investigations, mostly led by the pragmatic way it was used: from the use of forensic medicine and entomology in 13th century China (Sung, 1981) to the methodologies by Alphonse Bertillon to identify offenders based on physical measurements (Bertillon, 1893). In the beginning of the 20th century, the use of physical traces grew, arguably developed from the notion of uniqueness following the idea that "nature never repeats itself" by Quételet in the 19th century (Cole, 2001). For example, in addition to physical measurements, feature consistencies were used to interpret observations including bullets and firearms (Goddard, 1926), tool marks (May, 1930), and documents (Osborn, 1946). Many of those early identification fields resulted from practical needs by police and prosecutors rather than they were developed from, and validated by, conventional sciences such as in DNA-analysis (Saks and Faigman, 2008).

Scientists in the mid- and late 1800s and early 1900s such as surgeons Bell (1837–1911) and Doyle (1859–1930) started concerned themselves with the 'generalist view' of crime reconstruction and the philosophies that entail this, acknowledging the importance of observations and reasoning methods that were led by an evidence base (Chisum and Turvey, 2011). Similarly to Doyle, Gross (1847–1925), trained in the legal domain and therein experienced the unreliability of suspect-, eyewitness-, and victim accounts, emphasized the importance of scientific objectivity (Gross, 1906). As Gross articulated:

"A thousand mistakes of every description would be avoided if people did not base their conclusions upon premises furnished by others, take as established fact what is only possibility, or as a constantly recurring incident what has only been observed once." (Gross, 1906, p.22)

Locard (1872–1966), a scholar in medicine and law and encouraged by the ideas of Gross and Conan Doyle (Locard, 1930), found that at the beginning of the 20th century, none of the police laboratories he had visited across the United States and Europe had adapted scientific methodologies in crime reconstruction, and instead continued to use, for example, the (subjective) personal identification methodologies of Bertillon (Chisum and Turvey, 2011). His research therefore aimed at writing practice standards and methodologies, mainly in trace evidence and fingerprint examination (Locard, 1934). Kirk (1902–1970), a trained biochemist and author of the widely cited book *Crime Investigation* in 1953, was one of the first who specifically highlighted the importance of distinguishing between identification and individualisation (Kirk, 1953), where his thoughts were taken on as the principle of individualisation (Saks and Faigman, 2008). Although important notions, the ideas developed during these times on trace evidence dynamics and individualisations have sometimes been doubtfully applied, challenging the scientific scrutiny of arguments (see Chapter 5).

### 2.3.2 *The post-DNA era*

As the 19th century further progressed, scientific methods to underpin statements in an investigative and court setting significantly increased and fundamental changes took place in the forensic domain, enabled by methods such as DNA sequencing (Wilson et al., 1995), (bio)chemical analytical techniques including chromatography and spectroscopy (Kher et al., 2006; Maurer, 2005), the development of databases for evidence types such as DNA, shoeprints and drugs (Martin, Schmitter, and Schneider, 2001; Mikkonen, Suominen, and Heinonen, 1996; Perkal, Ng, and Pearson, 1994), and experimental studies to examine the behaviour of evidence (Morgan et al., 2009). Specifically, the growth of datasets containing feature frequencies in populations has continued to make a significant impact on case outcomes, such as the recent creative use of commercial DNA databases led to the arrest of the ‘Golden State Killer’ (The Sacramento Bee, 2018) and a mass DNA search led to the solving of a cold case through familial DNA searching (NFI, 2012).

In the 1980s and 1990s, the development within DNA analysis was quickly associated with the potential to expose miscarriages of justice in cases where biological material was available post-conviction (Cole, 2012), in, for example, sexual assault cases where initially only serological tests were performed. Paradoxically, these DNA analyses showed that forensic evidence could, in addition to being exposers, also be ‘contributors’ to why innocent individuals were convicted in the first place. Many of the early studies into the wrongful convictions that followed all concluded in the same way: leading factors include mistaken eyewitness identification, false confessions, guilty pleas, flaws surrounding forensic statements, and ignorance to exculpatory evidence (Garrett, 2015; Gross and Shaffer, 2012; Leverick and Chalmers, 2014). An extensive and systematic

study of misleading evidence can be found in Chapter 3, showing that the emerged issues are not just related to overstating findings with regards to attributing a sample to a specific source based on features, but are also related to an ignorance towards activity level hypotheses.

Introduced as the hierarchy of propositions in the late 1990s (Cook et al., 1998a) was the idea of defining propositions beyond a source level (e.g. ‘the glass sample on the suspect originates from the broken window’) to include an activity (‘the suspect broke the window’) and/or offence level hypothesis (‘the suspect burgled the house’). Where traditionally forensic experts would only concern themselves with the first proposition and let the Trier-of-fact deal with the other two, it has become clear in the last decade that a complete (new) body of knowledge is required on understanding evidence dynamics in order to address activity level hypotheses, which the scientific community should take responsibility for. With the growing acknowledgement that the significance of DNA evidence requires an assessment of activity level hypotheses, a quickly growing number of transfer and persistence studies have made their way into the forensic knowledge base (Meakin et al., 2015; Raymond et al., 2008). Arguably, the growing capabilities of extracting features from smaller traces make understanding these processes even more important. Studies on evidence dynamics within forensic domains more generally have been conducted as early as the 1970s on fibres (Pounds and Smalldon, 1975b), and later on traces including glass (Hicks, Vanina, and Margot, 1996), GSR (Fojtásek et al., 2003), paint (Buzzini et al., 2005), pollen (Morgan et al., 2014), diatoms (Scott et al., 2014), and fragrances (Gherghel et al., 2016). A more detailed scientific perspective on understanding the processes involved in evidence dynamics is discussed in Chapter 4. Despite the increased understanding that evidence dynamics play an important role in establishing the value of evidence for an hypothesis, results from Chapter 5 from a study that addresses the way in which these aspects are considered in practice, show that they are often ignored or their effects assumed to be diminished.

### 2.3.3 *The rise of governmental concerns*

Whilst specific studies were performed to develop a specific forensic evidence base (which includes both scientific evidence and expert knowledge based on experience), and shortcomings were slowly being recognised and addressed, critique has continued to grow. US and UK governmental reports highlighted that the forensic community over the years has not always seemed to succeed in translating the growing knowledge base into scientifically validated and reliable statements in court (National Research Council, 2009; The Law Commission, 2011). One of the key underlying reasons has argued to be the inability to interpret evidence in the light of growing analytical developments (Government Chief Scientific Adviser, 2015).

Where individual cases came to light following post-conviction DNA-testing, some large-scale issues have emerged over the years. In the US, these include the reveal of significant systematic errors in microscopic hair comparisons conducted by the FBI (FBI, 2015), issues with 180 cases of serology and DNA analysis in the Houston Police Department Crime laboratory between 1980 – 1992 (Bromwich, 2007), a lack of foundational scientific underpinnings and flawed testimonies in the compositional similarity of bullet-lead (National Research Council, 2009) and bite-mark features (PCAST, 2016), and intra- and inter-laboratory variations when interpreting DNA mixtures (Butler, Kline, and Coble, 2018). Individual cases receiving significant attention included the erroneous decision of linking Brandon Mayfield to the Madrid train bombing in 2004, arguably due to confirmation bias (Office of the Inspector General, 2006). A critical report by the US National Research Council discussing many of these cases in 2009 highlighted a number of concerns within a range of forensic (identification) disciplines, criticizing the validity of findings and, specifically, the association of traces to a specific source based on feature-comparisons (National Research Council, 2009). The PCAST report followed up on this initial report, specifically focussing on evaluating standards of foundational scientific validity and standards of how this has been applied practice (PCAST, 2016). However, the National District Attorneys Association argued that the PCAST report lacked in consulting adequate experts and that increasing the threshold of validity would hamper investigative processes (NDAA, 2016). Although recognition of the need for “empiricism, transparency, and a commitment to an ongoing critical perspective” has continued to be discussed mostly within the academic community (Mnookin et al., 2011), a challenge remains to recognise the need for such a research culture in every aspect of forensic reconstructions.

As presented in Chapter 3, concerns have also been raised in England and Wales in regard to the misinterpretation of evidence in many criminal cases, and the magnitude of this issue. In addition, large scale and systematic mistakes have surfaced more recently, such as the review of thousands of drug cases after possible deliberate data manipulations at Randox Testing Services (The Guardian, 2017b) and the possible misconduct of a scientist at the Metropolitan Police Service affecting over 30 cases (The Guardian, 2018b). Where the Law Commission of England and Wales argued for better admissibility standards and guidance for judges which were largely not incorporated at the time (Section 2.2.2), the UK government has recognised the need for ensuring scientific validity and good conduct. This is especially important given the closure of the Forensic Science Service in 2012, which was accompanied by a range of concerns including those on the implications for research, innovation, and criminal justice (House of Commons, 2011). In 2018, the UK Forensic Science Regulator, equipped with maintaining quality standards in forensic science, has warned that a shift in the general developments within the forensic science discipline due to the competitive commercial market and for difficulties in interpreting evidence due to analyses and interpretation of evidence being fragmented

(Tully, 2018). Moreover, the House of Lords Science and Technology Committee has just launched an inquiry into the strengths and weaknesses of the contribution of forensic science to the delivery of justice (House of Lords, 2018).

These governmental responses and discussions highlight that a fair and reasonable justice system has their widespread interest, with a growing interest in the (negative) role of the forensic sciences herein. The question, however, remains whether they could have been known about and minimised in the first place. This is not just important to know in order to rightly handle such past cases but also to consider ways of avoiding these in the future. One of the key factors in this is the ability to anticipate the growing understanding of uncertainties and how this is developing in the way that forensic reconstructions are approached, undertaken, and communicated.

#### 2.3.4 *An increased understanding of uncertainties*

Through a consideration of the law and historical aspects, it can be argued that there have been several conceptual aspects within forensic interpretations of which their importance is being recognised over time. Two of these aspects are discussed in this section, and include the notion of individualisations on a source level and the ability to incorporate activity-level hypotheses, and, secondly, the role of human judgement and decision-making in the forensic reconstruction process, as the development and application of knowledge will always, to a certain extent, involve human interaction.

Throughout history, a focus within the forensic sciences has been on the development of knowledge consisting of the occurrences of measurable characteristics in a population, from the physical measurements by Bertillon towards much more advanced possibilities of CT imaging to establish population characteristics (Giurazza et al., 2012). Similarly, within the interpretation of physical traces, such knowledge on population frequencies can come from commercial databases when dealing, for example, with shoe patterns or tire tracks. This can also be established from smaller scale studies to facilitate specific case contexts, such as establishing the background presence of glass on clothing (Coulson et al., 2001) or cocaine on banknotes (Wilson et al., 2014). With a growing knowledge base on specific populations, extreme care must be taken when, for example, feature frequencies, are generalised and used in contexts which differ in nature from the context in which the data has been collected. Moreover, the issues emerging from a consideration of individualisations originate from the notion of uniqueness, which has even been recognised by the International Association of Identification as a foundation to make the analysis “scientifically accepted and legally defensible” (International Association for Identification, 2007, p.1). Additionally, and possibly even more concerning, uniqueness has been the underlying argument for the accuracy of latent print analysis, a statement often referred to as the fingerprint examiner’s fallacy (Cole, 2009). The main

issue here is that in practice, rigorous claims of scientifically valid individualizations based on common characteristics are made in expert statements and court. However, from a scientific perspective, such a step (from a certain level of identification to an individualisation) is seen as a decision making process rather than as a direct conclusion (Garbolino, 2013), and therefore, some statistical decision theory should be applied (Champod, Evett, and Kuchler, 2001; Inman and Rudin, 2000). Saks and Koehler raise an important question, asking “how did the practice get so far ahead of the science?” (2008, p.210) and conclude that the current concept of individualisation has no scientific validity due to factors including a lack of data, the absence of random match probability estimates and unsupported claims of uniqueness (Saks and Koehler, 2008). It seems that the consideration of Kirk that individuality is limited by the present state of science (Kirk, 1963), in which ‘the science’ could be interpreted as both the science of differentiating between objects based on certain characteristics and the science of actually observing and analysing these characteristics, is often neglected.

Thus, while acknowledging inherent uncertainties, the more well known statements of Kirk such as “criminalistics is the science of individualization” (Kirk, 1963, p.236) together with the knowledge that, to this day, jurors have been shown to interpret an “identification” as belonging to an individual to the exclusion of all others (Swofford and Cino, 2018), captures the challenges forensic science continues to face. Others have argued that, if rightly questioned, Kirk’s principle of individualisation and Locard’s Exchange Principle could be used as ‘scientific laws’ whereby adding knowledge and thereby supporting legal reasoning. However, it has also been suggested that practitioners ignore or explain away falsifications (Saks and Faigman, 2008), suggesting that such ‘laws’ in forensic science do not undergo the scrutiny of ‘scientific laws’ in the conventional sciences. The idea that they have been falsified apparently does not, in this case, reject these laws (this is further discussed in Chapter 4). Perhaps, rather than presenting an ultimate answer and accompanying this with underlying limitations and assumptions, findings should be presented in a less idealistic manner, whilst preserving their scientific validity. Moreover, the notion of evidence dynamics has been increasingly studied in a number of forensic disciplines (Morgan et al., 2014; Raymond et al., 2008; Wilks, Morgan, and Rose, 2017). Where the variables and circumstances of such studies greatly differ, the results of those studies are no different a dataset of information than more structured datasets that follow from studying population frequencies. It would therefore be extremely valuable if the results from any findings within the domain would be presented and collected in a coherent and transparent framework to fully utilise the past knowledge base on build onto this in the future. This will be further discussed in Section 2.5.

Another growing point of focus in the past decades is how one should reason from observations. Introduced in Section 2.2.5 on judicial decision-making, the study of hu-

man reason has made its way into the forensic sciences. The process of understanding observations can, especially in subjective and/or ambiguous situations, become rather complex, which is why the human brain relies upon heuristic principles ('mental shortcuts') to reduce the complex reasoning task and quicken the reasoning process (Tversky and Kahneman, 1974). Even though such quick reasoning processes can be valuable in many situations, it makes individuals prone to unintentional reasoning errors, referred to as cognitive biases, which have shown to significantly affect expert' judgements, specifically when decisions are made under uncertainty (Dror and Rosenthal, 2008). For example, experts have shown to adjust their belief under contextual information, both in domains which rely upon methods with subjective aspects, including bite mark analysis (Page, Taylor, and Blenkin, 2012) handwriting analysis (Found and Ganas, 2013), and the analysis of skeletal remains (Nakhaeizadeh, Hanson, and Dozzi, 2014), as well as more objective fields including the interpretation of DNA evidence (Dror and Hampikian, 2011) and latent finger mark analysis (Dror et al., 2011). The growing body of research, together with the acknowledgement of their effects in previous erroneous convictions (including the previously mentioned Brandon Mayfield case (Kassin, Dror, and Kukucka, 2013)) in the forensic domain, it highlights, more generally, that many aspects within the forensic process can be prone to errors in which individuals (be it forensic scientists or legal professionals) are required to reason under (any sort) of uncertainty. To address the significant issue of heuristic effects in the forensic domain, it requires not only further research into their occurrence and effects on judgements (at all levels of the hierarchy of propositions), but simultaneously needs to be addressed by improving the acknowledgements of cognitive effects, and the development of reasoning frameworks to guide the reasoning processes. An example of such a framework is a Bayesian network, discussed in Section 2.4.3.

Where generally, the idea of acquiring knowledge and expanding ideas on how physical processes in the world might work can only lead to doing better in the future, the forensic science domain is a perfect example of the struggles and 'mistakes' which have to be faced before scientific advancements can fully be celebrated. Moreover, it is perhaps the excitement and applicability of forensic evidence through history which sometimes led to its rapid use rather than that they were scrutinised from the perspective of scientific standards and principles. In supporting the knowledge and progress which has been achieved so far, but whilst recognising the importance of establishing justifiable beliefs from such arguments, the next section will address the ways in which the forensic knowledge base can, and should, be used as in the reasoning towards beliefs in hypotheses and making fair and reasonable decisions from those beliefs.

## 2.4 REASONING PERSPECTIVE

In this section the interpretation of forensic evidence from a (logical) reasoning perspective is explored, that is, the drawing of valid inferences from observations towards conclusions. Several themes on the notion of reasoning are discussed, including evidence, hypotheses, the reasoning between these two, uncertainties, fallacies, and the use of Bayesian networks as a specific example on inference methods. The question which will be addressed is, in an idealistic world, how fair and reasonable inferences can be achieved within the forensic domain. Section 2.5 builds upon this exploration by examining the extent to which the knowledge required in making these arguments can (ever) be obtained and constructed into frameworks.

### 2.4.1 *Truth, beliefs, and reasoning*

The aim of any investigation or enquiry is to reconstruct past events, with the specific goal within the criminal justice system to guide subsequent decision-making at stages such as a crime scene, during analyses, and in a court of law. In order to achieve this, the current and past states of reality must be conceptualized. One of the ways of doing so is by forming statements (e.g. 'the earth rotates around the sun'), which are the result of reasoning processes from a set of premises. The reason why the representation of the reasoning processes of forensic experts is incredibly important is that it allows for the formalisation, and thereby the validation, of both the argument and its premises. Within the literature, inductive, abductive, and deductive processes have been introduced to reach conclusions from premises (Magnani, 2001; Pearl, 1988; Schum, 1994).

The first and classic cause-effect approach applied in scientific research, referred to as inductive reasoning, represents the process where some general laws are inferred from specific observations (Pearl, 1988). Secondly, abductive reasoning methods can be used to infer a (most likely) set of premises (causes) that can explain some observations made (the evidence) (Magnani, 2001), or in other words conduct "inference to the best explanation" (Coltheart, Menzies, and Sutton, 2010), see Figure 2.2. Moreover, despite the common belief that practical evidence-based reasoning processes (including the famous thoughts of Sherlock Holmes) are employing a third, deductive, inference method, this reasoning involves a logical implication of the conclusion, which can only be the case if no other alternatives are possible (Coltheart, Menzies, and Sutton, 2010). Within the complex and uncertain processes of crime reconstruction, such absolute claims with regards to the evidence can rarely be made and so will not be addressed here.

With respect to the forensic science discipline, inductive reasoning processes support the development of a general forensic knowledge base following the results of experimental studies and observations, and abduction reasoning processes can support de-

giving some explanations for observations based on this knowledge base. The extent to which such knowledge has been and can be developed is discussed in Section 2.5. Nevertheless, accepting that the absolute truth of statements is often uncertain, this thesis will use the more nuanced term belief, highlighting the difference between what is ultimately true and the extent to which such truths can be approached. Moreover, it highlights that these beliefs are situational in that it may differ depending on the information available to the mechanism expressing the belief (Lindley, 2014). The various ways of reasoning from evidence to beliefs has been introduced in Section 2.2.5, discussing the argumentative, narrative, and probabilistic-based methods within judicial decision-making. These ideas can be extended to many other stages of the criminal justice process and within the forensic chain, including at the crime scene, during the evidence submission processes, and in the analysis and interpretation of observations. For reasons discussed in the next section, the remainder of Section 2.4 focuses on the use of probabilistic approaches.

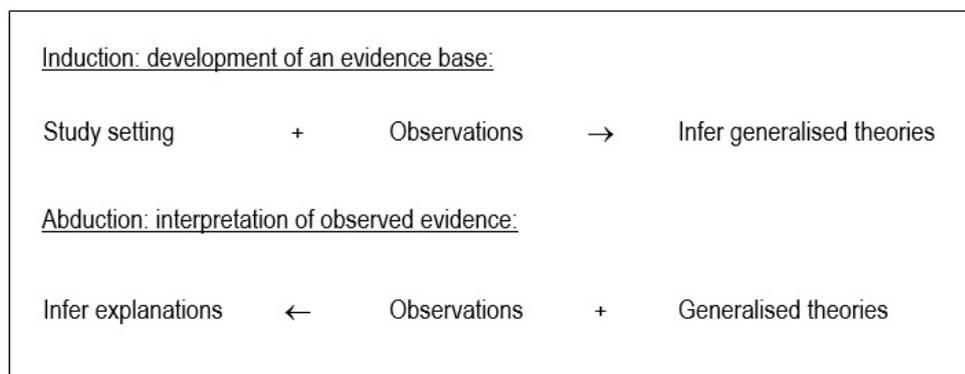


Figure 2.2: Summary of inductive and abductive reasoning processes

#### 2.4.2 Probabilistic-based reasoning

Although the use of probabilistic approaches in a legal setting have faced drawbacks mainly with regards to the use of arguably 'subjective' probabilities (see for a review of this discussion Prakken, 2014, Keppens, 2014, and Thompson, 2012) and the general difficulties in understanding such evidence (see Section 2.2), these approaches have become a more standardised way of interpreting findings by forensic experts (Asten, 2014). The position is taken in this thesis that the use of probabilities is unavoidable when aiming to understand, represent, and to a certain degree, minimise, levels of uncertainty in evidential reasoning. By acknowledging and representing uncertainties, assumptions, and dependencies, it hopes to shift the focus from the critique to this approach towards the possibilities of improving the understanding and use of inferential reasoning under some of the known (unavoidable) limitations. Moreover, in anticipation of the growing knowledge base (see Section 2.5), such methods would allow for drawing inferences from this complex body of knowledge.

Within the probabilistic approach to reasoning, there are two general viewpoints that can be taken when aiming to express some belief about (competing) propositions: the frequentist and Bayesian approach (Fienberg and Schervish, 1986; Taroni, Biedermann, and Bozza, 2016). The aim of this section is to introduce both approaches rather than to continue the (extensive) documented discussion between the two (Bland and Altman, 1998). In short, a frequentist considers the probability of a new observation as the long-term frequency of repeated observations, while to a Bayesian the probability of an event is the belief that it will occur (or has occurred), in the light of the previous (prior) belief and some new observations (E), resulting in a posterior probability (Curran, 2013; Taroni, Biedermann, and Bozza, 2016), expressed as:

$$P(H|E)$$

More specifically, it is possible to compare alternative explanations put forward (e.g. those put forward by the prosecution and defence), by expressing the ratio between the beliefs described by the posterior odds. The odds form of Bayes' theorem describes how this ratio can be obtained:

$$\frac{P(H_1|E)}{P(H_2|E)} = \frac{P(E|H_1)}{P(E|H_2)} \cdot \frac{P(H_1)}{P(H_2)} = \text{likelihood ratio} \cdot \text{prior odds}$$

Thus, Bayes' theorem is simply used to update the previous belief in hypotheses when some new information becomes available, of which the evidentiary value is described by the likelihood ratio (LR), to obtain the (new) posterior odds. By performing empirical studies, the LR can be determined (for those specific experimental conditions), which describes how much more likely it is to find the evidence under one proposition over the other. Within forensic practice, such hypotheses can be formulated for anything between sub-source and offence level.

Bayesian inference (the application of Bayes' theorem to update beliefs) in the forensic domain started to become more widely researched in the mid-1980s as a tool to support combining sources of information and highlighting the interaction of different aspects of a single item of evidence (Evetts and Buckleton, 1990). Work by Lindley and Evetts et al. describe the early stages of using solely a LR approach to compare characteristics of transferred glass (refractive index) and fibre particles (colour) (Evetts, 1984, 1986; Evetts and Buckleton, 1990; Grieve and Dunlop, 1992; Lindley, 1977) and the development of probability functions to describe this data (Evetts, Cage, and Aitken, 1987). Further work includes the within variability, prior distribution and transfer of fibres (Champod and Taroni, 1997; Wakefield et al., 1991). Currently, the determination of LRs in the interpretation of evidence has been a growing and widely discussed topic in published literature and forms in many cases the standard for presenting and evaluating expert opinion (Association of Forensic Science Providers, 2009).

However, the use of the Bayesian approach (used here to represent the application of Bayes' theorem) in forensic practice is facing a range of challenges. Firstly, in order to

avoid missing alternative explanations, there is a need to assess mutually exclusive and exhaustive propositions in the light of the evidence (Gould et al., 2013). However, these can often be very generic (e.g. 'someone else was the source of the trace'), making it difficult to assess the probability of observing the evidence under this hypotheses (Evetts et al., 2000), requiring the need to limit this to more accessible sub-populations. In relation to this, the LR can greatly differ depending on how the hypotheses are phrased, possibly causing evidence to be misinterpreted if this is not conveyed transparently (Meester and Sjerps, 2004). Moreover, in order to ease the interpretation for legal professionals, such LRs have been converted into a verbal scale to indicate the support of the evidence in court. Although the perceived strength of evidence based on a verbal scale has shown to be misunderstood by lay people (Mullen et al., 2014), it has been argued to be currently one of the most appropriate ways to communicate expert opinion (Aitken et al., 2011).

Additionally, one of the major points of debate regarding the Bayesian approach has been the need to establish a probability distribution representing the prior belief in hypotheses. Although much has been said about this, many conclude that the establishment of such a prior belief should follow case-specific information and logically justifiable arguments (Biedermann, Taroni, and Garbolino, 2007). This should arguably be the role of individuals dealing with the expert statements (such as the Trier-of-fact) rather than it being the role of the expert itself (Jackson et al., 2006). However, it can be questioned whether the Trier-of-fact is able to bridge the gap between probabilistic-based beliefs and their own need to establish guilt (e.g. beyond reasonable doubt)(Ligertwood and Edmond, 2012). An example of such a difficulty include the so-called prosecutors fallacy, where significant misunderstandings happen as the likelihood ratio is interpreted as being the posterior odds without considering any prior beliefs (Thompson and Schumann, 1987). With the difficulty of understanding and interpreting probabilistic evidence by lay people and legal professionals, this challenge will remain a topic for debate and requires a standardised and understandable approach to avoid misinterpretations.

Lastly, with a growth in complexity and dependencies within a case, such approaches are not intuitively applicable to calculate the joint probability distributions of the complete set of evidence in evaluating hypotheses (Pearl, 1988). With the discussion surrounding the best process by which this should be achieved is still on-going (Taroni, Biedermann, and Bozza, 2016), one of the ways to address some of these issues is by using a graphical representation of these probabilities, such as in a Bayesian network (Dawid and Evetts, 1997), especially with the rise in computational power (Keppens, 2014).

### 2.4.3 Bayesian networks

There are several steps involved in the use of Bayesian networks (BNs) to support the interpretation of evidence. First, it requires the development of the network structure, consisting of nodes and relationships between them representing the causal mechanisms between hypotheses and evidence: some evidence can be observed at a scene because some action (possibly related to the crime) caused this to be there. As the number of hypotheses and items of evidence increases, the structure of the networks may become unmanageable, especially when applied to complex and real cases (Fenton and Neil, 2000; Hepler, Dawid, and Leucari, 2007). The use of object-oriented Bayesian networks has therefore been proposed, reducing the network structure into multiple interrelated causal structures of a lower dimension (Hepler, Dawid, and Leucari, 2007), which can be formalised as idioms (Lagnado, Fenton, and Neil, 2013) in order to apply them to other, related, problems. This makes the process of constructing Bayesian networks transparent and logical, and allows for structured process improvement. Nevertheless, it is still crucial that more attention should be paid to the mechanisms of constructing Bayesian networks, providing standardised and accessible frameworks (Taroni et al., 2004; Taylor et al., 2018).

Second, although developing the network structure is not a straightforward task, the most challenging aspect of Bayesian networks is acquiring the quantifications for the relationship between nodes (Druzdzel and Gaag, 1995). For example, given the causal relationship between the proposition that the suspect was (somewhere else than) the crime scene and the evidence relating to soil evidence, probability distributions should provide insight in the following questions:

- what is the probability of finding (a certain level of) similarities between soil from the crime scene and soil from the suspects' shoe given that he was at the crime scene, and,
- what is the probability of finding (a certain level of) similarities between soil from the crime scene and soil from the suspects' shoe given that he was somewhere else than the crime scene.

These questions are essentially addressed through inductive reasoning processes (see Figure 2.2), given a study setting and some observations. These can be the result of generic or case-specific experimental studies, can follow from databases of population frequencies, or can be provided through expert experience. In a real case, following theories encapsulated in Bayesian networks and the presence of case specific observations, the joint probability over all nodes and thereby some most-likely explanations (posterior odds in hypotheses) to explain the presence of these observations can then be made (see Figure 2.2). The aim of using such a network in the interpretation of forensic trace evidence is not necessarily to solve problems algebraically, but mainly to provide a frame-

work to evaluate the evidence. Furthermore, it can act as a starting point to approach complex interpretation issues such as mixtures, missing data, and secondary and/or multiple transfer (Curran, 2013). Some of the early applications of Bayesian networks to forensic problems include the modelling of all offence and activity level hypotheses in the real Collins and Sacco & Vanzetti cases (Edwards, 1991; Schum, 1999), as well as source and activity level hypotheses of fibre, blood, and glass evidence (Aitken and Gamerman, 1989; Curran et al., 1998; Dawid and Evett, 1997; Garbolino and Taroni, 2002). Due to its applicability and its support to communicate findings, there has been a rapid increase in the published literature on the application of Bayesian networks to forensic inferences on (a combination of) source and activity-level hypotheses, of which some of the many examples include complex DNA-profiles (Dawid, Mortera, and Vicard, 2007), DNA cross-transfer (Aitken, Taroni, and Garbolino, 2003), fire incidents (Biedermann et al., 2005), and GSR-particles (Biedermann, Bozza, and Taroni, 2011). The general idea that follows from many of these studies is the possibility of representing and explaining complex interpretations, which possibly enables a better communication with decision-makers throughout the forensic chain.

#### 2.4.4 *Decision-making throughout the forensic chain*

Section 2.2.3 introduced the idea of legal decision-making, together with a range of thresholds which have been introduced in the legal domain describing normative processes of when decisions should be made (such as the criteria of belief beyond reasonable doubt). These are developed from the ideas of the consequences of, for example, acquitting those guilty and convicting those innocent. Although much has been said about the presentation of beliefs rather than decisions by forensic experts at a legal setting, they are required to make decisions continuously at other stages of the forensic process. Examples of these stages include a crime search strategy, the decisions as to which evidence types to collect from a crime scene and which of those are subsequently submitted to which laboratories, and which analytical methods to employ on the item of evidence and in what sequence. On the one hand, these questions are based upon decision thresholds, such as the determination of features in more subjective-based analysis processes. On the other hand, these can also require the decision based on the perceived added value of evidence such as in when deciding whether or not to collect a trace.

