Helping with inquiries: theory and practice in forensic science

Lawless, Christopher James

Helping With Inquiries: Theory and Practice in Forensic Science

Christopher James Lawless

Declaration

The contents of this thesis are produced solely for the qualification of Doctor of Philosophy at Durham University and consist of the author's original individual contributions with appropriate recognition of any references being indicated throughout.

Statement of Copyright

ACKNOWLEDGEMENTS

It seems like only yesterday that I first arrived in the Sociology Department at Durham to begin this PhD. Whilst the fact that three years has passed by so quickly can be partially attributable to the succession of deadlines which are part and parcel of a PhD, it is, more importantly, due to the fact that the completion of this research has also proven to be such a fruitful and absorbing experience. I must, above all, thank Professor Robin Williams for his continuing support, friendship and guidance throughout the course of this work, and for his insightful supervision, without which this thesis would not have been possible. I wholeheartedly look forward to working with him in the future on the topic of forensic science. I would also like to thank the Policy Ethics and Life Science (PEALS) group for providing much-needed financial support for myself and the work, and wish to thank the members of PEALS, especially Professor Erica Haimes, for their ongoing friendship and support.

I must, very importantly, thank my family for their continued love and support – Mum, Dad, plus Claire and Owain – and for having them put up with my opinionated ramblings about my PhD and the fortunes of Widnes Vikings Rugby League Club.

I would also like to thank Ustinov College for additional financial support, and for providing an environment in which to indulge my twin passions of soul music and beer, and for allowing me to inflict the former on fellow College members at regular intervals. Whilst at Durham I have been fortunate to get to know many fellow postgraduate researchers through my activities with the Institute of Advanced Study, and would like to thank them for their friendship and feedback. In particular, I extend my thanks to my fellow travellers of the *Kaleidoscope* – Ali, Beth, Shane and Ian – and wish them all the best in their own studies.

Last, but by no means least, I would like to thank a number of friends, old and new, for providing much-needed and welcome diversions from the rigours of study. In particular, I wish to thank Martin, Anthony, Sarah, Phil, Jon, Andrew, John, Kate, Katie, Charlotte, Rob, Rich and Joe, and hope that many of you will be able to join me in the pub now that this thesis is finished.

Declaration

The contents of this thesis are produced solely for the qualification of Doctor of Philosophy at Durham University and consist of the author's original individual contributions with appropriate recognition of any references being indicated throughout.

Statement of Copyright

CONTENTS

FIGURES AND TABLES

FIGURES **Page**

TABLES **Page**

ABSTRACT

This thesis investigates the reasoning practices of forensic scientists, with specific focus on the application of the Bayesian form of probabilistic reasoning to forensic science matters. Facilitated in part by the insights of evidence scholarship, Bayes Theorem has been advocated as an essential resource for the interpretation and evaluation of forensic evidence, and has been used to support the production of specific technologies designed to aid forensic scientists in these processes.

In the course of this research I have explored the ways in which Bayesian reasoning can be regarded as a socially constructed collection of practices, despite proposals that it is simply a logical way to reason about evidence. My data are drawn from two case studies. In the first, I demonstrate how the Bayesian algorithms used for the interpretation of complex DNA profiles are themselves elaborately constructed devices necessary for the anchoring of scientific practice to forensic contexts. In the second case study, an investigation of a more generalised framework of forensic investigation known as the Case Assessment and Interpretation (CAI) model, I show how the enactment of Bayesian reasoning is dependent on a series of embodied, experiential and intersubjective knowledge-forming activities.

Whilst these practices may seem to be largely independent of theoretical representations of Bayesian reasoning, they are nonetheless necessary to bring the latter into being. This is at least partially due to the ambiguities and liminalities encountered in the process of applying Bayesianism to forensic investigation, and also may result from the heavy informational demands placed on the reasoner. I argue that these practices, or 'forms of Bayes', are necessary in order to negotiate areas of ontological uncertainty.

The results of this thesis therefore challenge prevailing conceptions of Bayes Theorem as a universal, immutable signifier, able to be put to work unproblematically in any substantive domain. Instead, I have been able to highlight the diverse range of practices required for 'Bayesian' reasoners to negotiate the sociomaterial contingencies exposed in the process of its application.

CHAPTER ONE - INTRODUCTION

> 'Much has been said and published about the educated college policeman and detective and it is all bunk. Give me the practical detective with actual experience in handling criminals and criminal cases and with ten such men I will do more work than any college professor or so-called expert can do with one hundred of his trained nuts. Most of those that I have seen couldn't put a harness on a mule, let alone catch a crook...There is an overabundance of self-styled scientific detectives and crime experts in this country. They would have a gullible public believe they are so scientific that the crooks would respond to engraved invitations to visit police headquarters and surrender. Just how long the public will stand for this rot is a question' (Dunlap 1931)

Such were the brusque words of Captain Duncan Mathewson, the longstanding Chief of Detectives at the San Francisco Police Department, expressed at a conference of the International Association of Chiefs of Police. Captain Mathewson's remarks were originally quoted in an article published in *The American Journal of Police Science*, a leading periodical of its time in 1931, which addressed the issue of 'Science vs Practical Common Sense in Crime Detection' (Dunlap 1931).

The author of this particular article, Al Dunlap, the editor of *The Detective*, continues in an equally colourful vein, in his remarks concerning the supposed sophistication of his European police counterparts. Mention is made of reports from Europe, where the investigators of crime 'have some kind of magic wand called science, with which they are able to solve all crime mysteries', and that all their cops and detectives are scientists' (Dunlap 1931, p.322). Dunlap contrasts Europe, where 'every police station...is pictured as a great crime laboratory wherein the detective solves crime problems just as a chemist in America analyzes bootleg liquor', with the supposed inferiority of the 'old fashioned' and 'ignorant and incompetent' American detective, who should be 'displaced by a college-trained scientist' (Dunlap 1931, p.322).

The author is at pains to reassure his readers that, having had the chance to visit his counterparts in Europe, these tales of the 'alleged miracles performed by their scientific crime laboratories' are largely myths (Dunlap 1931, p.323). Instead, facilities for the scientific investigation of crime in Europe are portrayed as 'decidedly disappointing in view of the extravagant claims that have been published broadcast throughout the world' (Dunlap 1931, p.323). Of the much-vaunted crime laboratories that do exist, Dunlap rather bizarrely alleges that they bear a 'striking resemblance' to the 'old-fashioned dime museums originally started by P.T. Barnum and containing wax figures of characters such as Jack the Ripper and Jesse James' (Dunlap 1931, p.324). According to the author, the stories emanating from the Europe concerning the contribution of science to solve crime have been overblown:

> 'Much of the wrong impression now prevalent in America about the European miracles wrought by science are unquestionably due to the interesting reports of cases solved by analyzing the wax from the suspects ears and scrapings of his fingernails. These reports are no doubt, authentic and true; but the fact remains that probably not more than one case in ten thousand could be solved...by this particular method. As against these stories, our experienced detective can point to numerous cases solved purely through common-sense methods and without wasting valuable time examining ear-wax and fingernail scrapings.' (Dunlap 1931, p.324).

To add further weight to his argument, he cites a case involving the murder of female student which took place on the campus of Northwestern University, Illinois:

> 'It was a so-called baffling crime mystery. With