As a continuation of the Bayesian approach, a normative approach to decision-making in forensic practice has included so-called Bayesian decision analysis, using the belief in an hypothesis and the expected utility (Biedermann, Taroni, and Aitken, 2014). The formalisation and consideration of the perceived added value of the evidence can be used in the decision-process of whether or not to utilise the evidence (Bitzer et al., 2015). This idea has also been acknowledged in the late 1990s, when Cook et al. (1998) developed, within the UK Forensic Science Service, a general model to aid decision-making in an

operational forensic context. The evaluation of the questions that need to be addressed and the value of the evidence in this process is part of their so-called case pre-assessment stage (Cook et al., 1998b). Following the use of Bayesian networks, the added value of information and the relevance of individual items on evidence can be obtained following so-called sensitivity analyses. Moreover, in order to conclude that the evidence does not have a probative value ( $LR = 1$ ) and is therefore considered 'neutral', it is essential that the two propositions brought forward are both mutually exclusive and exhaustive (Fenton et al., 2014). This further highlights the importance of framing and communicating the propositions considered at every stage of the forensic process (Evetts, Jackson, and Lambert, 2000).

The decisions made following subjective analysis methods require a different approach. Just as individuals within the legal domain, forensic scientists do not always follow normative benchmarks either. The human role of expressing beliefs and subsequent processes of decision-making are prone to biasing effects, which are generally unwanted throughout forensic processes when aiming to make rational and scientifically founded claims. An introduction to these effects are discussed in Section 2.3.4, highlighting the growth of such research in a range of forensic domains. It is important to understand such descriptive processes, as (mis-)interpretations and wrongful decisions can cascade throughout the forensic chain, possibly leading to snowball effects where it may influence other lines of evidence (Dror et al., 2017). Therefore, it is extremely important that these decisions are also to some extent 'justified', including the need to justify 'inconclusive' decisions rather than utilising this as a simple excuse for avoiding this step (Dror and Langenburg, 2018).

This section has highlighted that the framework of forensic knowledge does not only consist of databases representing population frequencies of features and theories on physical processes such as transfer and persistence, but also on the growing body of knowledge representing the effect of human reasoning processes and the limitations of logical frameworks which may affect the ways in which evidence is interpreted, presented, and incorporated in legal decisions. The fourth perspective in this chapter therefore aims to make sense of this current and future body of knowledge.

## 2.5 KNOWLEDGE-BASED SYSTEMS IN THE FORENSIC SCIENCES

What follows from the legal, historical, and reasoning perspectives, is that the interpretation of forensic evidence should be based upon maximizing the use of objective observations and transparent communications, whilst acknowledging uncertainties and minimising the effect of unwanted biases. In other words, it requires an establishment of (the highest achievable) justified belief. Although one of the strengths of human reasoning is the capacity to combine information with different origins and varying degrees of

accuracy and probative value (Pearl, 1988), there are limits when it comes to interpreting complex sets of knowledge. Moreover, with a growing body of knowledge resulting from experience and experimental studies, it cannot be expected that an expert is able to incorporate all this when expressing belief in an hypothesis during their routine work. However, from a more research- and academic oriented view, it would not be unreasonable that with the current technological developments and push towards a cultural change within the forensic sciences (Mnookin et al., 2011), establishing an up-to-date, independent, and coherent body of knowledge is something to strive towards. Such a body would consist of an evidence-base containing, for example, published literature. Whilst this would constantly expand, the studies already incorporated in this evidence base will not change. In contrast, the knowledge base which can be derived from these studies will represent ideas and theories which follow from combining the knowledge in the evidence base. Therefore - such is the case with Locard's exchange principle - this might very well be adjusted when new information becomes available. In this thesis the term 'knowledge-base' will be used to represent both the evidence base as well as subsequent derived theories.

The conceptualisation of knowledge is certainly not new. In the context of computer systems to store and reason from knowledge, they have been referred to as knowledge-based systems (KBS) (Shortliffe, 1974). Rather than simple datasets with structured information stored in tables, the information in knowledge-based systems are conceptualised in more dynamic ontologies (Budin, 2005). The early knowledge-based systems were expert systems, providing inferences following specific questions which had to be answered by experts, such as the medical-decision support tool MYCIN (Shortliffe, 1974). One of the early specific uses of KBSs in the forensic domain included a knowledge base on casework conditions and observations regarding the classification of glass evidence (Evetts and Spiehler, 1987). However, there are also risks, specifically within casework, where updating the system with new findings (which are derived from this system) may make the system ignorant towards other knowledge (Aitken and Stoney, 1991). Nevertheless, when acknowledging such shortcomings, the major advantage of such a system in the forensic domain is the consistency between analyses and continuity of expertise (Aitken and Stoney, 1991). Moreover, it will be possible to explain the reasoning towards arguments from the knowledge used and apply knowledge across forensic domains.

Three phases comprise a KBS: knowledge representation, knowledge acquisition, and knowledge inference. Where there is a growing recognition within the forensic sciences that inference methods such as Bayesian networks are needed (Section 2.4.3), a solid understanding of the input knowledge in these systems is equally, if not more, important. Moreover, following the arguments discussed that some of the foundational beliefs used in argumentation are not always scientifically valid, it is not perhaps surprising that

the domain has been suggested to lack scientific integrity. In turn, if such a knowledge base has been adequately developed, it can, to an extent, be used for the automated construction of Bayesian networks rather than expert experience (Buntine, 1996). This section therefore focuses on the content and nature of this knowledge base, highlighting the extent to which the different sources of information can be conceptualised and evaluated.

### 2.5.1 *Knowledge sources and mechanisms*

In an ideal setting, the entire body of knowledge developed over the years would be incorporated in (some sort of) knowledge system. This structured framework can be updated when new information becomes available, which, in turn, could influence the current belief in arguments and theories. However, the knowledge can be derived and currently stored in a variety of ways, which is shown in Figure 2.3. Some of this knowledge can be referred to as explicit, consisting of a body of statements or arguments. Whilst some of the parameters might be known and deliberately chosen (such as a trace type and recipient object in transfer studies), the existence of other variables others may not be necessarily known (such as the effect of environmental conditions or human reasoning processes). Therefore, these statements have a varying degree of certainty (Feynman, 1955), and empirical studies require the acknowledgement of limitations of the research setting, allowing one to move away from presenting evidence as if the faith-based assumptions were true (Saks, 2010). Such explicit information can, on the one hand, include structured knowledge stored in databases. These include anything from national DNA and fingerprint databases until commercially available datasets containing, for example, shoe soles and tire tracks. Moreover, a shared use has been improved following the 2005 Prüm convention easing the exchange of DNA, fingerprint, and vehicle registration information within the European Union (Council of the European Union, 2005) as well as the development of transnational databases through the European Network of Forensic Science Institutes. However, on the other hand, the entire set of empirical findings published in the literature is in most cases unstructured (although systematic meta-analyses might be of a structured nature).

Moreover, the accumulated forensic knowledge base also contains knowledge which is not explicitly stated in arguments or probability functions, which can be referred to as tacit knowledge (Polanyi, 1966). In other words, “we know more than we can tell” (Polanyi, 1966, p.4). It is often the case that one relies upon common sense or intuition based on experience and content-dependent practical knowledge to form associations (Davies, 2015). Within the different forensic fields, many judgements rely upon a combination of experience-based knowledge which has been developed heuristically, and more explicit, scientific, knowledge gained through empirical research (National Research Council, 2009). Although judgements based upon experience can be very valu-

able in assessing hypotheses, a transparent presentation of the assumptions made and reasoning processes deployed is necessary to evaluate and/or contest these judgements (Thompson, 2012), rather than that these are simply assumed to be true (Mnookin et al., 2011). Therefore, the specific interest shift towards making the results of empirical studies more structured, as well as making expert experience both explicit and structured (see Figure 2.3). Section 2.5.2 and 2.5.3 will discuss these respectively.

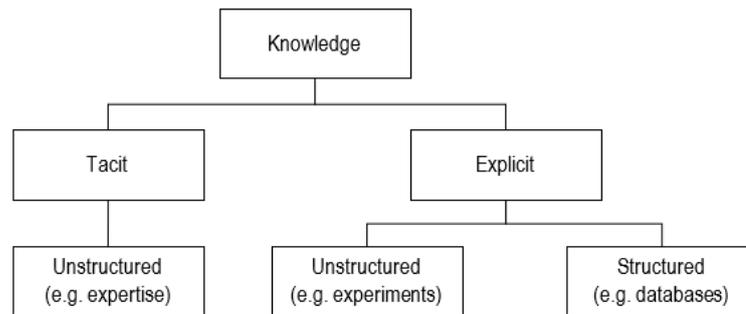


Figure 2.3: Simplified overview of knowledge types within the forensic sciences.

### 2.5.2 Structuring information from empirical studies

It must be accepted that the complex real-world cannot be completely conceptualised and modelled. However, by performing empirical studies and thereby objectifying the process of making observations about phenomena, the ideas of how these complexities work can be continuously updated, ideally obtaining a knowledge base approximating ‘truth’. Generally, experimental studies lead to the determination of the frequency  $P(E|H)$  of observing some evidence under fixed parameters of the model. Thus, while the hypothesis being tested for each data point is fixed, inferences can be drawn about the randomness of observing the evidence. While the parameters of the research setting are generally carefully chosen, it is possible that some are simply not known to exist or assumed to be accounted for. The disregard for secondary transfer and the possible influence of cognitive biases are examples of these (possibly unrecognised) uncertainties (see Section 2.3.4). While aleatory uncertainties are due to variability in the data (such as the throwing of a dice), epistemic uncertainties are the result of a lack of knowledge and can be reduced by gathering more data and by refining models (Kiureghian and Ditlevsen, 2009).

When aiming to conceptualise experimental studies and combine their findings into arguments within a single knowledge base, it is thus crucial that the parameters of the study design are known and given as well as that uncertainties are understood. First of all, the reliability of the method, which represents the extent to which the measured variable values are free of random error (Jacobe, Leitenberger, and Bergstresser, 2007)

should be understood. This does not just require the evaluation of the precision (variance of error, dependent upon both the repeatability and reproducibility) of the method, but also the variance of the true scores themselves (Streiner and Norman, 2006). Such random errors are different from systematic errors, which can occur through biases in the design or execution of the study (Jacobe, Leitenberger, and Bergstresser, 2007) and influences the validity of the method. This validity represents the extent to which the variable values obtained through the method correspond to the ‘truth’ (are accurate). Such truths can, for example, depict the internal validity (whether there is an actual causal relationship between the variable value and what one aims to establish) and external validity (the generalisability of the results) (Jiménez-Buedo and Miller, 2010).

The interpretation of fibres as evidence is taken as an example to understand the extent to which the findings from empirical studies could be combined into structured frameworks. First, to highlight the use of fibre evidence in court cases, a short study was undertaken for the purpose of this section, to understand the ways in which the evidence is presented in court, and to what extent the uncertainties surrounding fibre evidence are incorporated in the expression of beliefs. Figure 2.4 shows the result of collecting and analysing the 145 cases that appeared in the Court of Appeal of England and Wales, and contained the use of textile fibres (using the search term of ‘fibres’ and excluding irrelevant instances such as those containing asbestos). Despite these being a legal interpretation of the scientific observations, these results highlight the wide variation in terminology used. More specifically, the term “match” is used most frequently, such as in “a single fibre that matched the carpet in the stolen Ford Escort” ([2014] EWCA Crim 2047). What is not being highlighted in this case is whether this ‘match’ is based upon a commonality of feature values, or whether this also considers the population frequency, concluding that the stolen Ford Escort is the only possible source. The use of such a term implies that the scientist is not necessarily sure about population frequencies, but it can cause significant misinterpretations between scientists and lay people (Thompson, 2018). This is also the case for statements where “fibres from x were found on y”, such as the case where “fibres from [the suspect’s] cardigan had been found on the appellant’s jumper” ([2007] EWCA Crim 1041), suggesting an even stronger decision on the source of the fibre. It is important to understand this when discussing the knowledge base on fibres, as the results highlights the ways in which, at least in a court setting, uncertainties are incorporated in reasoning. Moreover, many of these statements seem to address activity level hypotheses, although little is being argued on the factors involving the evidence dynamics of fibres in these cases.

Following these observations, the questions is raised to what extent the knowledge base can contribute to these interpretations, not only by making them better supported by scientific findings, but also to ensure that the knowledge base is optimally used. In order to explain this point, ten studies have been randomly chosen which have stud-

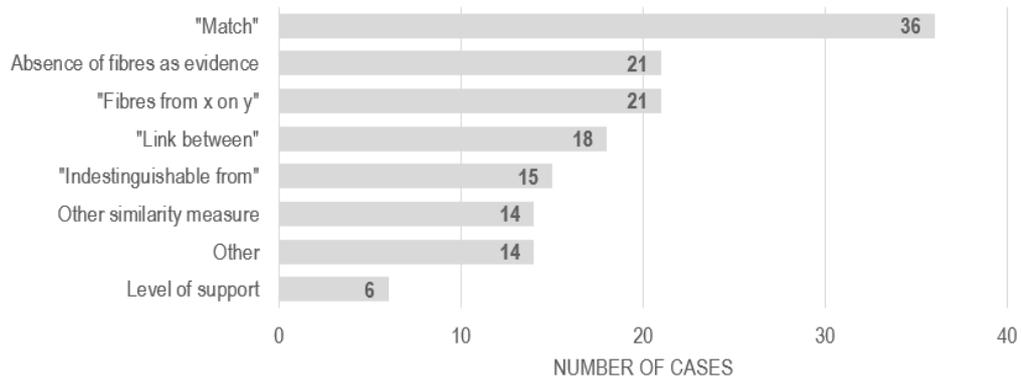


Figure 2.4: Results from the analysis of 145 transcripts from the Criminal Court of Appeal of England and Wales, describing the ways in which results from analyses on fibres as evidence were expressed

ied the population frequencies or the transfer and persistence of fibres. The aim of this analysis is to highlight how studies can be combined based on their aims and how well their parameter choices complement each other. Therefore, the number and types of parameters they varied within their study and the measurable outcomes which followed from the research design have been identified. These results are shown in Figure 2.5. Analysing these results backwards from the final observations, it can be observed that when this group of studies would expand, a significant understanding for each of these factors with regards to changes in fibre types, recipient objects, and environments, can be obtained. Whilst this is not surprising, it clearly highlights the possibility and advantage of combining these studies within a robust evidence base, in order to continuously develop generalised theories encapsulated in a knowledge base. However, the extent to which this can be achieved depends on the specific information that is provided on the factors surrounding, for example, the transfer mechanism within a study. For example, rather than varying parameters in isolation, it follows from Figure 2.5 that many studies do not only assess more variations of a single parameter (thicker lines), but also study multiple parameters (multiple lines). Depending on the specific research design, it makes the process of understanding specific causes for changes in observations more difficult. For example, it is not unthinkable that there would be less uncertainty introduced in inferences when a single study examines the effect of three different recipient garments than if there would be three different studies (i.e. contexts, locations, researchers), each considering one of these garments. It is therefore not only important to consider where the gaps in the current evidence base are, but also to what extent the results of each study can be incorporated in the already existing evidence base.

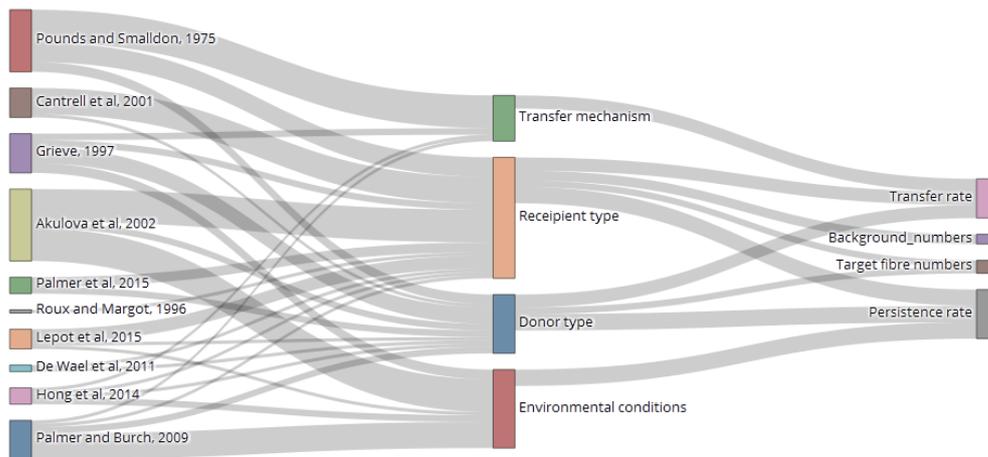


Figure 2.5: Sankey diagram highlighting the relation between the ten fibre test studies randomly selected for this example in terms of parameters they vary and characteristics they measure.

### 2.5.3 Making tacit expert reasoning explicit

In addition to the structuring of empirical findings, another major challenge within forensic science has been the conceptualisation of the wealth of knowledge currently ‘stored’ as expert experience. The dangers of having this knowledge, although extremely valuable, not presented in an explicit and structural form may raise concerns relating to the accuracy of the interpretation of findings in the case context and the transparency of presenting these results to the wider criminal justice system (Aitken 2009; Hamer 2012; PCAST 2016; The Law Commission 2011; Smit, Morgan, and Lagnado 2017). The consequence of this is as well that it is more challenging to discuss discrepancies between the findings of different scientists, as the components and assumptions within arguments are not explicit and can therefore not easily be compared. For example, the knowledge base used in forensic palynological interpretations does not only rely upon a combination of case-based and non-case based research and experience, but the practice of current interpretations have shown to include arguments supported extensively by the experience of an expert (Mildenhall, Wiltshire, and Bryant, 2006; Wiltshire and Black, 2006). Moreover, reasons for not often employing palynological evidence in, for example, the United States, include the reluctance of courts to use this as evidence and a lack of information and scepticism about its value (Bryant and Jones, 2006). It is in such cases crucial that the arguments and reasoning processes are made explicit to make the interpretation process, including assumptions and uncertainties, more transparent as well as to allow for a better integration of experience and the results of empirical studies. This is not only important to the validity of current arguments, but explicit knowledge is also much easier to incorporate in knowledge frameworks for future use.

## 2.6 THESIS' PERSPECTIVE

In the light of many concerns raised regarding the interpretation of forensic evidence, it has been argued that reasoning in the forensic science discipline needs to be of a normative character, to not only clarify good reasoning, but also to guard against logical inconsistencies (Biedermann, Taroni, and Aitken, 2014). As Finetti (1974) simply explains:

“To spend too much time on description is unwise when a normative approach exists, for it is like asking people’s opinion of  $2+2$ , obtaining an average of 4.31 and announcing this to be the sum. It would be better to teach them arithmetic” (Finetti, 1974, p.vii)

The question here is what the ‘arithmetic’ of the forensic science discipline should be. Because of the complexity of the ‘real world’ in which forensic reconstruction approaches must engage (Morgan, 2017a), the theories in the forensic sciences are not as straightforward as those of simple arithmetic, and defining norms, requirements or standards as part of a normative approach in reasoning under uncertainty is challenging. Nevertheless, progress can be made if there is recognition that theories are based on situational beliefs and can continue to evolve in the light of newly gained knowledge, constantly adding representations of small real world features (Jaynes, 2003). However, to highlight where such awareness is needed most within the forensic domain (and whether these issues are indeed present), the chapters of this thesis that follow aim to understand and evaluate the current interpretations in two ways: the interpretation and understanding of evidence from the perspective of the legal system (Chapter 3) and the interpretation of evidence from the perspective of the forensic scientific domain (Chapter 4 and 5). This thesis specifically aims to evaluate and understand whether the critique that the forensic sciences have faced (and will possibly continue to face) are justified or whether the domain, as a whole, could have performed better. Generally, despite a growing and elaborate understanding of many of the processes in play, concerns can be expressed with regards to the current practices of combining the current knowledge into valid beliefs and justified decisions.

## CHAPTER III

### MISLEADING EVIDENCE IN UNSAFE RULINGS IN ENGLAND AND WALES

#### 3.1 INTRODUCTION

Uncertainty is an inherent characteristic of many forensic analysis and interpretation processes. Therefore, advancing the role that the growing forensic knowledge base can play throughout the criminal justice system requires an understanding of how uncertainties are currently being dealt with and whether this is improving over time. This coincides with the significant number of concerns discussed in various government-led reports as discussed throughout Chapter 2, highlighting the need for research on the scientific validity of methods, judgements, and presentation methods in court (National Research Council, 2009; PCAST, 2016; The Law Commission, 2011). As will be discussed in the following sections in more detail, empirical research has increasingly focused on these topics, either by testing the validity of methods in experimental settings or by studying processes in mainstream casework, although a lack of studies highlighting the nature and significance of problematic evidence remains (Bono, 2011; Gross, 2008; Gross and O'Brien, 2008; Saks and Koehler, 2005). More specifically, the UK government has stated:

“It recognises the potential value of the proposed reliability test in reducing the risk of unsafe convictions arising from unreliable expert evidence. However, there is no robust estimate of the size of the problem to be tackled – either in terms of the number of cases where unreliable expert evidence is adduced, nor in the impact this has in terms of subsequently quashed convictions.” (UK Ministry of Justice, 2013, p.4)

This study begins to address this gap by systematically assessing the nature, extent, and consequences of ‘unreliable expert evidence’ in legal rulings in England and Wales, by studying the wider issue of misleading evidence within any ruling overturned by the Court of Appeal, Criminal Division. More generally, it presents a method which is applicable to other legislations and of which the results can be used to not only develop methods to avoid evidence being misleading in the future, but also to identify possible cases in which it has not surfaced yet.

### 3.1.1 *Unreliable expert evidence and uncertainties*

When forensic evidence is used throughout a criminal investigation, it is assumed to have some relation to the criminal act, and therefore has some ability to support the reconstruction of related events. More specifically, analysing an item of evidence aims to determine the value of parameters of this observed evidence (e.g. the refractive index of glass) which can subsequently be used in the interpretation stage to express a belief in hypotheses (e.g. possible sources). However, variation may exist between the true and observed parameter value, impacting subsequent interpretations (Christensen et al., 2014; Mnookin et al., 2011). As discussed in Chapter 2, such variation depends upon the method's accuracy or systematic error (determined by the specificity (true negative rate) and sensitivity (true positive rate), together with a threshold above which the method can be called 'reliable' (Cole, 2012)), as well as on information on its precision or random error (their repeatability and reproducibility) (PCAST, 2016; Streiner and Norman, 2006). In addition, factors may influence parameter values post-event, such as environmental conditions and collection strategies (Morgan et al., 2008; Scott et al., 2014). 'Unreliable expert evidence' then relates to the extent to which the meaning of the uncertainty caused by such factors are and can be considered in the interpretation and presentation stages of the forensic science process.

From a legal perspective, evidence admissibility standards should be the first safeguard against unreliable expert evidence in court (see Section 2.2.2). In England and Wales, judgements have suggested that expert evidence is admissible when adding expertise not available to the judge or jury ([1975] QB 834), when it is presented independently ([2006] EWCA Civ 1028) and by an expert in the field ([2006] EWCA Crim 2312), as well as that it must be be "sufficiently well-established to pass the ordinary tests of relevance and reliability" ([2002] EWCA Crim 1903). However, it has been argued that a statutory admissibility test and list of guidelines is needed in England and Wales to avoid allowing expert evidence without it being sufficiently scrutinised (The Law Commission, 2011). Moreover, Heaton (2013) has shown that neither the CCRC nor the Court of Appeal (Criminal Division) explicitly consider the possibility of individuals being 'innocent' throughout the post-conviction process. This study also led to the conclusion that, in order to ease the appeal process for those innocent, fresh evidence should be able to be received more readily by the Court (Heaton, 2013).

Moreover, a growth in empirical research allows for a greater understanding of uncertainties related to expert evidence, which, in turn, may make it easier to understand the relevance and reliability of evidence. For example, studies have highlighted factors involved in the dispersion, transfer, and persistence of many different trace evidence types including GSR and trace DNA (Maitre et al., 2017; Meakin and Jamieson, 2013). Additionally, a growing body of research has focussed on sources of potential bias in the

analysis processes of domains which rely upon methods with subjective aspects, such as handwriting analysis (Found and Ganas, 2013) and the analysis of skeletal remains (Nakhaeizadeh, Dror, and Morgan, 2014). Moreover, in addition to the growing analysis of error rates using past cases (Cole, 2005; Kloosterman, Sjerps, and Quak, 2014; Thompson, Tangen, and McCarthy, 2014), a recent shift can be observed, integrating blind testing programmes within mainstream case examinations (Kerkhoff et al., 2015). Current interpretation and presentation processes have been the focus of research after identification of misinterpretations of the frequency of features in populations, including those of fingerprints (Office of the Inspector General, 2006; Scottish Criminal Records Office, 2006), hair (FBI, 2015), and bullet-lead (National Research Council, 2004), and several high profile cases triggered research into the misleading aspects of statistical evidence (Buchanan, 2007; Meester et al., 2007). However, despite the growing knowledge base, uncertainties will remain and so does the need to acknowledge these and understand their effects.

### 3.1.2 *Misleading evidence and unsafe rulings*

Failing to acknowledge some of the discussed uncertainties that are associated with the evaluation of hypotheses (either due to a lack of knowledge or a misinterpretation) may result in an erroneous understanding of the evidential value of evidence, which, if sufficiently significant, results in “reporting support for a hypothesis that is not true” (Kerkhoff et al., 2015) while reporting opposition for a hypothesis that is true (to avoid misleading interpretations where evidence could also support an hypothesis that is true (e.g. a partial mark has common features with multiple prints (Thompson, 2008)). When doubts exist on the safety of the conviction, decisions made in the Crown Court can be appealed by the defence (Criminal Appeal Act 1995) or prosecution (Criminal Justice Act 2003). Although it not being a legal term, ‘misleading evidence’ will be used in this thesis to simply reflect any information presented which misled or had the potential to mislead decision-makers when expressing belief in, specifically in relation to forensic evidence, source and activity level hypotheses (see Section 2.3.2).

Doubts on the safety of a conviction (or acquittal) can vary in nature. Fresh evidence might come to light which the defence or prosecution deems relevant. Appeal court judges should subsequently quash the conviction if they argue rulings to be unsafe and may also order a substitute conviction or re-trial (Criminal Appeal Act 1995), following the identification of issues which in retrospect could have changed the decision of the jury ([2014] EWCA Crim 574, [2000] EWCA Crim 45). Despite the fact that the phrase ‘lurking doubt’ is sometimes used to discuss such cases, such beliefs are without clear evidence and reasoning not sufficient to allow an appeal Leigh, 2006. Additionally, as Blaxland (2017) discusses, the ‘jury impact test’ has not always been considered mandatory when the appeal court is faced with new evidence. Moreover, it has been suggested

that over time, the appeal court has become more liberal towards allowing appeals based on new evidence, although the rate of success in these appeals has decreased (Roberts, 2017). In addition to new evidence, procedural irregularities can also cast doubt on the safety of a conviction. Here, again, discussion has made its way into the jurisprudence highlighting circumstances under which this should (always) result in a successful appeal even if it strongly believed the defendant is guilty (Spencer, 2007). This includes gross misconduct that could relate to the forensic evidence throughout the investigation or in preparation for the trial. Dargue (2016) has shown that, specifically in the Court of Appeal (Criminal Division), the summing up, misuse of evidential discretion, and the presence of new evidence were the main grounds for appeal. This raises the question on whether these can be related to misinterpretations of the trial evidence.

The more specific studies on case judgements addressing the nature of misleading evidence focus on individual or small sets of high-profile cases (Gross and Shaffer, 2012), such as those in England and Wales (Roach and Trotter, 2005), the Netherlands (Brants, 2011), and Switzerland (Killias, 2008), often presenting specific recommendations following the outcome of each case. Some of the earliest more extensive empirical-based studies have reported issues related to eyewitnesses, informants, bad character, fingerprint forgeries, faked autopsies, and an inadequate defence (Borchardt and Lutz, 1932; Brandon and Davies, 1973; Connors et al., 1996; Frank and Frank, 1957; Fricker, 1993; Geller et al., 1999; Radin, 1964). Although it has been argued that knowledge is still lacking to draw strong inferences about the relationship between forensic evidence and wrongful rulings (Cole, 2012; UK Ministry of Justice, 2013), many of the more recent published studies on wrongful convictions include sections on the role of forensic evidence. An overview has been compiled for this study, see Table 3.1. Public attention rose in the USA by post-conviction DNA testing (see Section 2.3.2) through the Innocence Project (Scheck, Neufeld, and Dwyer, 2000), which has been referred to as the beginning of “the age of innocence” (Zalman, 2011, p.1499). This led to both the acceptance that wrongful convictions happen, as well as a growth in research identifying its causes and developing reforms (Leverick and Chalmers, 2014) through the use of more comprehensive data sets on population features (Cole, 2012). In many of the studies in Table 3.1, non-DNA evidence was initially used to narrow down the pool of suspects in what was later argued to be beyond what was scientifically valid (Connors et al., 1996). In the decade that followed, studies aimed at highlighting the general severity of problems with forensic evidence (Gross et al., 2005; Saks and Koehler, 2005), while others aimed more specifically in categorising these issues (Garrett, 2008, 2011), fuelling the debate on who to blame for these consequences; bad lawyering (Collins and Jarvis, 2009) or bad forensic science (Garrett, 2011). More recently, comparative studies have been performed, recognising ‘forensic error’ (Gould et al., 2013) and the amount of evidence types at the trial (Harmon, 2005) as predicting factors of unsafe convictions. What the results presented in Table 3.1 mostly show is that wrongful convictions are not always

just an issue of flawed science or bad lawyering, but rather, flawed communication and interpretation (Collins and Jarvis, 2009; Garrett, 2011; Griffin, 2015), an issue both sides should take responsibility for.

### 3.1.3 *Present study: a structural approach*

The idea that some of the evidence has the potential to be misleading throughout an investigation and in court is shown in the abstract representation in Figure 3.1. This study aims to contribute to understanding the nature of such misleading evidence by presenting and implementing a systematic content analysis. It aims to infer the nature of misleading evidence by identifying those instances where they were identified through concerns expressed by appeal judges on the safety of trial rulings (see Figure 3.1) in the Criminal Court of Appeal of England and Wales. The results will provide insights in both the type of evidence that is misleading (e.g. witness statements or DNA) as well as the nature of this misleading evidence (e.g. the evidence was not reliable or relevant). Of specific interest is whether the issues could have been known and subsequently could have been avoided in the trial. A greater understanding can drive subsequent research, allowing us to draw more general inferences about the complete set of misleading evidence in Figure 3.1, to include those who have not been recognised and to avoid similar issues in the future (Borchardt and Lutz, 1932; Gross and Shaffer, 2012; Rossmo, 2014).



Figure 3.1: Diagram showing the present study sample (overlap), a sub-sample of all misleading evidence, obtained through studying overturned unsafe rulings.

For the purpose of the study, we assume that the ‘correct’ rulings are those based upon the judgement of the appeal court judges, as the ground-truth cannot be known other than by running test-cases through the system. Additionally, allowing for the possibility that there are unsafe rulings which have not been recognised as such (Cole, 2012; Gross and Shaffer, 2012), the results can be used as a (highly liberal) snapshot of the current situation (assuming that rendering a ruling as unsafe is generally justified). Moreover, although it has been suggested that the frequency of (known or revised) errors of justice

Reference	Where	Data	N (F)	PC DNA	Categories of the issues with the forensic evidence
Borchardt & Lutz (1932)	USA (62)/ UK(3)	Significant cases	65 (8)	none	Wrongful conclusions (including handwriting, blood, and ballistics).
Bedau & Radelet (1987)	USA	Death penalty cases: various including NYT index and public inquests <sup>a</sup>	350 (46)	none	Erroneous diagnosis of the cause of death (16) and misleading circumstantial evidence (30)
Rattner (1988)	USA	Books, documents, newspapers	205 (3)	none	Forensic science errors (3)
Wilson (1989)	AUS	Newspaper and criminological literature <sup>b</sup>	20 (12)	none	Forensic expert as advocate (5), inconclusive expert evidence (5), circumstantial/suspect evidence (6)
Radelet et al. (1992)	USA	Death penalty cases <sup>a</sup>	416 (u)	none	Corrupt practices, rushing to judgments (unreliability in laboratory work).
Connors et al. (1996)	USA	Various searches	28 (22)	all	Too much weight on non-exclusion and scientific strength of non-DNA evidence (blood, semen, hair).
Sheck et al. (2000)	USA	Innocence project <sup>c</sup>	62 (u)	all	Serology inclusion (32), defective or fraudulent science (21), unreliability, including microscopic hair comparison (18), other forensic inclusions (5), DNA inclusions (1)
Saks & Koehler (2005)	USA	Innocence project <sup>c</sup>	86 (u)	all	Testing errors (63%), misleading expert testimony (27%)
Gross et al. (2005)	USA	Combination of media and website datasets <sup>d</sup>	340 (24)	144 (42%)	Perjury by a forensic scientist
Langdon & Wilson (2005)	AUS/NZ	Law databases, newspaper, books <sup>b</sup>	32 (21)	partly	Partisan expert testimony (22%), inconclusive expert evidence (31%), circumstantial/suspect evidence (44%)
Roach & Trotter (2005)	UK	IRA Cases	4 (4)	none	Over-estimation of probative value of test result for hypotheses (explosives).
Garrett (2008)	USA	First 200 DNA exonerations <sup>e</sup>	200 (113)	all	- Use of evidence with limited probative value (serology, hair) - Improper/misleading testimony by examiners (hair, fibres, bite marks)
Garrett & Neufeld (2009)	USA	DNA exonerations <sup>e</sup>	137 (82)	all	1. "invalid forensic testimony" (82) 1.1. The misuse of empirical population data 1.1.1 Non-probative evidence presented as probative (48) 1.1.2 Exculpatory evidence discounted (23) 1.1.3 Inaccurate frequency or statistic presented (13) 1.2. Conclusions on probative value unsupported by empirical data. 1.2.1 Statistics without empirical support (5) 1.2.2 Non-numerical statements without empirical support (19) 1.2.3 Conclusion that evidence originated from defendant (6) 2. Withholding exculpatory forensic evidence (3)

Reference	Where	Data	N (F)	PC DNA	Categories of the issues with the forensic evidence
Collins & Jarvis (2009)	USA	First 200 DNA exonerations <sup>e</sup>	200 (32)	all	Forensic science malpractice
Garrett (2011)	USA	First 250 DNA exonerations <sup>e</sup>	250 (155)	all	<ul style="list-style-type: none"> <li>- Unreliable forensic evidence (method that does not produce consistent or accurate results).</li> <li>- Invalid conclusions following the analyses. See [55].</li> </ul>
Judicial Commission of NSW (2011)	AUS	Conviction appeals in NSW	315 (u)	u	General categories included unreasonable or unsupported jury verdict, issues with the admission/rejection of evidence, and issues with the direction of the judge towards the jury.
Gross & Shaffer (2012)	USA	National registry of exonerations <sup>d</sup>	873 (210)	325 (37%)	False or misleading forensic evidence, ranging from simple mistakes to invalid techniques to fraud
Gould et al. (2013)	USA	Various searches and previous research	260 (88)	partly	Errors more in testimony or interpretation rather than scientific testing: <ul style="list-style-type: none"> <li>- Neglecting to provide the jury with key information</li> <li>- Overstating the inculpatory nature by providing inaccurate or non-existent statistics (e.g. hair)</li> <li>- Misstating the certainty of results when the technique does not allow for it</li> <li>- Poor communication between lab and police and prosecutor</li> <li>- Inadequate training on criminal justice officials</li> <li>- Police unaware of significance, evidence was not revealed</li> <li>- Tunnel vision: more a confirmatory role in the investigation</li> </ul>
Garrett (2015)	USA	First 330 DNA exonerations <sup>e</sup>	330 (234)	All	Concealed exculpatory evidence, erroneous analyses, vague testimony of similarity
Dioso-Villa (2015)	AUS	Ex gratia application cases	71 (22)	u	Forensic error or misleading forensic evidence
Griffin (2015)	UK	Forensic evidence in Court of Appeal (Criminal Division)	U (all)	all	<ul style="list-style-type: none"> <li>- DNA: new tests, reliability, instructions on experience of expert</li> <li>- Medical: uncertainties on cause of death, diagnostic standards in sexual abuse</li> <li>- Psychological: evidence on mental capacity at the time of the crime</li> <li>- Fingerprints: lack of expertise and scientific standards</li> <li>- Documents: new evidence showing document manipulation</li> </ul>

Table 3.1: An overview of studies into larger datasets and significant cases where issues with the forensic evidence were observed. It highlights similar datasets (superscripts), number of cases (N) and those containing forensic evidence (F), the percentage of post-conviction DNA tests, and unknowns (u).

is relatively low compared to all convictions (Gross and Shaffer, 2012), the disutility of even a single wrongful ruling is extremely high.

### 3.2 MATERIALS AND METHODS

Information from legal cases is often stored in unstructured text-based forms, impacting the applicability of analytical and predictive methods based on legal decisions (Stranieri and Zeleznikow, 2005). These processes have been improved through the development of semi-structured information from narrative judgements to address information-finding problems in case database analysis (Rose, 1994). Further studies have shown the applicability of a range of methods for the so-called ‘knowledge discovery from databases’ (KDD) (Fayyad, Piatetsky-Shapiro, and Smyth, 1996). These include classification and clustering results of, for example, similar case-types, outcomes, or length of the legal process (Fayyad, Piatetsky-Shapiro, and Smyth, 1996; Wilkins and Pillaipakkamnatt, 1997), as well as finding associations between case properties and legal outcomes (Dargue, 2016; Ivkovic, Yearwood, and Stranieri, 2001). In this study, text-analysis is applied. More specifically, in order to structurally study case documents of wrongful rulings to make valid inferences on the underlying themes, a content analysis approach was used (Drisko and Maschi, 2015; Krippendorff, 2007). This method has been applied to a variety of studies on legal judgments, mainly aiming at descriptive or analytical analyses (Evans et al., 2007; Hall and Wright, 2008). The steps that were undertaken include a systematic case selection, case coding, testing of coding reliability, and analysis of results (Hall and Wright, 2008).

#### 3.2.1 *Case selection*

The cases used in this study are a convenience sample of all relevant cases, accepting that this only includes misleading evidence identified through the used overturned rulings (see Figure 3.1)(Krippendorff, 2007). Although it has been argued that there is a lack of information on appeal outcomes (Jacobson, 2013), exacerbated after the discontinuation of Casetrack, appeal decisions were gathered from published appeal court judgments in the database Westlaw UK. They were selected from the ‘case analysis documents’ on the basis of having been heard by the Criminal Court of Appeal of England and Wales (EWCA Crim, further referenced as ‘AC’), and having been labelled with ‘criminal evidence’ in the database according to Sweet & Maxwell’s Legal Taxonomy (Scott and Smith, 2010). Westlaw UK was chosen as it contained much more judgments as compared to Bailii (over 10.000 versus approximately 1680 in the study period), and because the discontinuation of Casetrack would make this study otherwise less sustainable.

The dataset of relevant documents was further limited to appeals allowed with regards to the conviction or acquittal (rather than the sentencing), as this was believed to provide

more information on significant misleading evidence, following a belief by appeal judges that the trial ruling was unsafe (outlined in the Criminal Appeal Act 1995). The study term (of the appeal rulings) was a 7-year period January 2010 through December 2016.

### 3.2.2 *Coding categories and considerations*

This study aimed to identify the basis for the successful appeal (i.e. the reason why the trial conviction was unsafe), which is of a descriptive nature and reflects simply the statement of the appeal judge (In Vivo) rather than a normative evaluation of whether that reason was justified. Coding categories were determined using three different methods, combining deductive and inductive techniques. Firstly, codes were developed a priori based upon the results from previous studies presented in Table 3.1 and the admissibility and criteria of evidence from the Criminal Justice Act 2003 and the Police and Criminal Evidence Act 1984. Examples include issues with the validity of collection and analysis methods, and the determination of the relevance and probative value of evidence. Secondly, the coding framework was adjusted to categorise the issues based upon the possibility of developing methods to further understand and avoid these issues in the future. The most important aspect was differentiating between whether there was new evidence, whether the trial evidence had been re-evaluated by an expert, or whether the decision had been overturned without new evidence. Lastly, the framework was refined through the identification of inductively defined codes from analysing the case documents (Drisko and Maschi, 2015; Hall and Wright, 2008). Added categories include, for example, procedural errors, inconsistent decisions by the jury, and insufficient directions by the judge on the relevance of evidence. The final coding scheme is reflected in the decision scheme in Figure 3.2. The following terminology reflects some of the detailed coding categories:

- Issues with the relevance of the evidence have to do with the question whether it can be of value to address the hypothesis (e.g. and therefore should have been excluded/included).
- When considering the probative value of evidence, it is agreed that the evidence was relevant but it was misleading with respect to how much/little weight was given to the evidence.
- The validity of the evidence is used to categorise evidence which was argued by the appeal judge to not be (presented) accurately or precisely.

### 3.2.3 *Case attributes*

Together with the nature and type of the (misleading) evidence, attributes such as the appeal outcome, requests for re-trials, involvement of the Criminal Case Review Commission (CCRC), trial dates, and proposition levels were also considered.

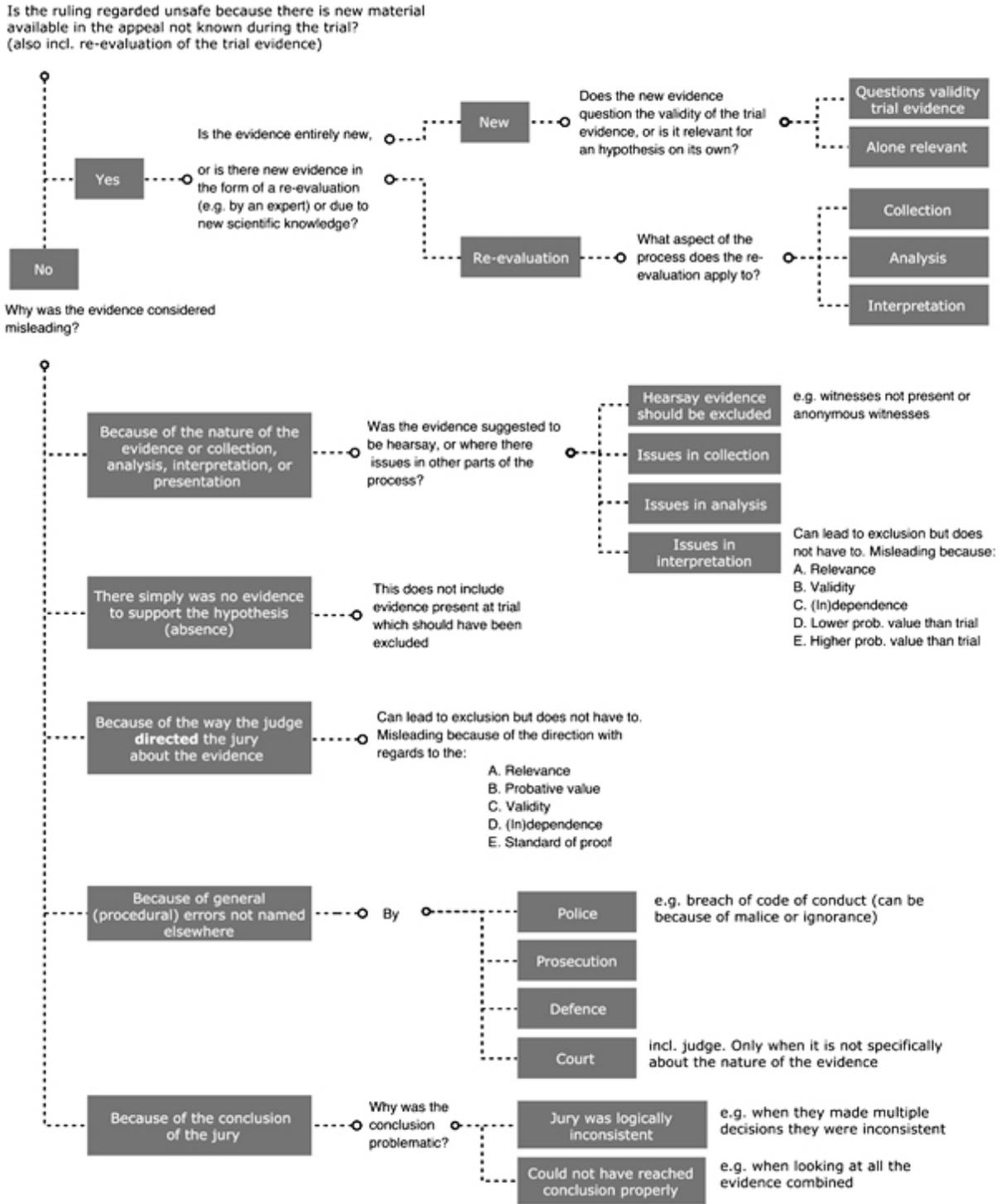


Figure 3.2: The coding scheme developed for this study to classify the reasons for rendering the trial ruling unsafe by the Court of Appeal judge(s).

3.2.4 *Inter-coder reliability*

To determine the reliability of the coding method, an inter-coder agreement index was developed based upon the coding of the nominal categories of the initial and two addi-

tional independent coders (as the initial coder was involved in the development of the coding rules which could impact the reliability (Lorr and McNair, 1966)), by determining Krippendorff's alpha coefficient (Krippendorff, 2007). The additional coders were given 20% of the total of cases studied and only relied upon the coding instructions presented in Figure 3.2 and the published appeal court judgments. They were required to identify the type of evidence questioned in the trial (and new evidence in the appeal if present), and the nature of why it was misleading (the chosen category following the coding procedure).

### 3.3 RESULTS AND DISCUSSION

#### 3.3.1 *Significance: number of cases*

In the 7-year period between 2010 and 2016, 10,859 cases were found in Westlaw UK as having been heard in the Court of Appeal of England and Wales. Because of the lack of structured data on whether these were appeals against conviction or sentence and whether these were successful or not, the first filter included the built-in label of 'criminal evidence'. Approximately 9% (996) of the cases remained to conduct further analyses on.

The challenge with the method developed for this study as compared to most of those discussed in Table 3.1 (which study already known unsafe rulings), is that only a small portion of all studied cases are of relevance. This is because the 10,859 cases are appeals against both sentencing and conviction, which also holds for the 996 cases labelled as 'criminal evidence'. Without studying these cases further it cannot be said what portion of these cases are successful appeals against conviction. More detailed ratios would require the need for more automated approaches of analysing the large number of case material – something that can only be explored once a more detailed picture exists of misleading evidence in published judgments as produced for this initial study. This also means that has not been further assessed what the details are of the excluded cases not labelled with 'criminal evidence'. Moreover, these initial results could be further compared with results of a similar study on other datasets, despite the difficulties with for example Casetrack (discontinued), Bailii (less judgements), or the comparability with international databases such as those of the U.S. Supreme Court. Both these approaches will provide an insight in the ability to generalise the findings from this study to the wider population by determining the representativeness of the ratio of appeals against sentencing/conviction, the rate of success on appeal, and the counts and evidence types which are included. The results must therefore be interpreted with caution when aiming to generalise findings.

Table 3.2: Details of the study set of published judgments in Westlaw UK between 2010 – 2016. Cases were first filtered on ‘criminal evidence’, followed by successful on appeal against conviction. The table includes the number of cases and individual counts within these cases, as well as the number of rulings, number of ‘problematic’ evidence types and the number of specific issues with these items of evidence.

	Count
Cases in Westlaw UK heard by AC between 2010 - 2016	10 859
Cases labelled with ‘criminal evidence’	996
Cases successful after appeal against conviction	218
Number of criminal counts	223
Number of rulings (suspects)	222
Conviction quashed	201
Conviction substituted	10
Acquittal quashed	10
Admission evidence	1
Problematic items of evidence	230
Identified issues with the evidence	235

Rulings in 218 of these 996 cases (22%) were argued unsafe because they contained misleading evidence, representing the overlapping area in Figure 3.1. Details of these cases are shown in Table 3.2. As argued previously by others (Cole, 2012; Collins and Jarvis, 2009; Gould et al., 2013), this study also suggests that the actual occurrence of misleading evidence is higher than only those identified through the current methodology. Namely, it can simply be the case that they have not gone into appeal or, as cases in the present dataset shown, it is possible that although errors are identified, they were not argued to “undermine the safety of the conviction” (e.g. see [2015] AC 1619 or [2016] AC 4, or where evidence was argued in the light of new evidence, to be ‘neutral’, see [2010] AC 2936). Therefore, these cases are not included in the present study, although understanding these issues is of great importance as such errors could potentially contribute to unsafe rulings in other contexts.

### 3.3.2 Further significance: time between rulings

The significance of the 218 successful appeal cases becomes more pertinent when considering the time between the original conviction and the decision by the appeal judges. This time difference in cases where this could be obtained ( $n = 208$ ) averaged 2.9 years but with a wide range ( $sd = 5.7$  years). Eighteen cases took over 10 years to be overturned, with a maximum time difference of 36.3 years (in [2014] AC 2047, which was quashed due to fresh evidence indicating police malpractice). Moreover, the method of data collection allowed for more detailed analyses. For example, comparing the median of the time difference (interquartile range) of 2.7(1.3 – 9.7) years for rulings where new

evidence had to be obtained with the 0.8(0.5 – 1.3) years for rulings which were overturned without having to acquire new evidence, shows a significant difference between the two distributions (Mann–Whitney  $U = 5684$ ,  $n_{new} = 47$ ,  $n_{notnew} = 179$ ,  $p < 0.001$ , two-tailed). This is consistent with the length of time it would take to acquire and consider new evidence in a case.

It followed from the case documents that of the 201 quashed convictions, no re-trial was ordered in 80 cases (in addition to 24 re-trials and 97 unknowns). Additionally, re-trials were requested following all 10 acquittals (including resumed proceedings), a possibility after the double jeopardy ban in the UK in 2003 (Criminal Justice Act 2003), see for example [2012] AC 414. In 10 cases, no re-trial was ordered as the conviction was substituted (e.g. murder to manslaughter). More generally, the time between the rulings does not always correspond with the time spent incarcerated, as, for example, the suspect may have been incarcerated before the trial date or have already completed their sentence before the appeal. The cases involved significant crimes, including those related to sexual assault ( $n = 50$ ), assault (34), murder (28), robbery (18), fraud (13), and drug offences (13). These numbers roughly compare to the distribution of counts in all crown court cases between 2014 - 2016, where the most common counts included violence against the person (including both assault and murder according to the Counting Rules for Recorded Crime (Home Office, 2019)) (22%), drug offences (14%), miscellaneous crimes against society (14%), theft offences (13%), sexual offences (13%), robbery (6%), and fraud (5%) (Criminal Justice System Statistics, Ministry of Justice, 2019). Especially sexual assault cases were more commonly found in this study as compared to the crown courts, facing mostly issues to do with witness statements, DNA and medical evidence, and evidence of bad character. Considering the problematic evidence type most commonly found in this study being witness evidence (Figure 3.3), it is perhaps not surprising that sexual assault often highly depends on this type of evidence. Nevertheless, whether a suspect was ultimately guilty or not does not change the fact that misleading evidence was still present in these trials in the first place.

### 3.3.3 Coding method reliability

The coding method reliability has been studied by determining the inter-coder agreement on 20% of the total number of cases coded for this study, resulting in a Krippendorff's alpha coefficient of 0.772 (47 subjects, 3 coders). This has been suggested as 'substantial' agreement (Landis and Koch, 1977), although others argue that above 0.8 would be considered reliable (Krippendorff, 2007). The coefficient presented here is, however, a conservative agreement as all categories are considered equally different, while issues with, for example, the probative value or the relevance are somewhat similar and not always easy to distinguish from the appeal court judgments. This does indicate that per-

haps a more detailed explanation of the coding categories and/or the reasoning of the judge could be valuable.

#### 3.3.4 *Overview of the results in each coding category*

Overtuns within 206 of the 218 cases were based upon a single key reason related to a single evidence type, highlighting the importance of avoiding any type of misleading evidence in trial cases. The major and minor categories in which the reason for the unsafe ruling was classified using the coding method (see Figure 3.2) can be found in Table 3.3. It follows that in a minority of instances (24%) there was entirely new evidence or newly gained knowledge (such as a re-evaluation by a forensic expert) which led to the unsafe ruling. In the majority (76%) of cases, however, the ruling was overturned given the same set of information available in the trial. These results do strongly suggest that there might be many more unsafe rulings beyond the post-conviction DNA testing cases summarised in Table 3.1. Additionally, from the 22 cases which were successful on appeal following a reference by the CCRC (according to the judgments), the majority (15 cases) were rendered unsafe due to the introduction of new evidence. This, again, suggests that perhaps not enough focus is directed on cases where there is not necessarily new evidence and adds on to the idea that the statutory test used by the CCRC should be reviewed (Kerrigan, 2006).

The reasons for why the evidence is misleading are studied from the point of the trial evidence. On the 19 occasions where there was new evidence which was relevant on its own (mostly witnesses ( $n = 4$ ), documents (3), medical evidence (3), and DNA (2)) and the 9 cases where there was an absence of evidence, no trial evidence was questioned. Additionally, there was no specific evidence associated with the unsafe ruling in some instances where there were issues with the direction given by the judge ( $n=6$ ), general errors (4), and issues with the decision of the jury (4). The remaining 193 reasons related to trial evidence are studied further. Here, the remaining cases where there was new relevant information questioning the trial evidence (c in Table 3.3) are considered together with the nature of the evidence as presented in court (a) as they both relate to the nature of the trial evidence, and only differ in the way the issue came to light ( $n = 122$ ). Moreover, the issues with the direction of the judge (b) indicate a misunderstanding of this similar nature of the evidence ( $n = 54$ ), an issue also identified by others (Judicial Commission of N.S.W., 2011). While such an issue with the direction of the judge might not in itself be 'misleading evidence', a misdirection on the value or relevance of the evidence or how this should be interpreted may also contribute to the expression of a 'wrongful' or less than ideal belief in an hypothesis by the jury. The presence of these revised categories are studied with respect to the evidence type groups shown in Figure 3.3.

Table 3.3: Overview of the results of each coding category from Figure 3.2 representing the reason for the unsafe ruling.

Coding category representing the reason for unsafe ruling	n (%)
a. Misleading nature of the evidence as presented in court	85 (36%)
Relevance	31
Probative value	18
Validity	17
Hearsay	16
Independence	2
Collection	1
b. Issues with the direction of the judge on	60 (26%)
Probative value	18
Standard of proof	17
Relevance	14
Validity	10
Independence	1
c. There is new relevant information	56 (24%)
New questions validity trial	26
New is relevant stand-alone	19
Expert re-evaluation of probative value	7
New questions probative value trial	3
Expert re-evaluation of collection method	1
d. General errors by prosecution (6), police (2), or court (10)	18 (8%)
e. Absence of evidence in the trial	9 (4%)
f. Issues with the decision of the jury	7 (3%)
	235

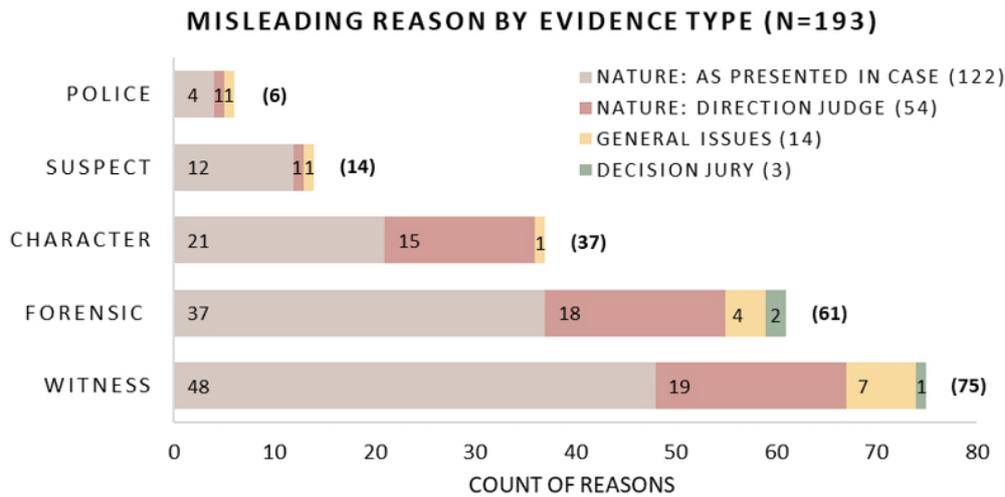


Figure 3.3: An overview of the evidence type groups and the reasons given why the evidence was misleading. 42 reasons did not specifically apply to an evidence type and are not included in this graph.

It follows from Figure 3.3 that witness (including victim) statements are the most problematic evidence type (39%), which has also been suggested as a major cause for concern by other scholars (Bedau and Radelet, 1987; Garrett, 2015; Gross and Shaffer, 2012). The second most-common evidence type is the grouped category of forensic evidence (32%), which will be a major focus of the subsequent sections. Although discussions have been ongoing as to whether (good and bad) character evidence (19%) has a higher risk of convicting the innocent and/or whether the underlying assumptions are fair towards a suspect (Spencer, 2016), especially as after becoming more easily admissible after the Criminal Justice Act (2003), research did not find evidence of unfair prejudice (Redmayne, 2002). The final two evidence types are police- and suspect statements. While false confessions have been identified as a significant factor in wrongful convictions in the USA (Gross and Shaffer, 2012) and can generally have a major impact in an investigation (Kassin, 2012), this study has only found four such cases, and are therefore either less common or were not identified in the present study. The four general reasons in Figure 3.3 are grouped in two sets and discussed in Sections 3.3.5 and 3.3.6.

### 3.3.5 *Issues with the nature of the evidence*

The two major categories related to the nature of the evidence are responsible for the majority of the overturned reasons found in this study ( $n = 176$  or 75%), see Figure 3.3. In other words, the common theme is that of possible erroneous beliefs in hypotheses in the light of such misleading evidence. In order to improve this process, it is necessary to understand what type of hypotheses are addressed with this evidence. This allows for the possibility to develop further knowledge bases with respect to each evidence type, such as within- and between source characteristics and evidence dynamics. The findings with respect to the hypotheses in question in each case is shown in Table 3.4.

It follows from the findings presented in Table 3.4 that the majority (116 or 66%) of items of evidence were used to address hypotheses related to an activity level, such as witnesses describing violent acts of suspects, victims describing sexual offences by suspects, or cases where evidence of bad character was used as evidence that a suspect was more likely to commit certain acts. The results highlight the importance of understanding trace evidence dynamics and the value and limitations of observing evidence such as DNA, fingerprints, and GSR when addressing these activity level hypotheses. Moreover, cases where there was a dispute about the lower, source level, hypotheses, were also commonly encountered whereby the evidence of witnesses and forensic evidence were misleading. In the remaining cases, the evidence was used to assess the validity of other evidence (this is not the same as issues with the validity in Table 3.3) and used to determine whether or not intent was involved.

Table 3.4: Overview of the of the number of cases of misleading evidence and the type of hypothesis that they addressed.

	Hypothesis related to:				
	Activity	Source	Validity	Intent	
<i>Witnesses evidence</i>	48	29	6	2	85
Police	2	3			5
Suspect	9	1	2	1	13
Other witness	37	25	4	1	67
<i>Forensic evidence</i>	39	10	4	1	55
Animal blood	1				1
Digital	3	1			4
DNA	4	4			8
Documents	11		1		12
Drugs	5		1	1	7
Fingerprints	4	2			6
Footwear marks		1			1
Glass		1			1
GSR	1				1
Handwriting		2			2
Medical	7		1		8
Weapon	3		1		4
<i>Character evidence</i>	29	2	5		36
	116	42	15	3	176

The final step in moving forward and addressing these issues by developing and better employing the forensic knowledge base is the need to understand the misleading aspect of the relationship between the evidence and the hypothesis. For example, was the evidence simply not relevant for the hypothesis? Was it given too much evidential weight? Or was the evidence itself not valid and therefore problematic for assessing the truth in the hypothesis? These results (using the sub-categories in Table 3.3) are shown in Table 3.5, grouped per general evidence category and hypothesis level (from Table 3.4). It follows from Table 3.3 that 16 cases were categorised as hearsay, following, for example, the argument that because a witness could not be cross-examined, the possibility of assessing their reliability was severely diminished ([2012] AC 1509). These cases were therefore categorised as questions on the (general) validity of the evidence. The three cases categorised under issues with the independence of evidence were rendered unsafe because of underlying issues of validity (whereby an identification was influenced by possible biasing information) and probative value (where two dependent items of evidence were presented as providing independent support for an hypothesis).

#### 3.3.5.1 *Relevance of the evidence*

First of all, the evidence must be relevant for hypotheses (Police and Criminal Evidence Act 1984). With regards to the character evidence, this assessment was the major issue, mainly whereby bad character was 'unfairly' used as evidence that a suspect could have

Table 3.5: Overview of the number of cases of misleading evidence and the detailed relationship between the different levels of hypotheses (see Table 3.4) and the nature of the misleading evidence, grouped by general evidence category.

	Misleading nature:				
	Relevance	Probative value	Validity	Standard of proof	
Witness evidence	8	14	60	3	85
Activity	4	7	35	2	
Source		6	22	1	
Validity	3	1	2		
Intent	1		1		
Forensic evidence	13	29	7	6	55
Activity	10	19	5	5	
Source	1	8	1	1	
Validity	2	1	1		
Intent		1			
Character evidence	24	6	4	2	36
Activity	20	4	3	2	
Source	1	1			
Validity	3	1	1		
	45	49	71	11	176

committed a criminal act. Evidence deemed irrelevant for the hypothesis in question was also observed in relation to several forensic evidence types, including cases where the presence of drugs was arguably wrongly used as evidence to support hypothesis in relation to handling or dealing (other types of) drugs (e.g. [2013] AC 901 and [2012] AC 2879).

In order to avoid such misinterpretations, it is crucial to define sub-hypotheses which are perhaps more clearly associated with the recovered evidence to provide transparency in the reasoning process. For example, when the presence of marijuana is used as evidence for dealing heroin, a sub-hypothesis can be defined on the dealing of marijuana, and how this is linked with the presence of marijuana on the one hand, and the dealing of heroin on the other. These relationships can then be assessed separately but interpreted as a whole.

### 3.3.5.2 *Probative value of the evidence*

After evidence is considered relevant for an hypothesis, the belief expressed in an hypothesis (with respect to another) is sometimes presented disproportionately to the actual probative value of an item of evidence. This issue was most often related to the forensic evidence (59%), and was also the major issue within the forensic evidence (53%). Examples on the activity level include medical evidence such as trauma, argued to be conclusive of penetration ([2012] AC 1433) or non-accidental trauma ([2010] AC 2847), which were on appeal revised considering new scientific findings. Within trace

evidence, the appeal highlighted questions such as those regarding the time when DNA was deposited ([2015] AC1732), to what extent the location of DNA ([2011] AC 460) and fingerprints ([2010] AC 2421) were suggestive of the criminal act, the possibilities of multiple transfer of blood ([2010] AC 2499), and the exact meaning of small quantities of GSR on clothing ([2014] AC 2507). Similarly, the appeal court rendered convictions unsafe where disproportionate weight was given to evidence when addressing questions of identifications or individualisations on a source level. Appeal rulings suggested the need to consider possible sources of glass unconnected to the crime ([2015] AC 1950), the misleading aspect of the trial judge failing to direct the jury about the meaning of a partial DNA profile ([2010] AC 1334), the importance of recognising disputes between fingerprint experts ([2011] AC 1296), and severe limitations in the notorious case of *R. v T* when determining the source of a shoe print ([2010] AC 2439).

In addition to identifying sub-hypotheses clarifying the relationships between evidence and hypotheses, determining the change in belief in an hypothesis following the evidence requires an in-depth analysis of the factors that can influence the state of the evidence given the specific hypothesis. As argued previously, it requires an extensive knowledge base on the factors influencing the dynamics of evidence and the abundance of observed characteristics. Although findings could possibly be generalised, it should be made clear under what assumptions these are applied in a case context and how this impacts the interpretation.

### 3.3.5.3 *Validity of the evidence*

When considering the witness statements, issues were most often associated with regards to their validity (e.g. the credibility of a witness). Although the validity of evidence is not directly linked to the relationship between evidence and an hypothesis (as compared to the relevance and probative value), a decrease in the validity of the evidence (whether it is the accuracy or precision) can, again, cause a misinterpretation of the probative value or relevance of the evidence (although these latter assessments could be justified themselves).

In relation to the witness statements, and to a lesser extent the character and forensic evidence with a questionable validity (including fingerprints, handwriting evidence, and digital evidence), it is pertinent that the factors influencing the validity are understood. These include, for example, the conditions under which the witness observations have been made (e.g. distance, weather) and the degree of subjectivism of forensic analysis methods and factors which can influence this process (as discussed in the introduction). Judges and jurors should be made aware of these effects and their implications for the belief in hypotheses, as they generally tend to consider forensic evidence as quite reliable (Garrett and Mitchell, 2016).

#### 3.3.5.4 *Standard of proof*

One of the roles of the judge is to direct the jury on the standard of proof (Judicial College, 2017). Part of this belief is an understanding of the factors discussed previously in this section. However, some of the issues do not fit in these categories and are related to the standard of proof more generally. Examples not related to a specific evidence type (see Section 3.3.4) include a lack of direction by the judge on what hypotheses they had to be sure of before they could convict (e.g. [2016] AC 1632). Moreover, the 11 cases where there was specific evidence (Table 3.5), the direction on the standard of proof related both to source and activity level hypotheses. It was, for example, argued that the judge failed to direct the jury that they could reach their own conclusions on the meaning of low-template DNA evidence ([2010] AC 549), and that the judge did not direct the jury on the law related to circumstantial drug-related evidence ([2015] AC 1733). Overcoming these issues from the point of the interpretation of the forensic evidence will mostly relate to a transparent presentation of the hypotheses the jury needs to assess, the uncertainties of the presented analyses or interpretations, and, together with that, an indication that the jury a) should reach their own conclusions based on these uncertainties and b) the extent to which the jury should be 'sure' of its conclusion.

#### 3.3.6 *General issues and the decisions by the jury*

In addition to the nature of the evidence which is (or has the potential to be) misleading in the truth-finding process, it follows from Figure 3.3 that in 14 cases there were 'general issues' with the evidence, including procedural errors by the court, failures to disclose statements by the prosecution to the defence, and breaches of code by police officers. These only form a small proportion of all cases and would require a more in-depth analysis of research targeted at these issues. Additionally, three cases were overturned where the appeal judges argued that the decisions of the trial jury were either logically inconsistent or where there was no possibility for them to reach a conclusion properly (e.g. [2010] AC 130). Research studying US juries has shown a variation in jurors' interpretation of the evidence following, for example, factors that might influence the validity of eyewitnesses (Schmechel et al., 2006), the so-called CSI-effect (Schweitzer and Saks, 2007), conditions affecting the interpretation of probabilistic evidence (Kaye and Koehler, 1991; Thompson, 1989), and a change in the method of presentation (e.g. qualitative versus quantitative) of identification evidence (McQuiston-Surrett and Saks, 2009). Because of the differences US and UK jury system, care must be taken when drawing generalised conclusions (Thomas; 2007). Nevertheless, understanding the idea that juries can make inconsistent decisions which are the grounds for a successful appeal raises further research questions on juror decision-making in combination with forensic evidence in the legal system in England and Wales.

### 3.4 CONCLUSION

Content analysis of published appeal court judgments can provide vast information on the presence and nature of misleading evidence. Although the sampling method causes a danger of over-analysing the data presented in this study, it does suggest that the relevance, probative value, and validity of evidence are often misunderstood and miscommunicated within a criminal trial when expressing beliefs in (competing) hypotheses. The consequences of these are severe and have caused many defendants to be wrongfully incarcerated. The results show that these issues can be addressed by clarifying sub-hypotheses on source and activity levels and by developing the knowledge base on, for example, the rarity of observed characteristics in a population and the factors affecting evidence dynamics. The fact that most of these rulings are overturned without including new evidence in the appeal suggests that they could have been avoided in the trial if these interpretations were improved. Accepting that uncertainty will remain present, it is important that the uncertainties are not caused by erroneous arguments and judgements but by inherent properties of the knowledge base, which, in turn, can be presented transparently.

The findings from this research support previous findings that the occurrence and nature of misleading evidence can no longer be attributed to simple individual 'bad apples' in the system (Thompson, 2010). Despite the small ratio of issues with misleading evidence (218 cases) compared to all cases (10,859), this study has identified structural issues which often could have been addressed at an earlier stage. This study provides a foundation of detailed issues related to the interpretation of forensic evidence and highlights the possibility of identifying these issues from appeal court judgments, fuelling the important debate on improving access to case documents post-conviction (Henneberg, 2017). On the one hand, the findings from this research can be validated by studying and improving the understanding of uncertainties involved in assessing the relevance, probative value, and validity of evidence on a more comprehensive level for individual evidence types. This also includes the use of methods to guide decision-making under uncertainty and with new information, such as Bayesian networks (Garbolino and Taroni, 2002; Pearl, 1988; Smit et al., 2016). On the other hand, there is great potential in applying similar methods to larger scale studies within England and Wales as well as in other jurisdictions, using, for example, more automated text-mining techniques. These can also be used to determine potential predicting factors of misleading evidence (see for general examples (Gould et al., 2013)), using specific factors related to the type of evidence and the type of hypotheses they address. The findings in this paper suggests that, as others have argued, "the scientific community can take the lead in reform efforts" (Garrett and Neufeld, 2009). It is therefore hoped that this research becomes more than just an 'academic exercise', and that these results allow for the much-needed continuous

evaluation of the use of forensic evidence in court and their contribution to both justified and unjustified rulings.

## CHAPTER IV

### TOWARDS A NORMATIVE APPROACH OF EVALUATING TRACE INTERPRETATIONS

#### 4.1 INTRODUCTION

It has already been suggested that forensic evidence can play a significant role in investigative and legal judgements by establishing and revising beliefs in hypotheses relevant for reconstructing past events (Houck, 2003; Robertson and Roux, 2010; Stoney and Stoney, 2015). However, the extent to which evidence can be meaningful depends on the ability to make context-specific interpretations, or in other words, “the drawing of rational and balanced inferences from observations, test results and measurement” (Cook et al., 1998b, p.152). Over time, the knowledge base used to support these interpretations is growing, meaning that analyses become more reproducible and reliable, and more insights are gained about the abundance of features and their changes under different conditions. Within the last decades, however, concerns have been expressed within and towards the forensic community with a common theme reflecting the philosophical and empirical underpinnings and limitations of analyses and interpretations (Christensen et al., 2014; Mnookin et al., 2011; National Research Council, 2009). These include concerns about over-estimations of the evidential value resulting in wrongful individualisations (Cole, 2005; FBI, 2015; National Research Council, 2004; Pretty and Sweet, 2001; Schwartz, 2005) and issues on the comprehension and transparency of forensic inferences presented and adopted within criminal justice system (Hamer, 2012). Moreover, there has been a growth in discussion of the meaning of various ‘errors’ in forensic processes (Christensen et al., 2014) and more attention has been paid to decision-making processes and the effect they might have on interpretations (Kerkhoff et al., 2015; Nakhaeizadeh, Hanson, and Dozzi, 2014).

Following the concerns discussed in Chapter 2, Chapter 3 has shown more specifically that misunderstandings and misrepresentations of the relevance, probative value, and validity of evidence have resulted in unsafe rulings on a large scale (FBI, 2015; Smit, Morgan, and Lagnado, 2017). Although the importance of robust and transparent interpretations and evaluations in the light of the recognition of limitations and the increasing amount and complexity of data has become clear, the manner in which this can be achieved has not always been so clear cut.

One of the evidence types facing complexities and challenges with interpretations is physical trace evidence. Physical traces refer to any particles which are created and/or displaced, and of which inferences can be made to support the spatial and temporal association of individuals, objects, and/or locations. A profound interpretation requires an identification of their features and possible provenance classes on a source level, and, crucially, understanding their spatial and temporal movements at an activity level (Cook et al., 1998a; Robertson and Roux, 2010; Smit, Morgan, and Lagnado, 2017). Combining the need for systematic approaches to evaluate interpretations in the forensic sciences more generally, and the complexity facing trace evidence specifically, this study aims to develop a guiding framework of the primary components of trace evidence interpretations. Chapter 5 will subsequently discuss the results of applying this framework to evaluate the current state of soil as evidence. These approaches are referred to as 'bottom up' and 'top down' approaches respectively. Where more general guiding principles have been developed previously, which includes the need for probabilistic approaches, limiting contextual bias, highlighting potential errors and assumptions, and being transparent in communications (FOS, 2017), this chapter develops principles on a much more detailed level of interpretations. It discusses the extent to which gaps found between the bottom up and top down approaches are and can be addressed from the current and future knowledge base. More specifically, this chapter aims to identify and articulate theoretical primary components of evidence interpretation, allowing for an improved ability to evaluate and interpret trace evidence in a forensic context and minimising misleading evidence.

Two important aspects are emphasized throughout these sections, namely inherent uncertainties and limitations of soil data and analytical methods on the one hand, and the nature and limitations of the interpretation methods on the other. Moreover, it focuses on three important aspects which have been shown to be misinterpreted at both a source and activity level (see the results in Chapter 3): the relevance (what questions can the evidence address?), the probative value (to what extent can these questions be answered from the observed evidence?) and the validity (how valid are analyses and interpretations?) of the evidence. The results from this chapter contribute to providing an understanding and systematic methodology to evaluate the current and future state of trace interpretations, and thereby supporting robust, transparent, and targeted interpretations to support decision-making processes.

Conceptualising seemingly straightforward fundamental ideas and identifying key components of interpretations is achieved in two ways. First, through the introduction and creation of the new concept of 'trace-scripts' which describe the path of traces over space and time. This is based upon the idea developed in cognitive science of using 'scripts' to represent procedural knowledge (Schank, 1982). As they capture generic patterns, they can be adapted and re-applied across varying scenarios and evidence types.

Moreover, this allows for a systematic analysis and identification of required assumptions and/or uncertainties within these processes. The trace script is developed by making use of logical concepts of possible particle movements and empirical studies on the behaviour of traces and their properties under different circumstances. Second, as a consequence of such a script, and in the light of the use of trace evidence (including associations and individualisations), by identifying the knowledge required for and inherent limitations of these interpretation methods. Key components identified through these processes are defined and summarised in Section 4.5. In sum, the components of this chapter and the relation to Chapter 5 are presented in Figure 4.1.

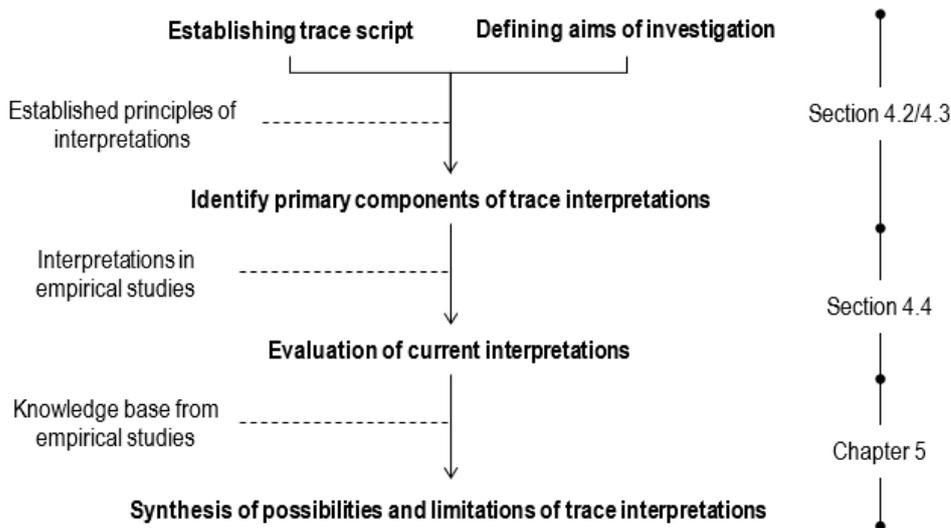


Figure 4.1: graphical view of the sections, aims, and methods in Chapter 4 and 5.

## 4.2 TRACE SCRIPTS

In order to infer the source of traces recovered from a crime scene, an understanding is required into the developmental and behavioural processes of traces over time, expressed in a new concept of ‘trace scripts’. Locard’s exchange principle (Locard, 1934) and Kirk’s principle of individualisation (Kirk, 1963) are traditionally referred to (possibly in derived form) when aiming to understand the meaning of transferred traces (Chan, Tan, and Wong, 2012; Crispino and Houck, 2016; Crispino et al., 2011). However, rather than attributing non-observed transfers to scientific limitations such as recovery rates and detection thresholds (meaning that the propositions are holding due to an argument from ignorance because of the lack of evidence for the contrary proposition), it is imperative that inferences are built upon empirical findings and substantiated theories. Recognising and elaborating on the complexity underpinning Locard’s exchange principle, the study of evidence dynamics (Chisum and Turvey, 2000) refers to the col-

lective effects which “add, change, relocate, obscure, contaminate, or obliterate physical evidence, regardless of intent” (Turvey, 2011, p.266), or more generally, those that influence its division, transfer, and persistence (Inman and Rudin, 2002). This implies that the agglomeration of particles (the sample) transferred is not necessarily the same as the sample that persists over time (Morgan et al., 2009). Additionally, the sample collected after persistence may include irrelevant (background) particles foreign to the original sample (resulting in a mixture from different sources (Morgan and Bull, 2007a)). In contrast, control soil samples from a crime scene, reference samples from previously established datasets, or alibi soil samples (Fitzpatrick and Raven, 2016) (here collectively referred to as ‘reference samples’) are assumed to not have had the transfer and persistence stages (see Figure 4.2). There are, however, challenges with regards to taking representative samples of a location of interest, especially when it is not very homogeneous (Pye et al., 2006). Similarly, the level of homogeneity is also important when assessing the comparability between the collected sample and the (partial) sub-sample which is eventually used for analysis. Furthermore, the validity of a method, expressed as degrees of accuracy (systematic error) and precision (random error) could logically impact results and should therefore be taken into consideration, recognising that the sample is eventually characterised and compared based upon the observed value of features rather than the true value (Christensen et al., 2014). Given these observations, interpretation can include comparisons between traces and/or reference samples, or processes whereby the provenance is being narrowed down (or ‘predicted’) from the general population using any type of database. A generic trace script that incorporates these features of the sample is presented in Figure 4.2 and the key components summarised in Table 4.1. The subsequent sections of this chapter investigate the interpretations made from sediment analysis further, distinguishing between the ability to differentiate locations based on soil features (4.3) and subsequently classifying unknown (trace) samples on the basis of the ability to differentiate between locations (4.4).

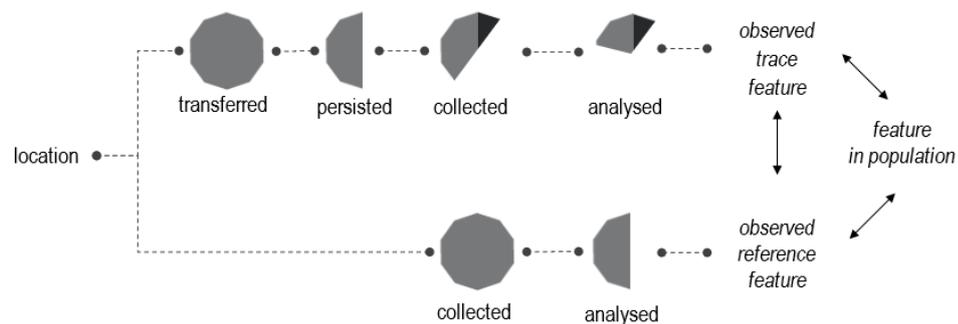


Figure 4.2: Generic trace script for the behaviour of a sample either developed as a trace or collected as reference sample.

### 4.3 DIFFERENTIATING BETWEEN REFERENCE SAMPLES BASED ON OBSERVED FEATURES

When the aim of a geo-forensic enquiry is to infer the origin of trace soil samples, there must be some belief that a (group of) source(s) can be associated with the trace based on some features. In addition, some belief must exist about the ability (to a certain extent) to differentiate between reference locations based on such features. Reference locations can vary spatially, for example based on their land-use and/or underlying bedrock geology. Soil is a complex mixture of materials and has a wide variety of properties, which can be understood according to the manner in which they are observed, using (a combination of) physical and chemical analysis methods. For example, non-intrusive methods ranging from simple microscopy to advanced electromagnetic spectroscopic techniques (e.g. MSP, laser granulometry, ATR-FTIR, LIBS, SEM-EDX, and XRF) can be used to analyse properties including colour (Woods et al., 2014b), particle size distribution (Chazottes, Brocard, and Peyrot, 2004), grain surface texture analysis (Bull et al., 2006), elemental composition (Woods et al., 2014a), mineralogical content (Ruffell and Wiltshire, 2004), radionuclide activity (Dragović and Onjia, 2006), IR absorption spectra (Suarez, Southard, and Parikh, 2015), electrical conductivity (Suarez, Southard, and Parikh, 2015), and magnetic susceptibility (Guedes et al., 2013). Moreover, the organic component of soil may contain forensically relevant indicators such as diatoms (Scott et al., 2014), pollen (Mildenhall, Wiltshire, and Bryant, 2006), fungi (Hawksworth and Wiltshire, 2011) and bacterial DNA (Quaak and Kuiper, 2011). Additionally, chemical assessment techniques using chromatography and mass spectrometry can be used to analyse properties including the elemental composition (Pye et al., 2006), stable isotope ratios (Pye et al., 2006), and plant wax compounds (Mayes et al., 2009). While the term property is used to refer to these general measurable characteristics (such as elemental composition), property features denote the measurements taken to describe these properties (such as the mass percent of a single element). Thus, the analysis of soil samples includes multiple properties which have, in turn, multiple features. Two aspects can be differentiated: the nature of soil and the data that can be obtained from analysing features on the one hand (4.3.1 – 4.3.3), and the nature of interpretations and presentations of such data on the other (4.3.4).

#### 4.3.1 *Diagnosticity of analysis and abundance of property features*

With the knowledge that some loci in human DNA contain sequences which tend to differ between individuals (Jeffreys, Wilson, and Thein, 1985), it is not surprising that those markers are used in forensic identification processes, as comparing DNA sequences known to be the same among humans is of little individualising value. Similarly, in geo-forensic investigations aiming to identify possible sample provenances, the diagnostic value of observed features influences how meaningful the outcome is. This principle

is highlighted in Figure 4.3. As a way of illustration, let the set  $L$  include six possible source locations. Using features of certain property A, these locations can not all be distinguished and are therefore clustered into  $\{1, 2, 3\}$  and  $\{4, 5, 6\}$  (clustering algorithms are discussed in Section 4.4.2.). Additionally, property B can differentiate the clusters  $\{1, 2\}$ ,  $\{3, 4\}$ ,  $\{5\}$  and  $\{6\}$ . By combining these, the clusters can be further fragmented:  $L_{(A,B)} = \{\{1, 2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$ . Evidently, this example is limited to the set of six reference locations, meaning that if other geologically similar locations are included in  $L$ , they could be clustered with existing locations. In the practice of individualisation, this number of other sources based on the observed features is not always considered (Bryant and Mildenhall, 1998; Cole, 2009; Pye and Croft, 2004), resulting in the so-called individualisation fallacy (Koehler and Saks, 2009). Moreover, in addition to considering the between location-variation of a feature, one should also consider its within-location variation, which could result in more overlap between locations. It follows from the example that property A is, based on these reference samples, less diagnostic (less clusters) than property B, and, given the same amount of samples, therefore generally has more objects (locations) within a single cluster (the features occur more often within the reference sample set). This links the case-specific diagnosticity of the analysis with the rarity of observed features (Bull and Morgan, 2006; Bull et al., 2006; Carvalho et al., 2013). When using such arguments to make generic statements about the diagnosticity of a property (compared to another), assumptions must be made about the generalisation of the reference set to the wider population. It must therefore also be accepted that the (alleged) infrequent occurrence of certain features cannot be used as proof for uniqueness (Saks and Koehler, 2008). These concepts are summarised in Table 4.1.

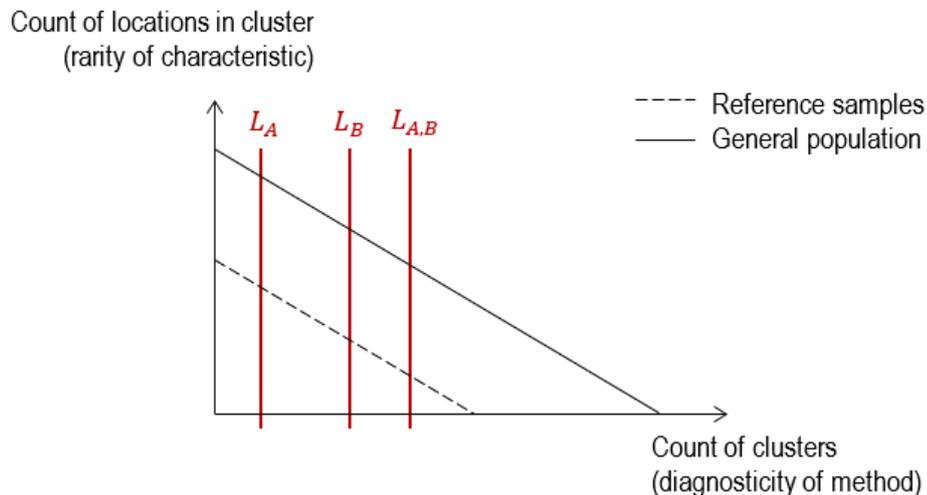


Figure 4.3: Graphical representation of the effect of combining analytical methods and using different populations in the process of clustering samples

#### 4.3.2 *Independence between features*

It must be understood whether there is dependency (e.g. correlation) between the observed features, both within one property (e.g. different elements in an elemental composition or multiple dimensions of colour) and across different properties. On the one hand, this is of value when determining a methodology to gain an understanding of the additional investigative and evaluative contributions of analyses (Cook et al., 1998b; Stoney and Stoney, 2015). Evidently, this can vary for a feature depending upon the geological differences between the reference samples, and the thresholds of when two samples are considered to be 'distinguishable'. Additionally, presenting the analyses of two dependent features as being independent provides a misrepresented view of the value of the interpretation. Whether two methods are dependent upon each other can either be known from a general physical-chemical knowledge base or can be derived from, for example, case-specific data analysis and feature-reduction methods. In relation to Figure 4.3, a sequence of dependent features could be observed when adding a feature does not greatly increase the number of clusters, depending on the difference in diagnosticity of the dependent features or validity of their analytical method.

#### 4.3.3 *Validity of analytical methods*

A difference in the validity of the method (i.e. its systematic error and/or random error) can also affect the way in which samples are (and can be) clustered. Firstly, the method's accuracy determines whether they make the 'correct' measurements (determined by the true negative rate and true positive rate). High error rates would then suggest that samples might be clustered 'wrong' (i.e. not the same as if this was based on the true value). Note that this does not have to be the case if all samples happen to face the same errors and are therefore somewhat comparable. Secondly, however, a significant random error means that the repeatability and/or reproducibility of a method is not very high (not consistent) between different measurements of the same sample. This also affects the final clusters of samples, because although (some) measurements might be very accurate, if over-all their reproducibility is not very high, one cannot rely on the fact that if you would perform the same process on different data (or introduce a new sample), similar clusters would result. In sum, a change in the validity suggests that these clusters are either bigger than they theoretically should be or the samples that have a true same source are not necessarily clustered together. Arguably, this is not only problematic when distinguishing between reference locations, but also when aiming to classify unknown traces (Section 4.4).

#### 4.3.4 *Data exploration and reduction*

The second part considers what can be learned from the discussed soil data with respect to supporting its interpretation, such as different ways of data exploration, dimension reduction, and data clustering. Given the complexity of the observed features, the data structure could be explored further, and, if valuable, reduced to fewer dimensions by excluding or combining features. Data reduction methods facilitate exploring the dependencies between features and highlighting the importance of certain features for that specific context. Moreover, it can be used to assess the added value of individual features, which has become an important part of focus within the geo-forensic sciences (Dawson and Hillier, 2010). For example, by maximising the difference between and minimising the difference within location groups (using supervised methods such as linear discriminant analysis (LDA)) or when aiming to achieve the greatest variance between all samples (without considering groups of samples, using unsupervised methods such as principle component analysis (PCA)). Given the relatively straightforward measurements in geo-forensic analyses, exploring further underlying latent factors which can be associated with one or more features, such as can be achieved using factor analysis, is not necessarily needed.

The extent to which locations can be differentiated based on the observations (either the original feature values or reduced components) can be explored using clustering methods. To ensure reproducibility and transparency of this process, it is important to highlight the details of the clustering algorithm. If known samples (groups) are used, the performance of the clustering algorithm can be tested. While clustering methods themselves do not aim at classifying unknown samples (see Section 4.4), nor do they test whether there are significant differences between groups, statistical tests can be undertaken to test the hypothesis that two clusters have different means based on the feature values. Examples include a t-test for two groups and a single variable, or its multivariate version Hotelling's  $T^2$ , although their results should be interpreted with care if the dataset is relatively small, which could often be the case within soil investigations.

#### 4.4 CLASSIFYING UNKNOWN TRACES

An understanding about whether one can differentiate (specific) reference samples, suggests that if a trace is found, comparisons between this sample and the reference samples can provide valuable information in an investigation. More specifically, this chapter has so far discussed the knowledge that a soil sample consists of many (dependent) features, which can be reduced in dimensions, and on which clustering algorithms can be applied to understand the usefulness of these features to distinguish these specific samples. This section aims to address the knowledge required for the ultimate aim in forensic investi-

gations: classifying an unknown sample and thereby providing information on possible sources of its provenance.

#### 4.4.1 *Hypotheses and the (strength of) evidence*

As the observed evidence can be regarded as the consequence of some activity related to the criminal act, the antecedents are described as hypotheses. For example, an hypothesis (H<sub>1</sub>) can be “location A is the source of the trace”, and its mutually exclusive hypothesis H<sub>2</sub>: “another location is the source of the trace”, in what is referred to as the classification problem (Zadora et al., 2013). While H<sub>2</sub> is mutually exclusive and exhaustive to H<sub>1</sub> assuming ‘another location’ is any location theoretically possible, they are no longer exhaustive when these other locations are only those consisting of reference samples taken (see Figure 4.3). Another way of expressing these hypotheses is by comparing specific samples in the so-called comparison problem, considering whether the samples have ‘the same’ or ‘a different’ origin (Zadora et al., 2013). Similarly, the evidence (observations from analyses in some form) and the diagnostic support given to an hypothesis by such evidence (the likelihood) can be expressed in different ways. For example, score-based methods can be used to assess the likelihood of observing the evidence as a difference between the features in some samples if they are from the same location versus if they are from a different location (Bolck, Ni, and Lopatka, 2014). In contrast, feature-based methods aim at assessing the likelihood of, for example, observing the evidence given that they come from specific location A rather than from B. Whilst the first problem requires knowledge purely on the differences or similarities in observations from samples from similar or different sources (and is therefore quite generic), the second problem requires an expectation about the specific feature values. Therefore, score-based methods lose the original feature values and therefore use less information as part of the interpretation, meaning that although feature-based models are more complex, they can be preferable in a forensic context (Bolck, Ni, and Lopatka, 2014). However, sometimes the results are presented as comparisons, for example when suggesting ‘categories of comparability’ including limited, moderate, and extremely strong to conclusive, depending upon the degree of comparability for a (combination of) characteristic(s) (Fitzpatrick and Raven, 2016).

Assuming a feature-based method, more complexities can arise. For example, it is known that ‘the trace’ which is collected and analysed can consist of an agglomeration of particles from different sources (Section 4.2). If this is the case, the hypotheses change to the hypothesis H<sub>1</sub> and H<sub>2</sub> presented below (Vooijs, Vergeer, and Weerd, 2015), which increases the chance of finding a random ‘association’ between certain features of the trace and reference location due to the increased numbers of comparisons, often referred to as the ‘two-trace problem’ (Stoney, 1994; Vooijs, Vergeer, and Weerd, 2015). Here, the

general term 'association' is used to refer to any measure used to express the extent of similarity of samples.

H1: location A is the source of some of the soil traces

H2: other locations are the source of the soil traces

The likelihood ratio (LR) expresses the ratio of likelihood of observing the evidence (e.g. feature values or a similarity measure) if H1 is true over H2 is true. Then, if multiple reference samples taken for comparison (e.g. H1 changes to 'one of the sampled locations is the source of some of the soil traces') the chance of finding a random association increases even further, meaning that the LR decreases (assuming independence). Moreover, an extra level of complexity is added when several of these sampled locations can be associated with some of the traces. The new hypothesis H1 "some of the reference materials are donors of some of the traces" then becomes more likely due to the fact that there is more than one association. Therefore, presenting these hypotheses and being explicit about the reference population and the way in which observations are used as evidence is an essential part of the interpretation process.

#### 4.4.2 *Classifications, exclusions, and posterior likelihoods*

Thus, the value of the evidence in an investigation is very much dependent upon the hypotheses that are put forward. The diagnostic support that the evidence (E) gives to the hypothesis (H) after it is observed is expressed as the likelihood  $P(E|H)$ , which can be estimated directly from the specific context without needing much more information from the knowledge base (Pearl 1988). Given two competing hypotheses H1 and H2 (for example, those put forward by the defence and prosecution), the ratio of their initial probabilities can be described by the prior odds. After a new item of evidence (E) becomes available (of which the evidentiary value is described by the LR), the posterior odds can be obtained using the odds form of Bayes' theorem as discussed in Chapter 2.

This belief updating makes intuitive sense, namely, in order to make a well-informed inference about the belief in an hypothesis given some evidence, it would be naïve to only consider what has been learned from observations and forget what was previously known. The potential of using the LR in determining the value of the evidence in light of an hypothesis has been discussed widely in the forensic literature (e.g. (Aitken and Lucy, 2004; Champod and Taroni, 1997; Evett, 1987; Wakefield et al., 1991)). However, real cases are comprised of a complex set of many (conditionally) dependent items of evidence and hypotheses. Obtaining the joint distribution requires a full inference over the entire set (Aitken and Taroni, 2004; Smit et al., 2016), which becomes unmanageable to a forensic scientists and incomprehensible for others in the criminal justice system as the set expands (Dawid and Evett, 1997). To overcome this problem, the potential use of a graphical representation of such relationships in Bayesian networks (Pearl, 1988) has

been introduced within the forensic domain (Aitken and Gammerman, 1989; Dawid and Evett, 1997), whereby domain knowledge is used in developing the structure of a BN by the inductive reasoning process of inferring consequences from some fixed premises in the so-called cause-effect approach (Pearl, 1988).

Within geo-forensic investigations, the observed data is used to classify the sample into a group of possible provenances (ideally a group that contains very little reference locations). A range of different classification methods can be applied, including decision trees, linear classifiers, kernel estimations, and neural networks. Because such analyses result in a posterior probability distribution over all possible sources (the prior probability is often automatically taken as 0.5, but should be acknowledged for full transparency), some transparent decision-making method should be applied when one wants to draw a final conclusion as to the source of the trace (e.g. the location which has the highest posterior odds in favour of the hypothesis that it is indeed the source).

As discussed previously, because of the nature and abundance of soil (data) it is often impossible to assess the possibility that the sample originates from, for example, “another location than site A”. It has therefore been suggested that the principle of exclusion is a more valid method for drawing conclusions (Bull & Morgan 2006). Here, an exclusion depicts the decision that the sample could not have come from a specific source based upon the differences in observed features, and a ‘non-exclusion’ suggests the decision that one is unable to differentiate a trace from (multiple) possible provenances (Bull and Morgan, 2006; Bull, Parker, and Morgan, 2006; Morgan et al., 2010; Pye and Croft, 2004). With regards to Figure 4.3, the ‘non-exclusionary’ value can, for example, be represented in the amount of locations remaining in a cluster, once all methods have been employed and none of the other locations in that cluster could have been excluded based on these features.

#### 4.5 SUMMARY OF PRIMARY COMPONENTS

In order to use the discussions within Sections 4.2 - 4.4 as guiding principles to evaluate and support the interpretation of (geo-)forensic traces, they have been summarised and conceptualised into ‘components of trace evidence interpretations’, see Table 4.1. There are four general component categories, each corresponding to different aspects of an interpretation. The sub-components within each group should be acknowledged and ideally significantly discussed in soil interpretations, both in terms of factors which can influence these components as well as the way they are ultimately considered in the final conclusion and possible decision-making steps. This framework now allows for a coherent and structural investigation and evaluation of the transparency, validity, and robustness of current interpretations. The result of this for interpretations within published soil studies is presented in Chapter 5.

Table 4.1: Key components which have been identified throughout this chapter which have to be discussed, understood, and/or made transparent as part of the interpretation of (soil as) trace evidence.

Components of trace evidence interpretations	
A.	Comparability of features between a collected reference sample and its trace
A1	Selective transfer of the trace
A2	Selective persistence of the trace
A3	Representativeness of the collected trace sample to the persisted trace
A4	Representativeness of the collected reference samples to the referenced site
B.	Validity of analytical method with respect to the collected sample
B1	Representativeness of the analyzed sample to the collected and/or original sample
B2	Validity of the analytical method and observed features
C.	Diagnosticity of observed features
C1	Independence between features (within and between properties)
C2	(Added) value of features
C3	Population frequencies of individual and combined features
D.	Method of interpretation
D1	Use and presentation of competing hypotheses
D2	Details of classification and/or clustering algorithm
D3	Use of individual and/or combined properties
D4	Likelihoods estimations
D5	Assumptions in decision-making process

#### 4.6 CONCLUSION

This chapter has shown the value of describing the logical sequence of events of trace evidence in a framework, referred to as a 'trace script'. The trace script developed in this chapter is based upon the current understanding of the behaviour of traces as causal relationships that follow from the results of empirical studies. Although the (forensic) knowledge base of these relationships and influencing factors is growing over time, knowledge concerning causal relationships can be applied across evidence types and between different phases within the same transcript. Generically developed, these frameworks can be applied to the interpretation of specific trace evidence types or individual cases. This fits in with the idea of using idioms which have previously been introduced in the context of Bayesian networks (Lagnado, Fenton, and Neil, 2013). Following these trace scripts, this chapter has shown the possibility of structurally identifying key components, breaking down (the evaluation of) interpretations into smaller sections which can be studied for each subject individually and subsequently evaluated across all case studies combined. Although the interpretations in the study subjects cannot be affected by the present research design, questions can be raised on the representativeness of the interpretation methods used in the published studies as compared to actual practice and detailed expert statements. Nevertheless, the findings presented in this chapter highlight the consequences of not adequately addressing key components to the strength of interpretations and the ability of making informed decisions. Moreover, such a frame-

work allows for organising the current knowledge base (and its gaps) and to highlight challenges within current practice. This can subsequently be used to efficiently and effectively express the belief in hypotheses using holistic methods of interpretation such as Bayesian networks, by highlighting aspects which require special attention and to support developing the network structure and the extent to which the observation of evidence impacts the belief in key hypotheses.



## CHAPTER V

### EVALUATING CURRENT INTERPRETATIONS OF SOIL EVIDENCE

#### 5.1 INTRODUCTION

While a range of uncertainties have to be considered for all trace evidence types during the interpretation process, the use of soil as evidence faces many of these to a significant extent, such as contamination and/or the presence of mixtures affecting features of a sample (Morgan et al., 2006; Wilks, Morgan, and Rose, 2017) and the abundance and variation of the complex set of soil features in a population (Ruffell and Sandiford, 2011). Therefore, forensic geoscience requires the adoption of fundamental philosophical and methodological principles from the classic geoscience field, including the exclusion of samples, the rarity and independence of features, and the nature of the analytical method (Morgan and Bull, 2007a). Moreover, deciding upon analytical and interpretative methods often depends to a great extent on the availability and opinions of an expert (Morgan and Bull, 2007b), suggesting that interpretations face a combination of explicit and tacit information, whereby beliefs in hypotheses are situational and depend on the information available to an expert at that moment (Lindley, 2014). Ultimately, the complexities underlying soil evidence in the light of future directions on combined approaches (Dawson and Hillier, 2010) and statistical analyses (Aitken, 2009) suggest that soil evidence could greatly benefit not only from best procedure from collection to analysis (see for example (Fitzpatrick and Raven, 2016)), but also from approaches to increase transparency and robustness when drawing inferences (Robertson, 2009), and, ultimately, best practices to evaluate those inferences.

Following the idealistic and theoretical bottom-up approach from Chapter 4, a range of empirical studies presenting soil as evidence is used to investigate the extent to which these key components are being addressed in published studies within this field. The aim of the top-down approach is meta-synthetic in nature: rather than increase certainty by combining findings of cause-effect studies, it aims to understand and explain phenomena (Walsh, 2005) and gain insights beyond the primary findings of those individual studies (Hoon, 2013). Such an analysis lies between a literature analysis and a meta-analysis combining the primary empirical findings, and is more qualitative in nature. It includes, after a framing of the research question, an identification and filtering of research studies, an extraction of the data, an analysis on a case-specific level, a

synthesis between-studies, and theory development and discussion (Hoon, 2013).

More specifically, the aim of this chapter is to apply the framework developed in Chapter 4 (see Table 4.1) in order to gain an understanding into the interpretation of soil as evidence. More specifically, it aims to evaluate the way in which geo-forensic evidence is currently interpreted in the light of the primary concepts, using a meta-synthetic approach on a set of empirical studies as case examples. The aim here is not to criticise these studies, but to use them to identify challenges and limitations (which could be) faced in these interpretations more generally.

## 5.2 METHODS AND MATERIALS

The data used in the present study consist of published studies because of their often detailed information and availability. Whilst they do not necessarily reflect the use of soil as evidence in criminal investigations more generally, these papers do often discuss real cases the authors have worked on, and it is assumed that the analysis- and interpretation methods can and are being employed in other contexts. This study therefore faces less limitations usually seen when using case studies, such as the limited possibility of generalising findings (Blaikie, 2010). The first set of research studies were selected from the first 200 'most-relevant' references in English of 'forensic' and 'soil' search results in Google Scholar, adding relevant papers that these studies cited as well as papers in which they were cited. Due to the complexity of soil and its wide range of measurable characteristics (hereafter called properties), the decision was made to only include studies that employ more than two properties, excluding those that did not include (any of the) more commonly used such as colour, particle size distribution, elemental composition, and mineralogy. Source provenances include references samples collected specifically for the purpose of the case or can be generic locations based on a number of factors such as flora and fauna, underlying bedrock geology, or land use. The unknown samples can include both the reference samples of unknown locations (and therefore skip transfer and persistence stages, see Figure 4.2), or can be actual trace samples. Depending on the type of samples used and number of (sub-) analyses performed, a range of information can be obtained from these studies, which are highlighted in Table 5.1.

The analysis of this data consists of an identification of noticeable challenges, inconsistencies, and/or positive observations within each case study and for each component group defined in Chapter 4. These observations are subsequently discussed for each component group across all studies and within the wider scope of the forensic science discipline moving forward.

Table 5.1: Overview of the different type of aims of a study depending on their samples used.

	No trace samples	Trace samples
Single location sample	Method validity	Method validity Evidence dynamics
Multiple location samples	Method validity Diagnosticity of features	Method validity Evidence dynamics Diagnosticity of features

### 5.3 RESULTS

The type of reference and trace samples were identified for each of the 19 case studies and can be found in Table 5.2, together with the soil properties which were analysed and information on whether it involved a real criminal case. For each individual study, each of the sub-components for each general component group (A through D) were studied. Noteworthy points were made on positive observations, acknowledgements and discussions of these components within each paper (the black boxes) as well as noteworthy points on aspects within interpretations which were not addressed fully, transparently, coherently, or in any other way to the extent possible (the white boxes). Table 5.2 provides an overview of these results, and these observations are discussed in more detail within each paper and across all papers for each component group in sections 5.4 through 5.7.

#### 5.4 COMPARABILITY OF FEATURES BETWEEN A COLLECTED REFERENCE SAMPLE AND ITS ORIGIN (A)

While some studies in Table 5.2 only include reference samples (Guedes et al., 2011; Suarez, Southard, and Parikh, 2015), and therefore do not consider the transfer and persistence stages, reference samples have been used to depict questioned samples as ‘blind locations’ (Bonetti and Quarino, 2014). Vice versa, trace samples have been interpreted as general reference samples (and therefore do not explicitly consider the effects of transfer and persistence discussed in Figure 4.2) (Woods et al., 2016). This could have implications on the decision presented in the paper that “neither of these scenes could have been the source” based on the observation that features “could be differentiated” (Woods et al., 2016, p.51), as the possible role of transfer and persistence on these differences are not considered. In the case where trace to scene discrimination is not possible, it could be of interest to study the trace to trace discrimination, in order to understand the extent to which those would be clustered into different or similar sources. Supporting the idea that more focus should be laid on addressing activity level propositions, Murray et al. (2016) consider the transfer pattern of soil, including the change in features within this pattern. Although not necessarily aiming to classify the trace samples back to a specific source, they do highlight the importance of ‘selective’ transfer for a range of features under different conditions. Other cases in Table 5.2, although being

Reference sample variations	Trace variables	Case	C	PSD	IR	Min	CC	pH	OM	A	B	C	D	Reference
										Comparability	Method validity	Features	Interpretations	
Geographic area	IED traces	-	-	+	-	+	-	-	+	■ secondary transfer			■ exclusion, classification	(Wilks et al. 2017)
Textures, wet/dry	Era, direction	-	+	+	-	+	-	-	+	■ transfer patterns			■ activity level	(Murray et al. 2016)
Sites	Shoes, shovels	-	+	-	+	-	+	-	-	■ collection method □ transfer	■ blind study	□ independence ■ within-site	□ clustering, decisions □ combined	(Woods et al. 2016)
County, land use	-	-	+	+	+	+	+	+	-	■ collection method	■ blind study	■ added value	■ combined	(Suarez et al. 2015)
State parks, seasons	Blind locations	-	-	+	-	-	-	+	+	■ collection method			■ classification □ decisions	(Bonetti & Quarino 2014)
Database, scenes	Shoes, shovels	+	+	-	-	+	+	+	-	■ procedure	■ procedure	■ population	□ likelihood, hypotheses	(Fitzpatrick & Raven 2012)
Crime scenes	Shoes, socks	+	+	+	-	+	-	-	-	■ transfer, persistence			■ hypotheses	(Ruffell & Sandiford 2011)
Sites, bedrock	-	-	+	+	-	-	-	-	-			□ within-site	■ combined	(Guedes et al. 2011)
Sites	Trousers	+	+	-	-	+	-	-	-	■ selective transfer		■ population	■ exclusion	(Ruffell & Dawson 2009)
Crime scene	Shoe	+	+	+	+	+	+	+	-			■ population	□ hypotheses	(Fitzpatrick et al. 2009)
Crime scene	Scene, clothes	+	+	+	-	+	+	-	+			■ population	□ likelihood, hypotheses	(Gradusova & Nesterna 2009)
Crime scene area	Slippers	+	+	-	-	+	+	-	+	■ size fraction				(Lombardi 2009)
Location, depth	-	-	-	+	-	+	+	-	+	■ within-site	■ procedure	■ added value	□ clustering	(Melo et al. 2008)
Sites, crime scene	Van, clothes	+	+	+	-	+	+	+	+	□ transfer, persistence		■ population	■ exclusion □ decisions	(Petrao et al. 2008)
Crime scene, database	Spade, rope, shoes	+	+	+	-	+	+	+	+	■ collection method	■ mixing	■ population □ independence	■ exclusion	(Morgan et al. 2006)
Crime scene, sites	Shoes	+	+	+	-	+	+	+	+	■ mixing		■ population	■ decisions	(Bull et al. 2006)
Within/betw. sites	-	-	+	+	-	-	+	-	-	■ collection method	■ sub-sampling	■ added value	□ combined	(Pye et al. 2006)
Location, bedrock	Shoes	-	+	+	-	-	+	-	-	■ transfer	■ sub-sampling	■ added value	□ hypotheses	(Croft & Pye 2004b)
Sites	Clothes, car	+	+	+	-	+	+	-	+			■ population	■ activity level	(Lombardi 1999)

Table 5.2: Overview of observations for each component group from the used case studies. Whether some of the key properties have been used are highlighted, including colour (C), particle size distribution (PSD), IR spectra (IR), mineralogy (Min), other chemical compositions (CC), pH, and organic matter (OM, not included in IR). Some papers use additional ones not shown. Where multiple case studies were used in a paper, all noticeable points are combined. Where a filled box suggests that key components have been thoroughly discussed, an empty box suggest that more could have been done or a better interpretation could have been made given the context and data.

the results of mimicked scenarios, specifically consider the effects of primary and even secondary transfer for the comparability of features, acknowledging their importance in the seek-and-find process (Wilks, Morgan, and Rose, 2017). Another approach to this issue is highlighting the assumptions with regards to the decision that selective transfer and persistence were not considered to be a major issue for the comparison process (Ruffell and Sandiford, 2011). Moreover, the comparability of samples can be considered inexplicitly, for example when a trace is only compared with a certain fraction of the site sample (Fitzpatrick, Raven, and Forrester, 2009). This, however, would have required the assumption that these fractions are more comparable after transfer, which was made more clear by others (Lombardi, 2009).

In some of the examples in the various cases discussed by Petraco et al. (2008), generic ‘significant differences’ for the reader to observe resulted in samples being excluded from a specific source. However, ‘differences’ in another example (although not ‘significant’ and within a context where it is suggested that more movements (i.e. evidence dynamics) took place) are suggested to be ‘easily accounted for’ due to such transfer and persistence affecting the features. Although generically understandable from the wording and case examples, what is unclear is where, for example, boundaries are between ‘significant’ and simply ‘different’, and when the case circumstances allow one to conclude that transfer and persistence could be an explanation. This is not to say that such arguments are easy to conceptualise, but at least these cases could be compared relatively to one another and their contexts (e.g. which features, difference in time, type of activity, et cetera), in order to place future observations on this scale.

With regards to the sampling strategy utilised in each study, the effect of the different approaches on the interpretation should be made clear. For example, taking the reference samples as multiple samples within a single location (Bonetti and Quarino, 2014; Woods et al., 2016) has the advantage of the ability to account for both within- and between site variabilities. Samples have also been taken at different depths at the same location to account for such variabilities (Melo et al., 2008; Morgan et al., 2006). Other cases in Table 5.2 show the added value of taking more samples within a grid to obtain the most optimal approximation of the mean value for the entire area (Pye et al., 2006). Although valuable for improving the collection process, the variability of soil makes it challenging to generalise and directly apply the most optimal approach in one case to others. Additionally, the effects of not being transparent about the sampling method has been made clear by Bonetti and Quarino (2014), showing that the results of the classification of the unknown samples in the blind study (using hold-one-out cross validation) varies depending on the distance between sample sites as well as, logically, how many other reference samples were considered as possibilities and how similar they were to the ‘correct’ location.

### 5.5 VALIDITY OF ANALYTICAL METHOD WITH RESPECT TO THE COLLECTED SAMPLE (B)

The second group in the interpretation framework (Table 4.1) includes both the representativeness of the analysed sample with respect to the collected sample as well as the validity of the analytical method. As there is general consensus about the validity of the most commonly used methods to analyse the features of a soil sample, the testing of method validity (i.e. systematic and random errors) is not always explicitly discussed in the trace-location association studies in Table 5.2. However, to avoid issues with the reproducibility of the analytical method itself, some studies highlight the procedures in detail (Fitzpatrick and Raven, 2012; Melo et al., 2008). Other studies which use known samples include blind trials to specifically test the validity of the method or protocol used (Bonetti and Quarino, 2014; Woods et al., 2016). However, if the samples used in the blind study were taken from a database rather than the collected samples (Woods et al., 2016), assumptions must be made about the comparability of the validity of the methods of the case study samples and database samples.

Similar to where the sampling strategy for the collection of reference samples in an area or on an object should consider the homogeneity of the area (e.g. a crime scene), the same should be the case when taking sub-samples of the collected sample for analysis. Taking such sub-samples and analysing each multiple times make it possible to argue about the validity of both the homogeneity of the collected sample as well as the validity of the analytical method (Pye et al., 2006). Moreover, where a sample consists of a mixture of different sources (either because it is a mixed trace or because the reference sample is a mixture of multiple samples), depending on the properties used, it must rightly be acknowledged that the analytical results could only support exclusionary decisions (Morgan et al., 2006).

### 5.6 DIAGNOSTICITY OF OBSERVED FEATURES (C)

Once samples are collected and different analytical methods are employed to identify feature values, different ways of combining these parallel lines of evidence have been used. In many of the studies in Table 5.2, the properties are interpreted individually (for example, using a clustering method) and only collectively discussed afterwards. However, where, for example, Petraco et al. (2008) discuss different properties arguing their conclusion is “clearly shown”, it is suggested that the individual properties add independent weight to the conclusion, although no arguments have been made about their dependencies. Similarly, where data is available and detailed interpretations are made on individual features (Woods et al., 2016), no attempts have been made to study the dependency between these features, which would clarify the strength of the final interpretation, rather than assuming their independence. Although in other cases clear

remarks are made that the variables involved were not dependent upon each other, clearly adding weight to the observation of similarity (Morgan et al., 2006), this argument would have been strengthened by highlighting how to study correlations between features based on the available data(bases). In cases where PCA and LDA spaces are used to classify the samples (Bonetti and Quarino, 2014), although not required for the purpose of the study, more insights could have been gained on the relationship between the features from these components.

Studying the dependency between features is closely related to the idea of studying the added value of individual features for, for example, the purpose of distinguishing locations. With an increase in features and ability to study (the abundance of) these features, deciding upon the most optimal sequence of analyses is becoming more important to advance the field and its use in investigations. Such information gain can be presented more generically, for example by presenting the possibilities of determining very generic geographic origins and the value of performing additional analyses in narrowing down such origins in a seek-and-find investigation (Wilks, Morgan, and Rose, 2017). This is studied more specifically in other cases in Table 5.2, from the added value of individual features after LDA, based on error rates gained through hold-one-out validation (Bonetti and Quarino, 2014), or by using the loadings following PCA to highlight the most important features to distinguish locations (Melo et al., 2008). Moreover, the diagnosticity of features when studying small scale spatial variability of individual features within a property (e.g. a single element) or combined features across properties has been highlighted from the coefficient of variation (Pye et al., 2006). It must be highlighted that in these cases, deciding upon the features with the highest diagnosticity depends on the hypotheses (Section 5.7). Stating the details of these analyses can avoid tunnel-vision, whereby, for example, selected features based on differentiating locations A and B are readily applied when a possible source C is added. More transparency in the analysis can also avoid the possibility to ‘direct’ the source of the trace towards a certain location simply by selecting some ‘favourable’ characteristics.

As argued in Chapter 4, one of the key parts in the interpretation process is acknowledging the population frequencies of observed features, whereby the definition (and size) of this population also depends on the defined hypotheses. For example, Gradusova and Nesterina (2009) rightly acknowledge that their comparison only includes the reference samples taken, and highlight the possibility that the sample originated from somewhere else not part of the chosen population (see Figure 4.3). However, no other information is given about these possible other sources, and while this is difficult to put into context as these locations are not specifically studied, these findings can be made more clear by highlighting specific hypotheses studied and, for example, transparent estimations of a (specific) wider population. When larger datasets are used as comparison, the value has been shown of not just looking at the population frequencies

of features (e.g. the presence of a group type of surface textures), but also, and crucially in this case, the population frequencies of a combined groups (Morgan et al., 2006). Additionally, where no specific reference samples are taken, generalised soil maps can be used to narrow the origin down to a generic area, to subsequently gather intelligence or take specific samples for comparison (Fitzpatrick and Raven, 2012; Ruffell and Dawson, 2009).

## 5.7 INTERPRETATION OF OBSERVED FEATURES (TRACE SAMPLE AND REFERENCE LOCATION SAMPLE) (D)

### 5.7.1 *Clustering and classifications*

As mentioned when discussing feature dependencies, many studies use the properties as parallel lines of evidence which only converge when drawing conclusions. While in some cases bringing these lines of evidence together in the initial interpretation stage is less obvious as no specific (numerical) results are presented for individual properties or features (Fitzpatrick and Raven, 2012), others could have achieved this as detailed individual results are available and can be normalised (Bull, Parker, and Morgan, 2006). Where combining the features of different properties, methods including hierarchical cluster analysis were used (Guedes et al., 2011). However, no within-site samples are included in the analysis, although they would have provided a stronger argument when differentiate these locations more generally, rather than attributing this difference to generic variability between samples, even on a small scale.

In many cases, the results of the PCA were directly used to draw conclusions about the level of separation between locations or populations, for example by arguing that “a clear separation was found” between populations (Pye et al., 2006) or by highlighting clusters which emerged (Melo et al., 2008). The matrix and PCA plots of Woods et al. (2016) explicitly base the separation of samples on comparing the intra- and inter-site variation of features to argue about the extent to which samples (locations) can be “clearly differentiated” and whether traces are “outside the scene variability” of a location, providing clusters within these samples. In all these cases, however, it has not been made clear what clustering algorithm (if any) was used. Suarez et al. (2015) also depicted the results as a matrix plot, stating the observation of some visual observations with regards to separation based on location type. However, visual observations without any thresholds or a lack in presenting clustering details make reproducing these findings and comparing these with the effects of other clustering methods not possible. Other cases have used both PCA and LDA to interpret the feature values (Bonetti and Quarino, 2014). However, the difference in results between these interpretations is not discussed in the light of the different supervisory nature of the methods (i.e. LDA often results in better differentiation between groups because it takes into consideration group

labels). Lastly, hypothesis testing (such as a t-test) has also been used to assess the differences between groups (Pye et al., 2006; Suarez, Southard, and Parikh, 2015). Although this will indicate a difference between mean values, it does not necessarily provide information about how different the groups are, which would be desirable when aiming to classify unknown samples.

In terms of such classifications, Bonetti and Quarino (2014) acknowledge that considering the unknown sample as a separate 'group' in the LDA rather than applying the transformation on the unknown sample at a later stage is a conservative approach. Whilst they use transparent distance measures as part of their classification, their arguments are also based on the very less-transparent visual method (Bonetti and Quarino, 2014). In contrast, Suarez et al. (2015) use PCA to provide visual 'signs of separation', and LDA to achieve a more detailed analysis, including the level of none to some misclassifications (possible due to using known samples), although it is not clear whether they, for example, used a leave-one-out method. Others have used the discriminating power to study the samples, evaluating the number of sample pairs which can be discriminated relative to all pairs compared (Smalldon and Moffat, 1973). However, to understand those results in context, for example when obtaining a different result from pairwise comparisons and clustering algorithms (Woods et al., 2014a), information must be given about the details of such methods and thereby the reason why these differences are shown in the interpretation.

### 5.7.2 Hypotheses

When presenting the results of 'correct' clusters or the classification of unknown samples, it must be highlighted that these are highly dependent upon the reference samples which are used, meaning that adding samples from other locations might change these results. It is therefore of key importance that the populations studied, and thereby the hypotheses examined are made clear. For example, Ruffell and Sandiford (2011) use soil as an additional item of evidence to "exclude or compare" traces on a victim to a possible crime scene. Although very few reference sites are used, it is made clear that traces were of "visually and mineralogically comparable makeup" to one of the locations, "to the exclusion of" comparison locations, including considering the possibility that the trace profile can be "coincidentally comparable" to unknown locations (Ruffell and Sandiford, 2011). It can also be the case that rather than attributing a specific origin to the samples, it is, for example, evaluated whether samples have come from similar or different (unknown) locations (Petrao, Kubic, and Petrao, 2008). Lastly, when studying significant differences between traces and their known origin (Croft and Pye, 2004b), the results would have more impact when also comparing the traces to other locations of which data is known. This need would have been more apparent if a clear presentation of the competing hypotheses was given.

### 5.7.3 *Conclusions and decisions*

A range of conclusion types are used to summarize the results of the interpretation of findings. For example, based on the consistency in the makeup of the samples, the sample “could have originated from” the source, whereas other reference samples were dissimilar to the questioned sample (Petraco, Kubic, and Petraco, 2008). Another example in the same study uses the terminology to “associate between” samples, suggesting that these arguments are presented as being different ‘levels’ of non-exclusions. Although it is unclear whether the subsequent statements made on the source of questioned samples involved more analyses, in any case, arguments about the association between samples requires more details about which samples were included in the comparison. Other conclusions are based upon significant similarities between samples, which were shown to be rare from a database in one instance and from extensive experience in another (Morgan et al., 2006). Moreover, Morgan et al. (2006) highlight the amount of data required to draw conclusions, also acknowledged by others, clarifying that “sufficient” data was acquired to be able to determine if they “compare” or “do not compare” (Fitzpatrick, Raven, and Forrester, 2009).

Following the observation of sample clusters, many conclusions were subsequently drawn without acknowledging that, because of the presence of uncertainties, this involved a decision-making step. For example, in Woods et al. (2016) the specific decision of the source location directly follows the dissimilarity-observation. This decision is therefore based upon the assumption that the sample came from one of the sampled sites. Additionally, in some cases it can be derived from the text what the decision-making steps (deciding upon the source of a sample) are based upon (e.g. the smallest Mahalanobis distance), the fact that a visual comparison was additionally used to come to the decision was not explained in the same way (Bonetti and Quarino, 2014). Much can also be learned from the way in which Gradusova and Nesterina (2009) draw conclusions for their different cases. For example, in the first case, it is argued that the trace soil samples were “with a very high degree of probability, from the place under the particular window in question, or from another place with soil that had the same features”, without giving any indication on such other places in the chosen population. Additionally, the terminology between the given examples are different, including statements such as “to a high degree of probability” and “to a high degree of certainty” (Gradusova and Nesterina, 2009), which challenging to compare if no context is provided. Others have also use probabilistic statements, such as arguing that “the soil from the shoe most likely sourced from” the sampled site based on a high degree of similarity between the samples (Fitzpatrick, Raven, and Forrester, 2009) and similar arguments (Fitzpatrick and Raven, 2012). However, there was little clarity about the alternative hypothesis (e.g. whether it most likely sourced from site A rather than that it did not source from site A, or that it most likely sourced from site A as compared to site B). In addition, these

statement suggest the presence of the prosecutor's fallacy, expressing the posterior probability as the likelihood ratio based on the evidence of similarities, without considering the prior probability. More bold statements have been discussed (Petrao, Kubic, and Petrao, 2008), whereby the expert concluded and testified in the case that, in his opinion, the traces were obtained from a specific location. It must be understood from the discussion in this paper that such statements are often beyond what can be known and should be accompanied by the limitations they face.

## 5.8 DISCUSSION AND CONCLUSION

It follows from the results in this chapter that the challenges specifically within interpreting soil as evidence arise in a) utilising the most suitable data analytical methods to soil evidence and b) being robust and transparent on the meaning of analyses in empirical studies and criminal cases. More specifically, this chapter has identified several main issues. First of all, although the idea of evidence dynamics (such as transfer and persistence) is often being acknowledged, many of the cases studied address these factors simply by stating assumptions, either that their effect can be diminished or that their influence can explain the observed differences. Not only should thresholds be given of where the boundaries are between these stages, or at least a justification for these choices, but they should also be supported by empirical arguments. Moreover, where the case context has been shown to be of value for such arguments (Morgan and Bull, 2007c), discussions and investigations into decision-making processes are needed within soil analysis given the effects of contextual information in analyses and interpretations in other forensic fields (Nakhaeizadeh, Dror, and Morgan, 2014; Osborne, Taylor, and Zajac, 2016).

To achieve the goal of making more evidence-based interpretations, the present and future data and knowledge from case-based inferences and empirical studies should effectively and efficiently be grouped into knowledge frameworks. It is therefore the responsibility for authors and experts to design experiments and present findings in ways to utilise these as best as possible. For example, where Croft and Pye (2004) determine instrument validity through the coefficient of variation (CV) within a group of several sub-samples each tested several times, it does not become clear whether the difference can be attributed to sample homogeneity (B1) or analytical method validity (B2). By comparing the CV between soils from different locations, however, arguments can be made about their differences in homogeneity, although assuming similar method validity (Croft and Pye, 2004a). Similarly, repeat analyses on the same partition of a sample (Guedes et al., 2009) can provide knowledge on method validity rather than its homogeneity. It also follows from these findings that crucial details and assumptions of interpretations are often lacking, whereby such unclarities could result in presenting

biased outcomes.

With a growth in the use of analytical methods (and thereby analytical data), and innovative methods such as bacterial DNA (Concheri et al., 2011), understanding the ways in which data can be combined, including the meaning of results and their limitations, are becoming increasingly more important. It can be concluded from this chapter that authors are not always being transparent on the clustering and classification methods, despite this being necessary to make interpretations comparable, reproducible, and justifiable. Additionally, it follows from these results that the independence between features (or properties more generally) is either simply assumed or not explicitly discussed, despite it often being possible given the amount and type of data. This is especially important within using soil as evidence due to the need for combining multiple properties and their contribution to the value of a comparison. For example, Morgan et al. (2006) discuss the idea that the combination of methods used were not able to exclude two samples (of which some individual properties were argued to be rare, or where some features were observed to be so different), making the argument that it was 'extremely unlikely' that they came from the same source. This suggests that there is some scale on which to place these observations, including thresholds between (levels within) exclusions and non-exclusions. Where attempts have been made to highlight categories of comparability based on various soil properties (such as 'none', 'moderate' and 'extremely strong to conclusive') (Fitzpatrick and Raven, 2016), these categories of comparisons which mostly rely upon the analytical findings still require the need to highlight the hypotheses which were compared and thereby other limitations which are part of the complete interpretation. Similarly, although the need to use reference samples is generally well addressed, the extent to which these arguments are reflected in established and clear hypotheses is often lacking. Lastly, in addition to providing a belief in an hypothesis, some authors have presented cases where additional subsequent decisions were made about the source of a sample, although information is lacking about the decision-making method, including assumptions on the prior belief and the fact that there is no other evidence.

## CHAPTER VI

### DISCUSSION

*All models are wrong, but some are useful*  
(Box and Draper, 1987, p.424)

The key aim of this thesis was to establish the nature and significance of some misinterpretations within the forensic reconstructions. Crucially, this includes the extent to which the expression of situational beliefs resulted in a misrepresentation of reality which could have been minimised at the time. Where it can be argued that all representations of reality are models which may never encapsulate the full complexity of the real world situation, it does not mean that they cannot adequately and reliably serve the purpose of reconstructions. A significant part of this is to understand and get a clearer idea of why making and communicating valid interpretations within various scientific domains can be such a challenge. Four studies were undertaken to provide an answer to these questions, which include a broadly theoretical approach (Chapter 2), an approach from legal practice (Chapter 3) and an approach from the practice by forensic scientists (Chapter 4 and 5). The following questions were addressed in each chapter respectively:

- How does the theoretical framework in which the forensic sciences operate create drawbacks and opportunities in making and communicating valid interpretations?
- To what extent can the misinterpretation of evidence be studied from published appeal court judgments, and if sufficient, what are the reasons for misleading evidence posed in these cases?
- To what extent can normative frameworks be developed to evaluate the arguments and beliefs expressed by forensic scientists in current practice and what are the limitations of taking such approaches?
- How is geo-forensic data currently being interpreted by forensic scientists, and is this living up to current scientific, logical, and legal standards?

Four cross cutting and hierarchical themes have emerged from the findings of these studies, as presented in Figure 6.1. At the top, this includes the descriptive aspects of understanding the nature of misleading evidence (6.1), which is directly based and dependent upon the methods applied to obtaining these results (6.2). Recognising the

importance of avoiding such misinterpretations, the third and fourth themes include the more normative approaches of how justified reconstructions can be achieved (6.3) and the ways in which the domain can look forward to the future (6.4). Each theme is outlined and discussed from specific findings to general implications, including the observations and arguments following the results in this thesis, how this fits in with current theories, and why this is (or is not) important and has implications in practice.

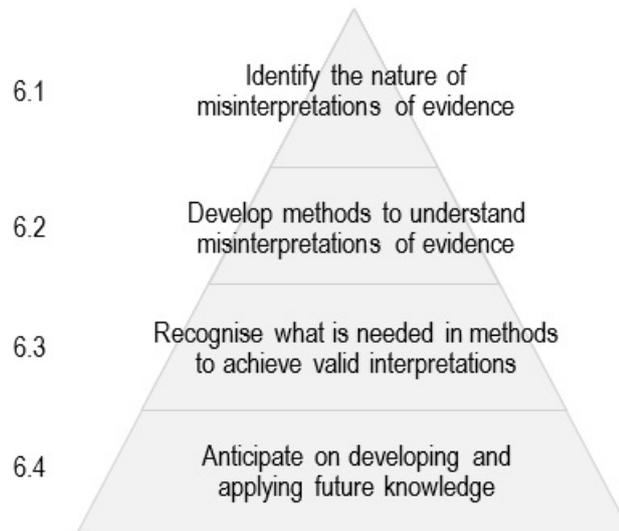


Figure 6.1: Schematic overview of the four themes which follow from the thesis

#### 6.1 THE MISINTERPRETATION OF FORENSIC EVIDENCE IS HAPPENING ON A CONSIDERABLE, UNACCEPTABLE, AND UNNECESSARY SCALE.

It follows from the analysis of legal documents and scientific observations (Chapters 3 and 5 respectively) that the reasons as to why forensic evidence has been misinterpreted are diverse. From a legal context, the most prevalent issues have in the present study been shown to be the direction of the judge (26%), the decisions by the jury (8%), or procedural errors (3%). However, the majority of issues found (36%) are related to the misleading nature of the forensic evidence itself as presented to the court. These instances include misinterpretations or misrepresentations of the probative value, relevance, or validity of evidence. The results from Chapter 5 provide more insights in the nature of the reasoning processes and arguments by forensic scientists, using the complex interpretations of geo-forensic evidence as an example. The results show that uncertainties caused by the nature of the analytical method, a lack of population knowledge on soil features, or from the ability to understand effects of evidence dynamics, are often acknowledged. Despite this knowledge, when it comes to the final steps of interpreting feature values in practice, the results show that the possible effect of these uncertainties have not always been considered when expressing the strength of the evi-

dence or belief in an hypothesis. More specifically, issues have been observed regarding the independence between analytical techniques, the application of statistical tools, and the presence of unsupported decisions of individualisations.

Moreover, it was found that many of the cases were overturned without the presence of any new evidence, suggesting that this could have been avoided in the trial. While this might sometimes be the case in theory, this does not mean it would have surfaced in practice. This could be more problematic in jurisdictions where, for example, there is a lack of clarity as to who should question the validity of the evidence (Vuille, 2013) and whether there is more than one expert involved (Roach, 2010). Nevertheless, the findings highlight for the first time from an empirical systematic study that there may well be more widespread issues than those found from post-conviction DNA-testing (Garrett and Neufeld, 2009).

The results from Chapters 3 and 5 therefore also provide some support for the theory that the presence of misinterpreted evidence is more prevalent than currently explicitly known. This is dependent upon the definition of the over-arching term of a 'misinterpretation'. Where Chapter 3 considers this issue from the perspective of whether or not a different presentation could have influenced the decision of the Trier-of-fact (Criminal Appeal Act 1995), Chapter 5 has argued for the need for valid interpretations that are independent from whether it is known that it led to unjustified legal outcomes. Observations from the court judgments (Chapter 3) highlight that Appeal Court judges acknowledge the presence of such misinterpretations (by an individual in any role), despite it not leading to a successful appeal. The importance for recognising this is the consideration of the context of the case - where in other circumstances - these could perhaps lead to unjustified outcomes. Nevertheless, it is possible to argue that the number of occurrences of misinterpreted evidence in cases, even on a relative note, is reason for concern given the potential impact in casework.

Moreover, the results presented in Chapter 3 where, in the majority of cases, there was no new evidence, would be very relevant to other jurisdictions where new information is one of the requirements to start an appeal (Broeders, 2018). Moreover, this has wider implications than just the conviction of innocent people. Namely, the dissemination of knowledge which have not undergone sufficient scientific scrutiny. This is especially important given, for example, recommendations to make publicly funded research Open Access (European Commission, 2018), and thereby increasing the need to be clear on findings given the often less scientifically trained audience. Nevertheless, forensic science requires adequate methods to identify these misinterpretations.

## 6.2 FORENSIC SCIENCE REQUIRES NORMATIVE FRAMEWORKS TO EVALUATE AND UNDERSTAND THE EXTENT AND NATURE OF MISINTERPRETATIONS

In addition to solely focussing on the outcomes of research, the results also provide some food for thought on the development of methods to evaluate misinterpretations. Where research in the past on misleading evidence in relation to wrongful convictions has mainly focussed on individual high-profile cases such as the IRA-bombings (Roach and Trotter, 2005) and the previously mentioned post-conviction DNA-exonerations, they would not have been sufficient to grasp the breadth of the results found in Chapter 3. This is not to say that the results from this thesis are sufficient, but an important consideration in developing the methods in Chapters 3 and 4 was its sustainability and comprehensiveness. In addition to performing targeted misinterpretation studies, population studies on misinterpretations are routinely needed.

In a similar manner to the need to develop normative benchmark in legal decision-making (Mitchell, 2010), benchmarks should also be developed for the scientific domain based upon scientific findings and valid arguments as argued in Chapter 2. The two frameworks to evaluate interpretations discussed in this thesis (Section 3.2 and Chapter 4) use different approaches to evaluate evidence interpretations. Table 6.1 highlights the differences in nature of these methods. Moreover, where the benchmarks within the legal domain are generally established (although being relatively broad), evaluating the scientific observations required the establishment of a clear and coherent set of key components. This has been achieved in the study in Chapter 4. Clarifying the difference between the two steps (establishing normative benchmarks and using those to evaluate findings from descriptive methods), allows a transparent understanding as to why these decisions were made. Moreover, the wide range of procedures highlighted in Chapter 4 to interpret geo-forensic observations suggests that these benchmarks are perhaps not well established enough within the forensic domain. Additionally, it offers an understanding of the consequences if certain normative benchmarks would change (for example if new knowledge became available).

One of the places where the legal and the scientific domains meet is that they both require the need for a great deal of human decision-making. Where this has been the topic of interest in the legal domain, such as in juror decision-making (Pennington and Hastie, 1992), it has only been more recently that the study of human reasoning has made its way into the mainstream forensic science research agenda (Dror, Charlton, and Péron, 2006). This is just one of the examples which form the complete breadth of what evaluative studies should encompass. More generally, the framework of normative approaches to studying evaluations requires a conjunction between understanding scientific and legal arguments. It is often the case, namely, that published appeal court judgments do not provide enough insight in the exact reasoning of an expert. Similarly, it cannot be

Table 6.1: Comparison between methods of evaluating interpretations within court judgments (Chapter 3) and published literature (Chapter 4 and 5)

	<b>Appeal court judgments</b>	<b>Published literature</b>
Nature of analysis	Top-down (descriptive)	Bottom-up (normative)
Benchmarks to identify possible issues	Ruling by Appeal Court judge on safeness conviction	Establish key components of robust interpretations (C4)
Identification of misinterpretations	Describe using a coding tree (C3)	Identify missing components from established list (C5)

assumed (as shown in Chapter 3) that valid scientific statements are appropriately used in court. The importance of evaluating scientific arguments has been recognised further after the expression of governmental concerns (National Research Council, 2009; The Law Commission, 2011), with a wide array of studies now continuously being applied to forensic practice. What should be remembered is that these evaluations, such as the observations of crime scene examiners (Eeden et al., 2018) or the integration of blind testing programmes in mainstream examinations (Kerkhoff et al., 2015) require transparent and pre-established normative benchmarks.

Where such frameworks do not only provide information on where interpretations are lacking (i.e. not living up to the normative benchmarks), it can also provide thoughts on what ‘good’ interpretations are. It can therefore, in turn, be used to update these and open up avenues for new research. Ideally, however, such evaluations would not be necessary and it is important to consider the requirements of what would be needed to achieve this.

### 6.3 FORENSIC INTERPRETATIONS SHOULD FOLLOW METHODS FROM OUTSIDE OF THE DOMAIN

The general aim of conducting any investigation is to express a belief in an hypothesis. In other words, it aims at collecting information in order to minimise the uncertainties surrounding initial beliefs in, ideally, two competing hypotheses (Lindley, 1977). Sections 6.1 and 6.2 have outlined that this process can result in misleading interpretations on a considerable scale. Moreover, the results from this thesis more generally have highlighted that these are either caused by an invalid set of premises (Chapter 2) or by questionable arguments from such premises (Chapters 3 and 5). Understanding what is needed, then, for avoiding these issues in practice and making justifiable arguments from the beginning, requires an understanding as to where these issues originate from. For example, a deliberate ignorance towards known uncertainties can be caused by a felt pressure to provide meaningful answers (Aitken and Taroni, 2004; Cole, 2013) or

by deliberate data manipulations (The Guardian, 2017b) or officer misconduct (Collins and Jarvis, 2009). As these are not necessarily directly related to scientific practice, the major challenge lies with those cases where there is a genuine lack of recognising what contains a valid interpretation.

In order to achieve valid interpretations from the outset it is crucial that research-logic- and reasoning approaches are applied from outside the domain. This may be considered to be an obvious statement given the belief that many forensic domains are based upon the application of established scientific domains to support legal questions. It is therefore important that fundamental philosophical and methodological principles from, for example, the classic geo-forensic field, are adapted to fit forensic practice (Morgan and Bull, 2007a). Moreover, Chapter 2 has discussed the idea that many forensic fields have developed from a pragmatic need, where there has been too little opportunity to develop methods to systematically reach valid interpretations and scrutinise foundational principles. What is meant with approaches outside of the domain, however, mostly relates to the use of reasoning processes and statistical approaches. This specifically follows from the results in Chapter 5 where, at least in published literature, many of the misinterpretations of data are simply embedded within current practice where the methodologies applied to make sense of raw data is often insufficient, unclear, and/or less than ideally applied. Moreover, it must be recognised that the forensic sciences are not the only field that faces these issues and that requires the decision-making under difficult and uncertain circumstances (West2002). What might be different to fields such as clinical decision-making, however, is the involvement of a wide array of individuals, each with their own understandings and goals, highlighting that communication remains key (Morgan, 2017b).

The ways in which valid interpretations can then be achieved is greatly dependent upon the ability to express the current knowledge base. Chapter 2 discusses the idea that despite being the foundation of most forensic arguments, both the experience of an expert as well as the results from empirical studies should be made more explicit and more structured respectively. A crucial - as well as perhaps the most challenging - part of this is the ability to incorporate the levels of uncertainty in structured arguments. In these cases, these will be mostly dependent upon the specific knowledge available to an expert and the details of experimental settings which form the premises of the conclusions. This is also important as, with a growing knowledge base and computational power, it perhaps takes away some grounds for miscommunication when the results of analytical studies are combined with research fields which are emerging within the forensic domain, including psychology, computer science, and applied mathematics and statistics.

The consequences of not making valid interpretations based on scientifically sound normative approaches is not only that it may lead to significant legal consequences, but it also overshadows the achievements and progress made in the past decades. Especially in cases where interpretations become increasingly complex, the method of communicating and understanding the value of the evidence becomes just as important as the sciences behind it (Howes, 2015). This perhaps requires a critical and innovative mindset for the future, which will be discussed in the next and final section.

#### 6.4 INNOVATIVE MINDSETS DRIVING THE FUTURE OF FORENSIC INTERPRETATIONS

Finally, the need and ability to keep an open mind with regards to the current knowledge base is critical for robust and effective forensic reconstructions. More specifically, this means that forensic scientists need to be aware that the available knowledge at a moment of analysis, interpretation, or decision, is situational and that it can change over time. Moreover, varying degrees of certainty can be obtained depending on the level of objectivity of the methods used and the characteristics of the attributes that are investigated. Additionally, these processes are limited by the need for context-specific knowledge within an investigative setting and the time pressure (Cole, 2013), as compared to the relative open-ended nature of research and practice in an academic setting.

Where this thesis has shown that many changes within forensic research and practice have been, logically, driven by emerging issues and extensive concerns on the validity of scientific arguments. Where there was widespread support for a 'research culture' in the forensic sciences (Mnookin et al., 2011), a change has taken place within the last decade where scientific scrutiny and quality standards are high on the agenda (Tully, 2015). In addition to recognising the importance of these issues by governing bodies, practising forensic scientists and legal professionals should be kept trained and up to date with changes within the forensic landscape and those fields concerning themselves with supporting valid interpretations (Government Chief Scientific Adviser, 2015).

One of the key arguments made in this thesis is that better use should be made from the current knowledge base. While understanding that it is important for individuals to grasp the methodologies to understand, combine, undertake, and interpret scientific research, the development of the forensic science discipline as a whole requires a development of a complete and coherent knowledge-base which form the basis for future research and applications. Moreover, in order to make the interpretation process more reproducible between scientists and/or legal professionals, the available information needs to be shared and combined in such a way that all individuals have 'access' to the 'most complete set' of information. Moreover, being critical towards this knowledge base, both in terms of the validity of premises (highlighting those perhaps based on disapproved methodologies or others that more generally lack a scientific basis) and the

validity of subsequent arguments (such as making unjustified decisions), using normative frameworks, allows for understanding where issues can possibly occur and where research should be focussed. In a general sense, it can be argued that the presentation and publication of findings and theories might never fully reflect what was aimed for (Woldinga, 1938). However, this does not mean that these results, given their limitations, cannot provide significant insights for future reconstructions.

## CHAPTER VII

### CONCLUSION

“You did [...] what you knew how to do,  
and when you knew better, you did better”

- *Maya Angelou*

Ever since forensic scientists have sought to reconstruct criminal events, forensic evidence has been acknowledged as both an extremely useful tool as well as a target of critique. To address these concerns, this thesis has shown the ability to evaluate current practice by combining normative and descriptive studies, both from the perspective of legal arguments in the shape of appeal court transcripts, as well as the arguments from forensic scientists as encapsulated in published empirical literature. Although there is an increasingly rich evidence base regarding, for example, population studies, analytical techniques, and trace evidence dynamics, the uncertainties that accompany the arguments which conceptualise such knowledge are often either not known or diminished in the expression of beliefs or in subsequent decision-making processes. This thesis has sought to understand the extent to which these interpretation processes could have been better, by addressing four research questions:

#### 7.1 HOW DOES THE THEORETICAL FRAMEWORK IN WHICH FORENSIC SCIENCE OPERATE CREATES DRAWBACKS AND OPPORTUNITIES IN MAKING AND COMMUNICATING VALID INTERPRETATIONS?

The findings from Chapter 2 of this thesis, which has explored the framework in which complex interpretations from forensic evidence are made from various perspectives, have generally shown that various aspects play a role in drawing justified inferences from observations. From a legal point of view, it has been shown that legal regulations such as evidence admissibility standards have not always resulted in argumentatively sound restrictions on evidence. Moreover, depending on the legal system, evidence is not always adequately questioned once admitted into court. Given the complexity of forensic interpretations, however, it cannot be expected from legal professionals to adequately address current critiques within the sciences and vice versa. Therefore, from a scientific perspective, interpretations should be communicated in a more understandable manner.

The results from exploring interpretations from a historical perspective have highlighted that, because of pragmatic needs and a lack of scrutinizing foundational principles, it is not surprising that misinterpretations have occurred throughout the past decades. However, it has been shown that critiques following these issues have changed forensic practice and its philosophical underpinnings in rigorous ways, allowing for a continuous improvement of interpretations. This does not suggest that current misinterpretations are acceptable, whereby the key issue has been identified as understanding and anticipating on (unknown) uncertainties, employing evaluative methods both preventively and repressively.

Understanding that forming sound argumentation is key in interpreting forensic evidence, the findings from this thesis suggest that a misunderstanding of reasoning processes often results in an over-statement of the value of findings. More specifically, it has been the introduction of probabilistic arguments which have caused concern from both the legal as well as the scientific community, despite this approach being the “most accurate” way of representing results. However, in acknowledging these issues, methods have been introduced within forensic interpretation, such as Bayesian networks, which can adequately support these processes if cultural and historical stances can be overcome.

A critical part of the framework of evidence interpretation is the evidence base which has been developed and the knowledge base which follows from combining these findings. The viewpoint that resulted from this chapter was that one of the major drawbacks of making informed interpretations is that there is a current inability to coherently combine the findings from unstructured published literature and tacit expert experience. By being more aware of this issue and by thinking of ways to develop a structured evidence base representing the full width of information from empirical findings, the knowledge base will be more scientifically sound and situational in nature.

## 7.2 TO WHAT EXTENT CAN THE MISINTERPRETATION OF EVIDENCE BE STUDIED FROM APPEAL CASE JUDGMENTS, AND IF SUFFICIENT, WHAT ARE THE REASONS FOR MISLEADING EVIDENCE POSED IN THESE CASES?

The findings presented in Chapter 3 have provided empirical data showing the nature of misleading evidence within criminal courts in England and Wales. Results from the systematic content analyses employed shows that a wealth of information can be obtained from case transcripts, despite them often not containing the full statements by forensic experts. The reasons for evidence to have been classified as misleading are mainly related to the nature of the evidence as presented in court, with the validity of witnesses, probative value of forensic evidence, and relevance of character evidence being the most prevalent combinations of issues which have been identified. The results show that mis-

interpretations of forensic evidence happen more often than only in a handful of cases, fuelling the discussion on justified rulings in the light of balancing between convicting innocent and acquitting guilty individuals in cases where misinterpreted forensic evidence might be present. This is especially the case as a critical question which was addressed in this study was whether these issues could have been minimised. The results show that the majority of overturned convictions did not contain explicit new evidence in the appeal, suggesting that many of these misleading aspects could have been prevented by providing more transparency in the relationship between evidence and hypotheses.

### 7.3 TO WHAT EXTENT CAN NORMATIVE FRAMEWORKS BE DEVELOPED TO EVALUATE THE ARGUMENTS AND BELIEFS EXPRESSED BY FORENSIC SCIENTISTS IN CURRENT PRACTICE AND WHAT ARE THE LIMITATIONS OF TAKING SUCH APPROACHES?

A major gap which follows from Chapter 2 is the lack of evaluating past and current interpretations. The findings from Chapter 4 have shown that it is possible to develop such evaluative normative frameworks, but that it requires a much more rigorous scrutiny of current arguments, beliefs, and foundational theories than has been currently articulated. The results also introduce the idea of trace-scripts which should form the basis of understanding physical effects on evidence and philosophical ideas of understanding the value of feature observations within an investigative context. Crucially, the resulting framework presented in this chapter and accompanying transparent methodology highlights the opportunities in developing thought-provoking frameworks based upon an analysis of previous theories and beliefs. The presented (idea of developing a) trace script, together with the identified key components of interpretations, allow for a systematic understanding of previous findings and an improved development of future interpretations.

### 7.4 HOW IS GEO-FORENSIC DATA CURRENTLY BEING INTERPRETED BY FORENSIC SCIENTISTS, AND IS THIS LIVING UP TO CURRENT SCIENTIFIC, LOGICAL, AND LEGAL STANDARDS?

Chapter 5 presented empirical findings derived from the evaluation of current interpretation approaches for geo-forensic evidence, which have demonstrated a range of issues. The findings have highlighted that geo-forensic experts are not always being transparent on the clustering and classification methods, despite this being necessary to make interpretations comparable, reproducible, and justifiable. Additionally, the results have shown that the independence between features (or properties more generally) has either been simply assumed or has not been explicitly discussed, despite that this would have been possible given the data. Moreover, Chapter 5 has provided a great example of the

need for knowledge-based systems, whereby the evidence base from case-based inferences and empirical studies should effectively and efficiently be combined, whereby the need has been shown for geo-forensic experts to take responsibility in providing these results in such manners. Lastly, the results of this study demonstrated that some individual experts have not only presented beliefs in hypotheses, but also made subsequent decisions on the source of a sample whilst lacking information on the decision-making method, assumptions, and prior beliefs.

#### 7.5 IMPLICATIONS OF THIS THESIS

This thesis has combined a variety of approaches to study, and thereby understand, the nature of why beliefs in hypotheses based upon forensic evidence have been the target of much critique. These approaches are novel in that they analyse publicly available data not used for this purpose previously, as well as that new normative approaches are developed in the form of trace scripts and key components of interpretations. Where the insights in misinterpretations of evidence were previously gained through generally high-profile isolated cases or found within individual laboratories, the nature of the approaches developed in this thesis have allowed for providing systematic and comprehensive insights into misinterpretations, both from a legal as well as a scientific perspective. Moreover, the findings from this thesis enable new research to be undertaken. For example, the misleading nature of evidence identified in Chapter 3 allows for applying advanced methodologies to these findings, such as machine learning, to automatically analyse cases on an even more significant scale. Moreover, the developed normative framework in Chapter 4, and the idea of developing trace scripts, can be readily applied to systematically studying a range of other forensic evidence types in a coherent and transparent way. Where this thesis has identified worrying and significant issues in the interpretation of a range of forensic evidence types, it has developed normative frameworks and presented philosophical and practical ideas which help pave the way to improve interpretations of evidence in the future, contributing to making fair and reasonable decisions in court.

## PUBLICATIONS

Smit, N.M., Morgan, R.M., Lagnado, D.A. (2018). *A systematic analysis of misleading evidence in unsafe rulings in England and Wales*. *Science & Justice*, 2(58).

FOS (2017). *Twelve Guiding Principles and Recommendations for Dealing with Quantitative Evidence in Criminal Law* Tech. rep. Cambridge: Isaac Newton Institute for Mathematical Sciences.

Smit, N.M., Morgan, R.M., Lagnado, D.A. (2018). *Five perspectives on making justifiable arguments from forensic evidence (review paper)*. *Forensic Science International*. In preparation.

Smit, N.M., Morgan, R.M., Lagnado, D.A. (2018). *Evaluating forensic interpretations (Part I): Towards a normative approach of evaluating trace interpretations*. *Science & Justice*. In preparation.

Smit, N.M., Morgan, R.M., Lagnado, D.A. (2018). *Evaluating forensic interpretations (Part II): Evaluating current interpretations of soil evidence*. *Science & Justice*. In preparation.

Earwaker, H., Nakhaeizadeh, S., Smit, N.M., Morgan, R.M. (2018). *A cultural change to enable improved decision-making in the forensic sciences: a six phased approach*. *Science and Justice*. Under review.



## BIBLIOGRAPHY

- Aitken, C.G.G. (2009). "Some thoughts on the role of probabilistic reasoning in the evaluation of evidence." In: *Criminal and Environmental Soil Forensics*. Dordrecht: Springer Netherlands, pp. 33–47.
- Aitken, C.G.G. and A.J. Gammerman (1989). "Probabilistic reasoning in evidential assessment." In: *Journal of the Forensic Science Society* 29.5, pp. 303–316.
- Aitken, C.G.G. and D. Lucy (2004). "Evaluation of trace evidence in the form of multivariate data." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 53.1, pp. 109–122.
- Aitken, C.G.G. and D.A. Stoney (1991). *The use of statistics in forensic science*.
- Aitken, C.G.G. and F. Taroni (2004). *Statistics and the evaluation of evidence for forensic scientists*.
- Aitken, C.G.G., F. Taroni, and P. Garbolino (2003). "A graphical model for the evaluation of cross-transfer evidence in DNA profiles." In: *Theoretical Population Biology*.
- Aitken, C.G.G. et al. (2011). *Expressing evaluative opinions: A position statement*.
- Akulova, V, D Vasiliauskiene, and D Talaliene (2002). "Further insights into the persistence of transferred fibres on outdoor clothes." In: *Science & Justice* 42.3, pp. 165–71.
- Association of Forensic Science Providers (2009). "Standards for the formulation of evaluative forensic science expert opinion." In: *Science & Justice* 49.3, pp. 161–164.
- Asten, A.C. van (2014). "On the added value of forensic science and grand innovation challenges for the forensic community." In: *Science & Justice* 54.2, pp. 170–9.
- Baum, L. (2008). *Judges and their Audiences: A Perspective on Judicial Behavior*. Princeton University Press.
- Bedau, H.A. and M. Radelet (1987). "Miscarriages of justice in potentially capital cases." In: *Stanford Law Review* 40.1, pp. 21–179.
- Bell, E. (2010). "Judicial Assessment of Expert Evidence." In: *Judicial Studies Institute Journal* 55.
- Bertillon, A. (1893). *Identification anthropométrique: Instructions signalétiques*.
- Biedermann, A., S. Bozza, and F. Taroni (2011). "Probabilistic evidential assessment of gunshot residue particle evidence (Part II): Bayesian parameter estimation for experimental count data." In: *Forensic Science International* 206.1-3, pp. 103–10.
- Biedermann, A., F. Taroni, and C.G.G. Aitken (2014). "Liberties and constraints of the normative approach to evaluation and decision in forensic science: a discussion towards overcoming some common misconceptions." In: *Law, Probability and Risk* 13.2, pp. 181–191.
- Biedermann, A., F. Taroni, and P. Garbolino (2007). "Equal prior probabilities: can one do any better?" In: *Forensic Science International* 172.2-3, pp. 85–93.
- Biedermann, A., F. Taroni, O. Delémont, C. Semadeni, and A.C. Davison (2005). "The evaluation of evidence in the forensic investigation of fire incidents. Part II. Practical examples of the use of Bayesian networks." In: *Forensic Science International* 147.1, pp. 59–69.
- Bitzer, S., N. Albertini, E. Lock, O. Ribaux, and O. Delémont (2015). "Utility of the clue — From assessing the investigative contribution of forensic science to supporting the decision to use traces." In: *Science & Justice* 55.6, pp. 509–513.
- Blaikie, N.W.H. (2010). *Designing social research: the logic of anticipation*. Polity Press Ltd, p. 298.

- Bland, J.M. and D.G. Altman (1998). "Statistics notes: Bayesians and frequentists." In: *BMJ* 317:7166, pp. 1151–1160.
- Blaxland, Henry (2017). "Sappers and Underminers: Fresh Evidence Revisited." In: *Criminal Law Review* 7, pp. 537–542.
- Bolck, A., H. Ni, and M. Lopatka (2014). "Evaluating score-and feature-based likelihood ratio models for multivariate continuous data: Applied to forensic MDMA comparison." In: *Law, Probability and Risk* 14.3, pp. 243–266.
- Bonetti, J. and L. Quarino (2014). "Comparative Forensic Soil Analysis of New Jersey State Parks Using a Combination of Simple Techniques with Multivariate Statistics." In: *Journal of Forensic Sciences* 59.3, pp. 627–636.
- Bono, J.P. (2011). "Commentary on the Need for a Research Culture in the Forensic Sciences." In: *UCLA L. Rev.* 58.3.
- Borchardt, E.M. and R.E. Lutz (1932). *Convicting the Innocent: Sixty-five Actual Errors of Criminal Justice*. Garden City, New York: Garden City Publishing Company.
- Box, G.E.P. and R.N. Draper (1987). *Empirical model-building and response surfaces*. Wiley, p. 669.
- Brandon, R. and C. Davies (1973). *Wrongful Imprisonment: Mistaken Convictions and Their Consequences*.
- Brants, C. (2010). "The vulnerability of Dutch criminal procedure to wrongful conviction." In: *Wrongful Conviction: International Perspectives on Miscarriages of Justice*. Ed. by C.R. Huff and M. Killias. Temple University Press, pp. 157–182.
- (2011). "Wrongful Convictions and Inquisitorial Process: The Case of the Netherlands." In: *University of Cincinnati Law Review* 80, pp. 1069–1114.
- Broeders, A.P.A. (2018). "Misleidend bewijs." In: *Expertise & Recht* 1.1.
- Bromwich, M.R. (2007). *Summary of Recommendations of the Independent Investigator for the HPD Crime Lab and Property Room*. Tech. rep.
- Bryant, V.M. and G.D. Jones (2006). "Forensic palynology: Current status of a rarely used technique in the United States of America." In: *Forensic Science International* 163.3, pp. 183–197.
- Bryant, V.M. and D.C. Mildenhall (1998). "Forensic palynology: a new way to catch crooks." In: *Contributions Series-American Association of Stratigraphic Palynologists* 33, pp. 145–155.
- Buchanan, M. (2007). "Statistics: Conviction by numbers." en. In: *Nature* 445.7125, pp. 254–255.
- Budin, G. (2005). "Ontology-driven translation management." In: *Knowledge Systems and Translation*. 7.
- Bull, P.A. and R.M. Morgan (2006). "Sediment fingerprints: a forensic technique using quartz sand grains." In: *Science & Justice* 46.2, pp. 107–124.
- Bull, P.A., A. Parker, and R.M. Morgan (2006). "The forensic analysis of soils and sediment taken from the cast of a footprint." In: *Forensic Science International* 162.1-3, pp. 6–12.
- Bull, P.A., R.M. Morgan, A. Sagovsky, and G.J.A. Hughes (2006). "The transfer and persistence of trace particulates: experimental studies using clothing fabrics." In: *Science & Justice* 46.3, pp. 185–95.
- Buntine, W. (1996). "A guide to the literature on learning probabilistic networks from data." In: *IEEE Transactions on Knowledge and Data Engineering* 8.2.
- Butler, J.M., M.C. Kline, and M.D. Coble (2018). "NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): Variation observed and lessons learned." In: *Forensic Science International: Genetics* 37, pp. 81–94.

- Buzzini, P. et al. (2005). "Survey of crowbar and household paints in burglary cases-population studies, transfer and interpretation." In: *Forensic Science International* 152.2-3, pp. 221-34.
- Cantrell, S, C Roux, P Maynard, and J Robertson (2001). "A textile fibre survey as an aid to the interpretation of fibre evidence in the Sydney region." In: *Forensic Science International* 123.1, pp. 48-53.
- Carvalho, Á. et al. (2013). "Organic matter characterization of sediments in two river beaches from northern Portugal for forensic application." In: *Forensic Science International* 233.1-3, pp. 403-415.
- Champod, C., I.W. Evett, and B. Kuchler (2001). "Earmarks as evidence: a critical review." In: *Journal of forensic sciences* 46.6, pp. 1275-1284.
- Champod, C. and F. Taroni (1997). "Bayesian framework for the evaluation of fibre transfer evidence." In: *Science & Justice* 37.2, pp. 75-83.
- Chan, K., G. Tan, and R.C.S. Wong (2012). "Looking at forensic intelligence from the metaphysical perspective: citing illicit heroin profiling as an example." en. In: *Australian Journal of Forensic Sciences* 44.3, pp. 227-242.
- Chazottes, V., C. Brocard, and B. Peyrot (2004). "Particle size analysis of soils under simulated scene of crime conditions: The interest of multivariate analyses." In: *Forensic Science International* 140.2-3, pp. 159-166.
- Cheng, E.K. and A.H. Yoon (2005). "Does Frye or Daubert matter? A study of scientific admissibility standards." In: *Virginia Law Review* 91.2, pp. 471 -513.
- Chisum, J.W. and B.E. Turvey (2000). "Evidence Dynamics: Locard's Exchange Principle & Crime Reconstruction." In: *Journal of Behavioral Profiling* 1.1.
- (2011). *Crime Reconstruction*. Academic Press, p. 704.
- Christensen, A.M., C.M. Crowder, S.D. Ousley, and M.M. Houck (2014). "Error and its meaning in forensic science." In: *Journal of forensic sciences* 59.1, pp. 123-6.
- Cole, S.A. (2001). *Suspect identities: a history of fingerprinting and criminal identification*. Harvard University Press, p. 369.
- (2005). "More than zero: Accounting for error in latent fingerprint identification." In: *The Journal of Criminal Law and Criminology* 95.3, pp. 985-1423.
- (2009). "Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification." In: *Law, Probability and Risk*.
- (2012). "Forensic Science and Wrongful Convictions: From Exposer to Contributor to Corrector." In: *New England Law Review* 46.4.
- (2013). "Forensic culture as epistemic culture: the sociology of forensic science." In: *Studies in history and philosophy of biological and biomedical sciences* 44.1, pp. 36-46.
- Cole, S.A. and R. Dioso-Villa (2009). "Investigating the 'CSI Effect' Effect: Media and Litigation Crisis in Criminal Law." In: *Stanford Law Review* 62.2, p. 125. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Collins, J.M. and J. Jarvis (2009). "The Wrongful Conviction of Forensic Science." In: *Forensic Science Policy & Management: An International Journal* 1.1, pp. 17-31.
- Coltheart, M., P. Menzies, and J. Sutton (2010). "Abductive inference and delusional belief." In: *Cognitive Neuropsychiatry* 15.1-3, pp. 261-287.
- Concheri, G. et al. (2011). "Chemical Elemental Distribution and Soil DNA Fingerprints Provide the Critical Evidence in Murder Case Investigation." In: *PLoS ONE* 6.6.

- Connors, E., T. Lundregan, N. Miller, and T. McEwen (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial*. Tech. rep. National Institute of Justice, p. 85.
- Cook, R., I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert (1998a). "A hierarchy of propositions: deciding which level to address in casework." In: *Science & Justice* 38.4, pp. 231–239.
- (1998b). "A model for case assessment and interpretation." In: *Science & Justice* 38.3, pp. 151–6.
- Cooper, J., E.A. Bennett, and H.L. Sukel (1996). *Complex Scientific Testimony: How Do Jurors Make Decisions?* Tech. rep. 4.
- Coulson, S.A., J.S. Buckleton, A.B. Gummer, and C.M. Triggs (2001). "Glass on clothing and shoes of members of the general population and people suspected of breaking crimes." In: *Science & Justice* 41.1, pp. 39–48.
- Council of the European Union (2005). *Prüm Convention*. Tech. rep.
- Crispino, F. and M.M. Houck (2016). "Principles of Forensic Science." In: *Materials analysis in forensic science*. Ed. by Max M. Houck. Academic Press, p. 530.
- Crispino, F., O. Ribaux, M.M. Houck, and P. Margot (2011). "Forensic science – A true science?" In: *Australian Journal of Forensic Sciences* 43.2-3, pp. 157–176.
- Croft, D.J. and K. Pye (2004a). "Colour theory and the evaluation of an instrumental method of measurement using geological samples for forensic applications." In: *Geological Society, London, Special Publications* 232.1, pp. 49–62.
- (2004b). "Multi-technique comparison of source and primary transfer soil samples: an experimental investigation." In: *Science & Justice* 44.1, pp. 21–8.
- Curran, J.M. (2013). "The Frequentist Approach to Forensic Evidence Interpretation." In: *Encyclopedia of Forensic Sciences*, pp. 286–291.
- Curran, J.M., C.M. Triggs, J.S. Buckleton, K.A.J. Walsh, and T. Hicks (1998). "Assessing transfer probabilities in a Bayesian interpretation of forensic glass evidence." In: *Science & Justice* 38.1, pp. 15–21.
- Dargue, Paul (2016). "The Impartiality of the England and Wales Court of Appeal (Criminal Division): A Quantitative Analysis." PhD Thesis. University of Northumbria.
- Daubert v. Merrell Dow Pharmaceuticals (1993). 509 U.S. 579.
- Davies, M. (2015). "Knowledge (Explicit, Implicit and Tacit): Philosophical Aspects." In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pp. 74–90.
- Dawid, A.P. and I.W. Evett (1997). "Using a Graphical Method to Assist the Evaluation of Complicated Patterns of Evidence." In: *Journal of Forensic Sciences* 42.2, pp. 226–231.
- Dawid, A.P., J. Mortera, and P. Vicard (2007). "Object-oriented Bayesian networks for complex forensic DNA profiling problems." In: *Forensic Science International* 169.2-3, pp. 195–205.
- Dawid, A.P., W. Twining, and M. Vasilaki (2011). "Evidence, inference and enquiry." In: *British Academy*.
- Dawson, L.A. and S. Hillier (2010). "Measurement of soil characteristics for forensic applications." In: *Surface and Interface Analysis* March, pp. 363–377.
- De Wael, K., C. Baes, L. Lepot, and F. Gason (2011). "On the frequency of occurrence of a peculiar polyester fibre type found in blue denim textiles." In: *Science & Justice* 51.4, pp. 154–62.
- Dioso-Villa, R. (2015). "A Repository of Wrongful Convictions in Australia: First Steps Towards Estimating Prevalence and Causal Contributing Factors." In: *Flinders Law Journal* 17, pp. 163–202.

- Dragović, S. and A. Onjia (2006). "Classification of soil samples according to their geographic origin using gamma-ray spectrometry and principal component analysis." In: *Journal of Environmental Radioactivity* 89.2, pp. 150–158.
- Drisko, J. and T. Maschi (2015). *Content Analysis*. Oxford University Press.
- Dror, I.E., D. Charlton, and A.E. Péron (2006). "Contextual information renders experts vulnerable to making erroneous identifications." In: *Forensic Science International* 156.1, pp. 74–8.
- Dror, I.E. and G. Hampikian (2011). "Subjectivity and bias in forensic DNA mixture interpretation." In: *Science & Justice* 51.4, pp. 204–8.
- Dror, I.E. and G. Langenburg (2018). "'Cannot Decide': The Fine Line Between Appropriate Inconclusive Determinations Versus Unjustifiably Deciding Not To Decide." In: *Journal of Forensic Sciences*.
- Dror, I.E. and R. Rosenthal (2008). "Meta-analytically quantifying the reliability and biasability of forensic experts." In: *Journal of forensic sciences* 53.4, pp. 900–3.
- Dror, I.E. et al. (2011). "Cognitive issues in fingerprint analysis: inter- and intra-expert consistency and the effect of a 'target' comparison." In: *Forensic Science International* 208.1-3, pp. 10–7.
- Dror, I.E., R.M. Morgan, C. Rando, and S. Nakhaeizadeh (2017). "Letter to the Editor - The Bias Snowball and the Bias Cascade Effects: Two Distinct Biases that May Impact Forensic Decision Making." In: *Journal of Forensic Sciences* 62.3, pp. 832–833.
- Druzdzel, MJ and LC Van Der Gaag (1995). "Elicitation of probabilities for belief networks: combining qualitative and quantitative information." In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*.
- Edmond, G., J.M. Tangen, R.A. Searston, and I.E. Dror (2015). "Contextual bias and cross-contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals." In: *Law, Probability and Risk* 14.1, pp. 1–25.
- Edwards, H.T. and M.A. Livermore (2009). "Pitfalls of Empirical Studies that Attempt to Understand the Factors Affecting Appellate Decision Making." In: *Duke Law Journal* 58.1895.
- Edwards, W. (1991). "Influence Diagrams, Bayesian Imperialism, and the Collins Case: An Appeal to Reason." In: *Cardozo Law Review* 13.
- Eeden, C.A.J. van den, J. Ost, C.J. de Poot, and P.J. van Koppen (2018). "Solving the puzzle: The effects of contextual information and feedback on the interpretation of a crime scene." In: *Journal of Investigative Psychology and Offender Profiling* 15.2, pp. 109–123.
- Eighth Circuit Judicial Committee (2017). *Manual of Model Criminal Jury Instructions for the District Courts of the Eighth Circuit*. Tech. rep.
- European Commission (2018). *Recommendation on access to and preservation of scientific information*. Tech. rep.
- Evans, Michael, Wayne McIntosh, Jimmy Lin, and Cynthia Cates (2007). "Recounting the Courts - Applying Automated Content Analysis to Enhance Empirical Legal Research." In: *Journal of Empirical Legal Studies* 4.1007.
- Evett, I.W. (1984). "A quantitative theory for interpreting transfer evidence in criminal cases." In: *Applied Statistics* 33.1, pp. 25–32.
- (1986). "A Bayesian Approach to the Problem of Interpreting Glass Evidence in Forensic Science Casework." In: *Journal of the Forensic Science Society* 26.1, pp. 3–18.
- (1987). "On meaningful questions: a two-trace transfer problem." In: *Journal of the Forensic Science Society* 27.6, pp. 375–381.

- Evett, I.W. and J.S. Buckleton (1990). "The interpretation of glass evidence. A practical approach." In: *Journal of the Forensic Science Society* 30.4, pp. 215–223.
- Evett, I.W., P.E. Cage, and C.G.G. Aitken (1987). "Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases." In: *Applied statistics*.
- Evett, I.W., G. Jackson, and J.A. Lambert (2000). "More on the hierarchy of propositions: exploring the distinction between explanations and propositions." In: *Science & Justice* 40.1, pp. 3–10.
- Evett, I.W. and E.J. Spiehler (1987). *Rule Induction in Forensic Science*. Tech. rep.
- Evett, I.W., G. Jackson, J.A. Lambert, and S. McCrossan (2000). "The impact of the principles of evidence interpretation on the structure and content of statements." In: *Science & Justice*.
- FBI (2015). *FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review*.
- FOS (2017). *Twelve Guiding Principles and Recommendations for Dealing with Quantitative Evidence in Criminal Law*. Tech. rep. Cambridge: Isaac Newton Institute for Mathematical Sciences.
- Fayyad, U, G Piatetsky-Shapiro, and P Smyth (1996). "From Data Mining to Knowledge Discovery in Databases." In: *AI Magazine* 17.3, pp. 37–54.
- Fenton, N.E. (2011). "Science and law: Improve statistics in court." In: *Nature* 479.7371, pp. 36–37.
- Fenton, N.E. and M. Neil (2000). "The Jury Observation Fallacy and the use of Bayesian Networks to present Probabilistic Legal Arguments." In: *Mathematics Today (Bulletin of the IMA)* 36.6, pp. 180–187.
- Fenton, N.E., D. Berger, D.A. Lagnado, M. Neil, and A. Hsu (2014). "When 'neutral' evidence still has probative value (with implications from the Barry George Case)." In: *Science & Justice* 54.4, pp. 274–287.
- Feynman, R.P. (1955). "The value of science." In: *Engineering and Science* 19.3, pp. 13–15.
- Field, S (2009). "Fair trials and procedural tradition in Europe." In: *Oxford Journal of Legal Studies* 29.2, pp. 365–387.
- Fienberg, S.E. and M.J. Schervish (1986). "The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decisionmaking." In: *BUL Rev.*
- Findley, K.A. (2011). "Adversarial Inquisitions: Rethinking the Search for the Truth." In: *SSRN Electronic Journal*.
- Finetti, B. (1974). *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons Australia, Limited, p. 300.
- Fitzpatrick, R.W. and M.D. Raven (2012). "How Pedology and Mineralogy Helped Solve a Double Murder Case: Using Forensics to Inspire Future Generations of Soil Scientists." In: *Soil Horizons* 53.5, p. 14.
- (2016). *Guidelines for conducting criminal and environmental soil forensic investigations*. Tech. rep. Centre for Australian Forensic Soil Science, p. 39.
- Fitzpatrick, R.W., M.D. Raven, and S.T. Forrester (2009). "A systematic approach to soil forensics: Criminal case studies involving transference from crime scene to forensic evidence." In: *Criminal and Environmental Soil Forensics*. Dordrecht: Springer Netherlands, pp. 105–127.
- Fojtášek, L., J. Vacinová, P. Kolář, and M. Kotrlý (2003). "Distribution of GSR particles in the surroundings of shooting pistol." In: *Forensic Science International* 132.2, pp. 99–105.
- Found, B. and J. Ganas (2013). "The management of domain irrelevant context information in forensic handwriting examination casework." In: *Science & Justice* 53.2, pp. 154–158.
- Frank, J. and B. Frank (1957). *Not Guilty*. Garden City, New York: Doubleday, p. 261.

- Fricker, R.L. (1993). "Pathologist's Plea Adds to Turmoil: Discovery of Possibly Hundreds of Faked Autopsies Helps Defense Challenges." In: *ABA Journal* 79, p. 24.
- Frye v. United States (1923). 293 F. 1013.
- Garbolino, P. (2013). "The Scientification of Forensic Practice." In: *New Challenges to Philosophy of Science*. Springer Science & Business Media, pp. 287–297.
- Garbolino, P. and F. Taroni (2002). "Evaluation of scientific evidence using Bayesian networks." In: *Forensic Science International* 125.2-3, pp. 149–155.
- Garrett, B.L. (2008). "Judging innocence." In: *Columbia Law Review* 108.1, pp. 55–142.
- (2011). *Convicting the innocent: Where criminal prosecutions go wrong*.
- (2015). "Convicting the Innocent Redux." In: *Wrongful Convictions and the DNA Revolution*. Ed. by Daniel S. Medwed. Cambridge: Cambridge University Press, pp. 40–56.
- Garrett, B.L. and G. Mitchell (2016). "Forensics and Fallibility: Comparing the Views of Lawyers and Judges." In: *West Virginia Law Review* 119.621-710, pp. 2016–63.
- Garrett, B.L. and P.J. Neufeld (2009). "Invalid forensic science testimony and wrongful convictions." In: *Virginia Law Review* 95.1, pp. 1–97.
- Geller, B., J. Almog, P. Margot, and E.A. Springer (1999). "A Chronological Review of Fingerprint Forgery." In: *Journal of Forensic Sciences* 44.5, pp. 963–968.
- Gherghel, S., R.M. Morgan, C.S. Blackman, K. Karu, and I.P. Parkin (2016). "Analysis of transferred fragrance and its forensic implications." In: *Science & Justice* 56.6, pp. 413–420.
- Gilliéron, G (2013). "Wrongful Convictions in Switzerland: A Problem of Summary Proceedings." In: *University of Cincinnati Law Review*.
- (2014). "Wrongful Convictions in the Inquisitorial System." In: *Encyclopedia of Criminology and Criminal Justice*. Springer New York, pp. 5576–5584.
- Giurazza, F. et al. (2012). "Determination of stature from skeletal and skull measurements by CT scan evaluation." In: *Forensic Science International* 222.1-3, 398.e1–9.
- Goddard, CH (1926). "Scientific Identification of Firearms and Bullets." In: *Journal of the American Institute of Criminal Law and Criminology* 17.2.
- Gould, J.B., J. Carrano, R.A. Leo, and J.K. Young (2013). "Predicting erroneous convictions: A social science approach to miscarriages of justice." In: *Univ. of San Francisco Law Research Paper* 2013-20.
- Government Chief Scientific Adviser (2015). *Forensic Science and Beyond: Authenticity Provenance and Assurance*. Tech. rep. London: The Government Office for Science, p. 14.
- Gradusova, O. and E. Nesterina (2009). "The current status of forensic soil examination in the Russian Federation." In: *Criminal and Environmental Soil Forensics*, pp. 61–73.
- Greene, E. et al. (2002). "Jurors and juries: a review of the field." In: *Taking psychology and law into the twenty-first century*. Ed. by James R. P. Ogloff. New York: Kluwer Academic, pp. 225–284.
- Grieve, M.C. and T.W. Biermann (1997). "Wool fibres – transfer to vinyl and leather vehicle seats and some observations on their secondary transfer." In: *Science & Justice* 37.1, pp. 31–38.
- Grieve, M.C. and J. Dunlop (1992). "A practical aspect of the Bayesian interpretation of fibre evidence." In: *Journal of the Forensic Science Society* 32.2, pp. 169–175.
- Griffin, Lissa (2015). "Forensic Evidence and the Court of Appeal for England and Wales." In: *Brit. J. Am. Legal Stud.* 4, pp. 619–647.
- Groscup, J.L. and S.D. Penrod (2002). "The effects of Daubert on the admissibility of expert testimony in state and federal criminal cases." In: *Psychology, Public Policy, and Law* 8.4.

- Gross, Hans (1906). *Criminal Investigation: A Practical Handbook for Magistrates, Police Officers, and Lawyers*. Madras: A. Krishnamachari.
- Gross, S.R. (2008). "Convicting the innocent." In: *Annual Review of Law and Social Science* 4.1, pp. 173–192.
- Gross, S.R. and B. O'Brien (2008). "Frequency and Predictors of False Conviction: Why We Know So Little, and New Data on Capital Cases." In: *Journal of Empirical Legal Studies* 5.4, pp. 927–962.
- Gross, S.R. and M. Shaffer (2012). "Exonerations in the United States, 1989 – 2012." In: *U of Michigan Public Law Working Paper* 277, p. 103.
- Gross, S.R., K. Jacoby, D.J. Matheson, N. Montgomery, and J. Daniel (2005). "Exonerations in the United States 1989 through 2003." In: *The Journal of Criminal Law & Criminology* 95.2, pp. 523–560.
- Guedes, A., H. Ribeiro, B. Valentim, and F. Noronha (2009). "Quantitative colour analysis of beach and dune sediments for forensic applications: A Portuguese example." In: *Forensic Science International* 190.1-3, pp. 42–51.
- Guedes, A. et al. (2011). "Characterization of soils from the Algarve region (Portugal): A multi-disciplinary approach for forensic applications." In: *Science & Justice* 51.2, pp. 77–82.
- Guedes, A. et al. (2013). "The potential application of magnetic susceptibility as a technique for soil forensic examinations." In: *Geological Society, London, Special Publications* 384.1, pp. 65–73.
- Hall, Mark A. and Ronald F. Wright (2008). *Systematic content analysis of judicial opinions*.
- Hamer, D. (2012). "Discussion paper: The R v T controversy: forensic evidence, law and logic." In: *Law, Probability and Risk* 11.4, pp. 331–345.
- Harmon, T. R. (2005). "Too Late for Luck: A Comparison of Post-Furman Exonerations and Executions of the Innocent." In: *Crime & Delinquency* 51.4, pp. 498–520.
- Hastie, Reid (1993). *Inside the Juror: The Psychology of Juror Decision Making*. New York: Cambridge University Press.
- Hawksworth, D.L. and P.E.J. Wiltshire (2011). *Forensic mycology: The use of fungi in criminal investigations*.
- Heaton, S.J. (2013). "A Critical Evaluation of Using Innocence as a Criterion in the Post-Conviction Process." DPhil Thesis. University of East Anglia.
- Henneberg, M. (2017). "Worlds Apart: Cold Case Reviews and Investigations into Alleged Wrongful Convictions in England and Wales." In: *Journal of Cold Case Review* 3.1.
- Hepler, A.B., A.P. Dawid, and V. Leucari (2007). "Object-oriented graphical representations of complex patterns of evidence." In: *Law, Probability and Risk*.
- Hicks, T., R. Vanina, and P. Margot (1996). "Transfer and persistence of glass fragments on garments." In: *Science & Justice* 36.2, pp. 101–107.
- Hong, Sungwook, Aleum Han, Sojung Kim, Dasom Son, and Heewon Min (2014). "Transfer of fibres on the hands of living subjects and their persistence during hand washing." In: *Science and Justice* 54.6, pp. 451–458.
- Hoon, Christina (2013). "Meta-Synthesis of Qualitative Case Studies." In: *Organizational Research Methods* 16.4, pp. 522–556.
- Horowitz, I.A. (1997). "Reasonable Doubt Instructions: Commonsense Justice and Standard of Proof." In: *Psychology, Public Policy, and Law* 3.2 - 3, pp. 285–302.

- Houck, M.M. (2003). *Trace Evidence Analysis: More Cases in Forensic Microscopy and Mute Witnesses*. Elsevier, p. 300.
- House of Commons (2011). *Science and Technology Committee - Seventh Report*. Tech. rep.
- House of Lords (2018). *Forensic science: call for evidence*. Tech. rep. London.
- Howes, L.M. (2015). "The communication of forensic science in the criminal justice system: A review of theory and proposed directions for research." In: *Science & Justice* 55.2, pp. 145–154.
- Inman, K. and N. Rudin (2000). *Principles and Practice of Criminalistics: The Profession of Forensic Science*. CRC Press, p. 392.
- (2002). "The origin of evidence." In: *Forensic Science International* 126.1, pp. 11–16.
- International Association for Identification (2007). *IAI Position Concerning Latent Fingerprint Identification*. Tech. rep. Mendota Heights, MN.
- Ivkovic, S., J. Yearwood, and A. Stranieri (2001). "Discovering Interesting Association Rules from Legal Databases." In: *Information and Communication Technology Law* 11.1, pp. 35–47.
- Jackson, G., S. Jones, G. Booth, C. Champod, and I.W. Evett (2006). "The nature of forensic science opinion—a possible framework to guide thinking and practice in investigation and in court proceedings." In: *Science & Justice* 46.1, pp. 33–44.
- Jacobe, H.T., J.J. Leitenberger, and P.R. Bergstresser (2007). "Understanding clinical trial outcomes: design, analysis, and interpretation." In: *Dermatologic therapy* 20.2, pp. 77–85.
- Jacobson, J. (2013). *By mistakes we learn? — A review of criminal appeals against sentence*. Tech. rep. Justice Reform.
- Jaynes, E.T. (2003). *Probability Theory: the Logic of Science*.
- Jeffreys, A.J., V. Wilson, and S.L. Thein (1985). "Individual-specific 'fingerprints' of human DNA." In: *Nature* 316.6023, pp. 76–79.
- Jiménez-Buedo, María and Luis M. Miller (2010). "Why a trade-off? The relationship between the external and internal validity of experiments." In: *Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia* 25.3, pp. 301–321.
- Judicial College (2017). *Crown Court Compendium*. Tech. rep.
- (2018). *The Crown Court Compendium Part I*. Tech. rep., p. 447.
- Judicial Commission of N.S.W. (2011). *Conviction appeals in New South Wales*. Tech. rep., p. 318.
- Kahneman, Daniel (2003). "A Perspective on Judgment and Choice Mapping Bounded Rationality." In: *American psychologist* 58.9, p. 697.
- Kassin, S.M. (2012). "Why confessions trump innocence." In: *The American psychologist* 67.6, pp. 431–45.
- Kassin, S.M., I.E. Dror, and J. Kukucka (2013). "The forensic confirmation bias: Problems, perspectives, and proposed solutions." In: *Journal of Applied Research in Memory and Cognition* 2.1, pp. 42–52.
- Kaye, D.H. and J.J. Koehler (1991). "Can Jurors understand probabilistic evidence?" In: *J.R.Statist. Soc. A* 154.1, pp. 75–81.
- Kehl, Danielle, Priscilla Guo, and Samuel Kessler (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*. Tech. rep.
- Keller, Helen and Alec. Stone Sweet (2008). *A Europe of rights: the impact of the ECHR on national legal systems*. Oxford University Press, p. 852.
- Keppens, J. (2014). "On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning." In: *Artificial Intelligence and Law* 22.3, pp. 239–290.

- Kerkhoff, W. et al. (2015). "Design and results of an exploratory double blind testing program in firearms examination." In: *Science & Justice* 55.6, pp. 514–519.
- Kerrigan, Kevin (2006). "Miscarriage of justice in the magistrates' court: The forgotten power of the Criminal Cases Review Commission." In: *Criminal Law Review* 2, pp. 124–139.
- Kerstholt, José H. and Janet L. Jackson (1998). "Judicial decision making: order of evidence presentation and availability of background information." In: *Applied Cognitive Psychology* 12.5, pp. 445–454.
- Kher, A., M Mulholland, E. Green, and B. Reedy (2006). "Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis." In: *Vibrational Spectroscopy* 40.2, pp. 270–277.
- Killias, M. (2008). "Wrongful Convictions in Switzerland: The Experience of a Continental Law Country." In: *Wrongful Conviction: International Perspectives on Miscarriages of Justice*. Temple University Press, pp. 139–156.
- Kirk, PL (1953). *Crime Investigation; Physical Evidence and the Police Laboratory*.
- (1963). "The Ontogeny of Criminalistics." In: *Journal of Criminal Law and Criminology* 54.2, pp. 235–238.
- Kiureghian, Armen Der and Ove Ditlevsen (2009). "Aleatory or epistemic? Does it matter?" In: *Structural Safety* 31.2, pp. 105–112.
- Kloosterman, A., M.J. Sjerps, and A. Quak (2014). "Error rates in forensic DNA analysis: Definition, numbers, impact and communication." In: *Forensic Science International: Genetics* 12, pp. 77–85.
- Koehler, J.J. and M.J. Saks (2009). "Individualization claims in forensic science: Still unwarranted." In: *Brook. L. Rev.*
- Koehler, J.J., N.J. Schweitzer, M.J. Saks, and D. McQuiston (2016). "Science, technology, or the expert witness: What influences jurors' judgments about forensic science testimony?" In: *Psychology, Public Policy, and Law* 22.4, pp. 401–413.
- Koppen, P. van and S.D. Penrod (2012). *Adversarial versus Inquisitorial Justice: Psychological Perspectives on Criminal Justice Systems*. Springer Science & Business Media, p. 437.
- Krippendorff, K. (2007). *Content analysis: An introduction to its methodology*. SAGE, p. 413.
- Kulesza, Cezary (2011). *Criminal plea bargains in the English and the Polish administration of justice systems in the context of the fair trial guarantees*. Wydawnictwo Temida 2, p. 243.
- Kumho Tire Co. v. Carmichael (1999). 526 U.S. 137.
- Lagnado, D.A., N.E. Fenton, and M. Neil (2013). "Legal idioms: a framework for evidential reasoning." In: *Argument & Computation* 4.1, pp. 46–63.
- Landis, JR and GG Koch (1977). "The measurement of observer agreement for categorical data." In: *biometrics*.
- Landsman, Stephan and Richard F. Rakos (1994). "A preliminary inquiry into the effect of potentially biasing information on judges and jurors in civil litigation." In: *Behavioral Sciences & the Law* 12.2, pp. 113–126.
- Langdon, J. and P.R. Wilson (2005). "When justice fails: A follow-up examination of serious criminal cases since 1985." In: *Current Issues Crim. Just.* 17.2, pp. 179–202.
- Leigh, L.H. (2006). "Lurking doubt and the safety of convictions." In: *Criminal Law Review* 809, pp. 809–816.

- Lepot, L., T. Vanden Driessche, K. Lunstroot, F. Gason, and K. De Wael (2015). "Fibre persistence on immersed garment-Influence of knitted recipient fabrics." In: *Science & Justice* 55.4, pp. 248–253.
- Leverick, F and J Chalmers (2014). "Causes of wrongful conviction." In: *Post-Corroborator Safeguards Review Report of the Academic Expert Group*. Edinburgh: The Scottish Government, pp. 30–43.
- Lieberman, J.D. and J. Arndt (2000). "Understanding the limits of limiting instructions: Social psychological explanations for the failures of instructions to disregard pretrial publicity and other inadmissible evidence." In: *Psychology, Public Policy, and Law* 6.3, pp. 677–711.
- Ligertwood, A. and G. Edmond (2012). "Expressing evaluative forensic science opinions in a court of law." In: *Law, Probability and Risk* 11.4, pp. 289–302.
- Lindley, D.V. (1977). "A Problem in Forensic Science." In: *Biometrika* 64.2, p. 207.
- (2014). *Understanding Uncertainty*. 2nd. John Wiley & Sons, p. 424.
- Locard, E (1930). "The analysis of dust traces. Part I." In: *American Journal of Police Science* 1.3, pp. 276–298.
- (1934). *La Police et les Méthodes Scientifiques*. Tech. rep. Paris: Impr. des Presses universitaires de France.
- Lombardi, G. (1999). "The contribution of forensic geology and other trace evidence analysis to the investigation of the killing of Italian Prime Minister Aldo Moro." In: *Journal of forensic sciences* 44.3, pp. 634–42.
- (2009). "The Death of Countess Agusta in Portofino (Northern Italy) and the Soil from Two Mismatched Slippers." In: *Journal of Forensic Sciences* 54.2, pp. 395–399.
- Lorr, M and DM McNair (1966). "Methods relating to evaluation of therapeutic outcome." In: *Methods of research in psychotherapy*.
- Magnani, Lorenzo (2001). *Abduction, reason, and science: Processes of discovery and explanation*.
- Maitre, M., P. Kirkbride, M. Horder, C.P. Roux, and A. Beavis (2017). "Current perspectives in the interpretation of gunshot residues in forensic science: A review." In: *Forensic Science International* 270, pp. 1–11.
- Martin, Peter D., Hermann Schmitter, and Peter M. Schneider (2001). "A brief history of the formation of DNA databases in forensic science within Europe." In: *Forensic Science International* 119.2, pp. 225–231.
- Maurer, Hans H (2005). "Multi-analyte procedures for screening for and quantification of drugs in blood, plasma, or serum by liquid chromatography-single stage or tandem mass spectrometry (LC-MS or LC-MS/MS) relevant to clinical and forensic toxicology." In: *Clinical biochemistry* 38.4, pp. 310–8.
- Maveety, Nancy (2003). *The Pioneers of Judicial Behavior*. University of Michigan Press.
- May, LS (1930). "The Identification of Knives, Tools and Instruments A Positive Science." In: *American Journal of Police Science* 1.3, pp. 246–259.
- Mayes, R.W., L.M. Macdonald, J.M. Ross, and L.A. Dawson (2009). "Discrimination of Domestic Garden Soils Using Plant Wax Compounds as Markers." In: *Criminal and Environmental Soil Forensics*. Dordrecht: Springer Netherlands, pp. 463–476.
- McQuiston-Surrett, D. and M.J. Saks (2009). "The testimony of forensic identification science: What expert witnesses say and what factfinders hear." In: *Law and Human Behavior* 33.5, pp. 436–453. arXiv: [arXiv:1002.2562v1](https://arxiv.org/abs/1002.2562v1).

- Meakin, G. and A. Jamieson (2013). "DNA transfer: review and implications for casework." In: *Forensic science international. Genetics* 7.4, pp. 434–43.
- Meakin, G.E., E.V. Butcher, R.A.H. van Oorschot, and R.M. Morgan (2015). "The deposition and persistence of indirectly-transferred DNA on regularly-used knives." In: *Forensic Science International: Genetics Supplement Series* 5, e498–e500.
- Meehl, PE (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of The Evidence*.
- Meester, R., M. Collins, R. Gill, and M. van Lambalgen (2007). "On the (ab)use of statistics in the legal case against the nurse Lucia de B." In: *Law, Probability and Risk* 5.3-4, pp. 233–250.
- Meester, Ronald and M.J. Sjerps (2004). *Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence*. Tech. rep., pp. 51–62.
- Melo, V.F., L.C. Barbar, P.G.P. Zamora, C.E. Schaefer, and G.A. Cordeiro (2008). "Chemical, physical and mineralogical characterization of soils from the Curitiba Metropolitan Region for forensic purpose." In: *Forensic Science International* 179.2-3, pp. 123–134.
- Mikkonen, S., V. Suominen, and P. Heinonen (1996). "Use of footwear impressions in crime scene investigations assisted by computerised footwear collection system." In: *Forensic Science International* 82.1, pp. 67–79.
- Mildenhall, D.C., P.E.J. Wiltshire, and V.M. Bryant (2006). "Forensic palynology: why do it and how it works." In: *Forensic science international* 163.3, pp. 163–72.
- Mitchell, Gregory (2010). "Evaluating Judges." In: *The Psychology of Judicial Decision Making*. Oxford University Press, pp. 221–244.
- Mnookin, J.L., S.A. Cole, I.E. Dror, and B.A.J. Fisher (2011). "The Need for a Research Culture in the Forensic Sciences." In: *UCLA L. Rev.* 58, pp. 725–780.
- Morgan, R.M. (2017a). "Conceptualising forensic science and forensic reconstruction. Part I: A conceptual model." In: *Science & Justice* 57.6, pp. 455–459.
- (2017b). "Conceptualising forensic science and forensic reconstruction. Part II: The critical interaction between research, policy/law and practice." In: *Science & Justice* 57.6, pp. 460–467.
- Morgan, R.M. and P.A. Bull (2007a). "Forensic geoscience and crime detection." In: *Minerva Medicolegale* 127, pp. 73–89.
- (2007b). "The philosophy, nature and practice of forensic sediment analysis." In: *Progress in Physical Geography* 31.1, pp. 43–58.
- (2007c). "The use of grain size distribution analysis of sediments and soils in forensic enquiry." In: *Science & Justice* 47.3, pp. 125–35.
- Morgan, R.M., P.E.J. Wiltshire, A. Parker, and P.A. Bull (2006). "The role of forensic geoscience in wildlife crime detection." In: *Forensic Science International* 162.1-3, pp. 152–62.
- Morgan, R.M. et al. (2008). "The preservation of quartz grain surface textures following vehicle fire and their use in forensic enquiry." In: *Science & Justice* 48.3, pp. 133–40.
- Morgan, R.M. et al. (2009). "The relevance of the evolution of experimental studies for the interpretation and evaluation of some trace physical evidence." In: *Science & Justice* 49.4, pp. 277–85.
- Morgan, R.M., J. Robertson, C. Lennard, K. Hubbard, and P.A. Bull (2010). "Quartz grain surface textures of soils and sediments from Canberra, Australia: A forensic reconstruction tool." In: *Australian Journal of Forensic Sciences* 42.3, pp. 169–179.

- Morgan, R.M., E. Allen, T. King, and P.A. Bull (2014). "The spatial and temporal distribution of pollen in a room: Forensic implications." In: *Science & Justice* 54.1, pp. 49–56.
- Mullen, Carrie, Danielle Spence, Linda Moxey, and Allan Jamieson (2014). "Perception problems of the verbal scale." In: *Science & Justice* 54.2, pp. 154–8.
- Murphy, Peter (2007). *Murphy on Evidence*. Oxford University Press, UK, p. 707.
- Murray, K.R., R.W. Fitzpatrick, R.S. Bottrill, R. Berry, and H. Kobus (2016). "Soil transference patterns on bras: Image processing and laboratory dragging experiments." In: *Forensic Science International* 258, pp. 88–100.
- NDAA (2016). *Response to 'Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*.
- NFI (2012). *DNA Match in Vaatstra Case*.
- Nakhaeizadeh, S., I.E. Dror, and R.M. Morgan (2014). "Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias." In: *Science & Justice* 54.3, pp. 208–14.
- Nakhaeizadeh, S., I. Hanson, and N. Dozzi (2014). "The power of contextual effects in forensic anthropology: a study of biasability in the visual interpretations of trauma analysis on skeletal remains." In: *Journal of forensic sciences* 59.5, pp. 1177–83.
- National Research Council (2004). *Forensic Analysis: Weighing Bullet Lead Evidence*. Washington, DC: The National Academies Press.
- (2009). *Strengthening Forensic Science in the United States: a Path Forward*. Tech. rep. Washington D.C.: The National Academies.
- Office of the Inspector General (2006). *A review of the FBI's handling of the Brandon Mayfield Case*. Tech. rep., p. 21.
- Osborn, AS (1946). *Questioned Documents*. Boyd, p. 1076.
- Osborne, N.K.P., M.C. Taylor, and R. Zajac (2016). "Exploring the role of contextual information in bloodstain pattern analysis: A qualitative approach." In: *Forensic Science International* 260, pp. 1–8.
- PCAST (2016). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Tech. rep. Executive Office of the President President's Council of Advisors on Science and Technology, p. 174.
- Page, M., J. Taylor, and M. Blenkin (2012). "Context effects and observer bias—implications for forensic odontology." In: *Journal of forensic sciences* 57.1, pp. 108–12.
- Palmer, R. and H.J. Burch (2009). "The population, transfer and persistence of fibres on the skin of living subjects." In: *Science & Justice* 49.4, pp. 259–264.
- Palmer, R. et al. (2015). "The prevalence of two 'commonly' encountered synthetic target fibres within a large urban environment." In: *Science & Justice* 55.2, pp. 103–106.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. 2nd. Vol. 88. San Francisco, CA: Morgan Kaufmann Publishers Inc., p. 552.
- Pennington, Nancy and Reid Hastie (1992). "Explaining the Evidence: Tests of the Story Model for Juror Decision Making." In: *Journal of Personality and Social Psychology* 62.2, pp. 189–206.
- Perkal, M., Y.L. Ng, and J.R. Pearson (1994). "Impurity profiling of methylamphetamine in Australia and the development of a national drugs database." In: *Forensic Science International* 69.1, pp. 77–87.
- Petraco, N.D.K.K., T.A. Kubic, and N.D.K.K. Petraco (2008). "Case studies in forensic soil examinations." In: *Forensic Science International* 178.2-3, e23–e27.

- Plous, Scott (1993). *The psychology of judgment and decision making*, p. 302.
- Podlas, Kimberlianne (2005). "The CSI Effect: Exposing the Media Myth." In: *Fordham Intellectual Property, Media & Entertainment Law Journal* 16.2, pp. 429–465. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- Polanyi, Michael (1966). *The Tacit Dimension*. Garden City, New York: Doubleday & Company, Inc., p. 128.
- Pounds, C.A. and K.W. Smalldon (1975a). "The Transfer of Fibres Between Clothing Materials During Simulated Contacts and Their Persistence During Wear. Part I—Fibre Transference." In: *Journal - Forensic Science Society* 15.1, pp. 17–27.
- (1975b). "The Transfer of Fibres Between Clothing Materials During Simulated Contacts and Their Persistence During Wear. Part II—Fibre Persistence." In: *Journal - Forensic Science Society* 15.1, pp. 29–37.
- Prakken, H. (2014). "On direct and indirect probabilistic reasoning in legal proof." In: *Law, Probability and Risk* 13.3-4, pp. 327–337.
- Pretty, I.A. and D. Sweet (2001). "The scientific basis for human bitemark analyses—a critical review." In: *Science & Justice* 41.2, pp. 85–92.
- Pye, K. and D.J. Croft (2004). "Forensic geoscience: introduction and overview." In: *Geological Society, London, Special Publications* 232.1, pp. 1–5.
- Pye, K., S.J. Blott, D.J. Croft, and J.F. Carter (2006). "Forensic comparison of soil samples: Assessment of small-scale spatial variability in elemental composition, carbon and nitrogen isotope ratios, colour, and particle size distribution." In: *Forensic Science International* 163.1-2, pp. 59–80.
- Quaak, F.C.A. and I. Kuiper (2011). "Statistical data analysis of bacterial t-RFLP profiles in forensic soil comparisons." In: *Forensic science international* 210.1-3, pp. 96–101.
- Quinn, K.M. (2012). "The Academic Study of Decision Making on Multimember Courts." In: *California Law Review* 100.6, pp. 1493–1501.
- R v T (2010). *EWCA Crim* 2439.
- Radelet, M.L., H.A. Bedau, and C.E. Putnam (1992). *In spite of innocence: Erroneous convictions in capital cases*. Northeastern University Press, p. 400.
- Radin, Edward D. (1964). *The innocents*. New York: Morrow, p. 256.
- Rattner, Arye (1988). "Convicted but innocent - Wrongful conviction and the criminal justice system." In: *Law and Human Behavior* 12.3, pp. 283–293.
- Raymond, J.J. et al. (2008). "Assessing trace DNA evidence from a residential burglary: Abundance, transfer and persistence." In: *Forensic Science International: Genetics Supplement Series* 1.1, pp. 442–443.
- Redmayne, M. (2002). "The Relevance of Bad Character." In: *The Cambridge Law Journal* 61.3, pp. 684–714.
- Roach, Kent (2010). "Wrongful convictions: adversarial and inquisitorial themes." In: *North Carolina journal of international law and commercial regulation* 35.
- Roach, Kent and Gary Trotter (2005). "Miscarriages of Justice in the War Against Terrorism." In: *Penn State Law Review* 109.4.
- Robbers, Monica L P (2008). "Blinded by Science." In: *Criminal Justice Policy Review* 19.1, pp. 84–102.
- Roberts, P. (2013). "Renegotiating forensic cultures: between law, science and criminal justice." In: *Studies in history and philosophy of biological and biomedical sciences* 44.1, pp. 47–59.

- Roberts, Stephanie (2017). "Fresh Evidence and Factual Innocence in the Criminal Division of the Court of Appeal." In: *Journal of Criminal Law* 81.4, pp. 1–30.
- Robertson, J. (2009). "Soils ain't soils: Context and issues facing soil scientists in a forensic world." In: *Criminal and Environmental Soil Forensics*. Dordrecht: Springer Netherlands, pp. 3–12.
- Robertson, J. and C.P. Roux (2010). "Trace evidence: here today, gone tomorrow?" In: *Science & Justice* 50.1, pp. 18–22.
- Rose, Daniel (1994). *A Symbolic and Connectionist Approach To Legal Information Retrieval*. New York: Psychology Press, p. 336.
- Rossmo, D. Kim (2014). "Case rethinking: a protocol for reviewing criminal investigations." In: *Police Practice and Research* 4263.May 2015, pp. 1–17.
- Roux, C.P. and P. Margot (1997). "The population of textile fibres on car seats." In: *Science & Justice* 37.1, pp. 25–30.
- Ruffell, A. and L.A. Dawson (2009). "Forensic geology in environmental crime: Illegal waste movement & burial in Northern Ireland." In: *Environmental Forensics* 10.3, pp. 208–213.
- Ruffell, A. and A. Sandiford (2011). "Maximising trace soil evidence: an improved recovery method developed during investigation of a \$26 million bank robbery." In: *Forensic science international* 209.1-3, e1–7.
- Ruffell, A. and P.E.J. Wiltshire (2004). "Conjunctive use of quantitative and qualitative X-ray diffraction analysis of soils and rocks for forensic analysis." In: *Forensic Science International* 145.1, pp. 13–23.
- Saks, M.J. (2010). "Forensic identification: From a faith-based "Science" to a scientific science." In: *Forensic Science International* 201.1-3, pp. 14–7.
- Saks, M.J. and D.L. Faigman (2008). "Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It." en. In: *Annual Review of Law and Social Science* 4.1, pp. 149–171.
- Saks, M.J. and J.J. Koehler (2005). "The Coming Paradigm Shift in Forensic Identification Science." In: *Science* 309.5736, pp. 892–895.
- (2008). "The individualization fallacy in forensic science evidence." In: *Vanderbilt Law Review* 61, pp. 199–219.
- Schank, R.C. (1982). *Dynamic memory: a theory of reminding and learning in computers and people*. Cambridge University Press, p. 234.
- Schmechel, Richard S., Timothy P. O'Toole, Catharine Easterly, and Elizabeth F. Loftus (2006). "Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence." In: *Jurimetrics* 46, pp. 177–214.
- Schum, D.A. (1994). *The evidential foundations of probabilistic reasoning*.
- (1999). "Inference networks and the evaluation of evidence: alternative analyses." In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 575–584.
- (2005). *A Wigmorean interpretation of the evaluation of a complicated pattern of evidence*. Tech. rep.
- (2009). "A science of evidence: contributions from law and probability." In: *Law, Probability and Risk* 8.3, pp. 197–231.
- Schwartz, A. (2005). "A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification." In: *Science and Technology Law Review*, pp. 1–42.
- Schweitzer, N.J. and M.J. Saks (2007). "The CSI effect: popular fiction about forensic science affects the public's expectations about real forensic science." In: *Jurimetrics* 47.3, pp. 357–364.

- Scott, K.R., R.M. Morgan, V.J. Jones, and N.G. Cameron (2014). "The transferability of diatoms to clothing and the methods appropriate for their collection and analysis in forensic geoscience." In: *Forensic Science International* 241, pp. 127–137.
- Scott, M. and N. Smith (2010). "Legal Taxonomy From Sweet and Maxwell." In: *Legal Information Management* 10.03, pp. 217–222.
- Scottish Criminal Records Office (2006). *McKie Report*. Tech. rep.
- Segal, Jeffrey A. and Harold J. Spaeth (2002). *The Supreme Court and the Attitudinal Model Revisited*. Cambridge: Cambridge University Press.
- Sheck, B., P.J. Neufeld, and J. Dwyer (2000). *Actual innocence: five days to execution, and other dispatches from the wrongly convicted*. Doubleday, p. 245.
- Shortliffe, E.H. (1974). "A Rule-based Computer Program for Advising Physicians Regarding Antimicrobial Therapy Selection." In: *1974 annual ACM conference*. Vol. 2. June. New York, New York, USA: ACM Press, pp. 739–739.
- Smalldon, K.W. and A.C. Moffat (1973). "The Calculation of Discriminating Power for a Series of Correlated Attributes." In: *Journal of the Forensic Science Society* 13.4, pp. 291–295.
- Smit, N.M., R.M. Morgan, and D.A. Lagnado (2017). "A systematic analysis of misleading evidence in unsafe rulings in England and Wales." In: *Science & Justice* 58.2, pp. 128–137.
- Smit, N.M., D.A. Lagnado, R.M. Morgan, and N.E. Fenton (2016). "Using Bayesian networks to guide the assessment of new evidence in an appeal case." en. In: *Crime Science* 5.1, p. 9.
- Smith, L.L., R. Bull, and R. Holliday (2011). "Understanding juror perceptions of forensic evidence: investigating the impact of case context on perceptions of forensic evidence strength." In: *Journal of forensic sciences*.
- Spencer, J.R. (2007). "Quashing Convictions for Procedural Irregularities." In: *Criminal Law Review* Nov, pp. 835–848.
- (2016). *Evidence of Bad Character*. 3rd. Hart Publishing, p. 344.
- Sperling, C. and S.L. Cooper (2013). "When Science Changes, How Does Law Respond." In: *Encyclopedia of Forensic Sciences*. Elsevier, pp. 476–480.
- Spottswood, M. (2013). "Bridging the gap between Bayesian and story-comparison models of juridical inference." In: *Law, Probability and Risk* 13.1, pp. 47–64.
- (2015). "Ordering Proof: Beyond Adversarial and Inquisitorial Trial Structures." In: *SSRN Electronic Journal*.
- Stoney, D.A. (1994). "Relaxation of the assumption of relevance and an application to one-trace and two-trace problems." In: *Journal of the Forensic Science Society* 34.1, pp. 17–21.
- Stoney, D.A. and P.L. Stoney (2015). "Critical review of forensic trace evidence analysis and the need for a new approach." In: *Forensic Science International* 251, pp. 159–170.
- Stranieri, Andrew and John Zeleznikow (2005). *Knowledge Discovery from Legal Databases*. Ed. by Springer. Dordrecht.
- Streiner, D.L. and G.R. Norman (2006). "'Precision' and 'accuracy': two terms that are neither." In: *Journal of clinical epidemiology* 59.4, pp. 327–30.
- Suarez, M.D., R.J. Southard, and S.J. Parikh (2015). "Understanding Variations of Soil Mapping Units and Associated Data for Forensic Science." In: *Journal of Forensic Sciences* 60.4, pp. 894–905.
- Sung, Tz'u (1981). *The Washing Away of Wrongs*. Ed. by Brian McKnight. Ann Arbor, MI: University of Michigan Press.

- Swofford, H.J. and J.G. Cino (2018). "Lay Understanding of "Identification": How Jurors Interpret Forensic Identification Testimony." In: *Journal of Forensic Identification* 68.1.
- Taroni, F., A. Biedermann, and S. Bozza (2016). "Statistical hypothesis testing and common misinterpretations: Should we abandon p-value in forensic science applications?" In: *Forensic Science International* 259, e32–6.
- Taroni, F., A. Biedermann, P. Garbolino, and C.G.G. Aitken (2004). "A general approach to Bayesian networks for the interpretation of evidence." In: *Forensic Science International* 139.1, pp. 5–16.
- Taylor, D., A. Biedermann, T. Hicks, and C. Champod (2018). "A template for constructing Bayesian networks in forensic biology cases when considering activity level propositions." In: *Forensic Science International: Genetics* 33, pp. 136–146.
- The Conversation (2017). *DNA databases: it's still far from clear how effective they are in fighting crime.*
- The Guardian (2017a). *DNA in the dock: how flawed techniques send innocent people to prison.*
- (2017b). *Manchester lab's drug tests may have been manipulated.*
- (2018a). *Forensic science: the tip of the iceberg?*
- (2018b). *Met urgently reviews suspected forensic science blunders.*
- The Law Commission (2011). *Expert Evidence in Criminal Proceedings in England and Wales.* Tech. rep. London: HMSO.
- The Law Gazette (2018). *Degraded in a decade.*
- The Sacramento Bee (2018). *'Open-source' genealogy site provided missing DNA link to East Area Rapist, investigator says.*
- Thomas, C. (2008). "Exposing the myths of jury service." In: *Criminal Law Review* 6, pp. 415–430.
- (2010). *Are juries fair?* Tech. rep. London: Ministry of Justice Research Series.
- (2017). "Ethnicity and the fairness of jury trials in England and Wales 2006-2014." In: *Criminal Law Review* 11, pp. 860–876.
- Thompson, M.B., J.M. Tangen, and D.J. McCarthy (2014). "Human Matching Performance of Genuine Crime Scene Latent Fingerprints." In: *Law and human behavior* 38.1, pp. 84–93.
- Thompson, W.C. (1989). "Are juries competent to evaluate statistical evidence?" In: *Law and Contemporary Problems.*
- (2008). "The Potential for Error in Forensic DNA Testing (and How That Complicates the Use of DNA Databases for Criminal Identification)." In: *Forensic DNA Databases and Race: Issues, Abuses and Actions*, pp. 1–49.
- (2010). "Beyond bad apples: analyzing the role of forensic science in wrongful convictions." In: *Southwestern University Law Review* 1, pp. 971–994.
- (2012). "Discussion paper: Hard cases make bad law—reactions to R v T." In: *Law, Probability and Risk* 11.4, pp. 347–359.
- Thompson, W.C. and E.L. Schumann (1987). "Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy." In: *Law and Human Behavior* 11.3, pp. 167–187.
- Thompson, William C. (2018). "How Should Forensic Scientists Present Source Conclusions ?" In: *SETON HALL LAW REVIEW*, pp. 773–813.
- Tinsley, Yvette (2001). "Juror decision-making: a look inside the jury room." In: *The British Criminology Conference: Selected Proceedings* 4.
- Tully, Gillian (2015). *Forensic Science Regulator Annual Report.* Tech. rep., p. 32.

- Tully, Gillian (2018). "Forensic science in England & Wales, a commentary." In: *Forensic Science International*.
- Turvey, B.E. (2011). *Criminal Profiling: An Introduction to Behavioral Evidence Analysis*. Academic Press, p. 728.
- Tversky, A and Daniel Kahneman (1974). "Judgment Under Uncertainty: Heuristics and Biases." In: *Science* 185.4157.
- Twining, W. (1985). *Theories of Evidence: Bentham and Wigmore*. Stanford University Press, p. 265.
- (2003). "Evidence as a multi-disciplinary subject." In: *Law, Probability and Risk* 2.2, pp. 91–107.
- (2011). *Moving beyond law: interdisciplinarity and the study of evidence*. Tech. rep.
- UK Ministry of Justice (2013). *Government response to Law Commission report on expert evidence*. Tech. rep.
- (2018). *Criminal Procedure Rules*. Tech. rep., p. 831.
- Verheij, B. (2014). "To catch a thief with and without numbers: arguments, scenarios and probabilities in evidential reasoning." In: *Law, Probability and Risk* 13.3-4, pp. 307–325.
- Vooijs, C., P. Vergeer, and J. van der Weerd (2015). "Towards source level evaluation of the evidential value of fibre examinations." In: *Forensic science international* 250, pp. 57–67.
- Vuille, Joëlle (2013). "Admissibility and appraisal of scientific evidence in continental European criminal justice systems: past, present and future." en. In: *Australian Journal of Forensic Sciences* 45.4, pp. 389–397.
- Wakefield, J.C., A.M. Skene, A.F.M. Smith, and I.W. Evett (1991). "The evaluation of fibre transfer evidence in forensic science: a case study in statistical modelling." In: *Applied statistics*.
- Walton, Douglas (2010). *Legal Argumentation and Evidence*. Vol. 1. Penn State Press.
- Wigmore, J.H. (1937). *The Science of Judicial Proof, as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials*. Little, Brown, p. 1065.
- Wilkins, D. and K. Pillaipakkamnatt (1997). "The Effectiveness of Machine Learning Techniques for Predicting Time to Case Disposition." In: *Proceedings of Sixth International Conference on Artificial Intelligence and Law*, pp. 39–46.
- Wilks, B., R.M. Morgan, and N.L. Rose (2017). "An experimental study addressing the use of geoforensic analysis for the exploitation of improvised explosive devices (IEDs)." In: *Forensic Science International* 278, pp. 52–67.
- Wilson, A., C.G.G. Aitken, R. Sleeman, and J.F. Carter (2014). "The evaluation of evidence relating to traces of cocaine on banknotes." In: *Forensic Science International* 236, pp. 67–76.
- Wilson, M.R., J.A. DiZinno, D. Polanskey, J. Replogle, and B. Budowle (1995). "Validation of mitochondrial DNA sequencing for forensic casework analysis." In: *International Journal of Legal Medicine* 108.2, pp. 68–74.
- Wilson, P.R. (1989). "When justice fails: a preliminary examination of serious criminal cases in Australia." In: *The Australian Journal of Social Issues* 24.1.
- Wiltshire, Patricia E.J. and Sue Black (2006). "The cribriform approach to the retrieval of palynological evidence from the turbinates of murder victims." In: *Forensic science international* 163.3, pp. 224–30.
- Winter, R.J. and E. Greene (2007). "Juror decision making." In: *Handbook of applied cognition*.
- Wistrich, A.J., C. Guthrie, and J.J. Rachlinski (2005). "Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding." In: *U Penn L Rev* 4.1251.
- Woldinga, G.J. (1938). *Xenophons Symposium*. J. Schipper.

- Woods, B., C. Lennard, P. Kirkbride, and J. Robertson (2014a). "Soil examination for a forensic trace evidence laboratory—Part 1: Spectroscopic techniques." In: *Forensic Science International* 245, pp. 187–194.
- Woods, B., P. Kirkbride, C. Lennard, and J. Robertson (2014b). "Soil examination for a forensic trace evidence laboratory—Part 2: Elemental analysis." In: *Forensic Science International* 245, pp. 195–201.
- Woods, B., C. Lennard, P. Kirkbride, and J. Robertson (2016). "Soil examination for a forensic trace evidence laboratory—Part 3: A proposed protocol for the effective triage and management of soil examinations." In: *Forensic Science International* 262, pp. 46–55.
- Zadora, G., A. Martyna, D. Ramos, and C.G.G. Aitken (2013). *Statistical Analysis in Forensic Science: Evidential Values of Multivariate Physicochemical Data*.
- Zalman, M. (2011). "An Integrated Justice Model of Wrongful Convictions." In: *Alb. L. Rev.* September 2010, pp. 1465–1525.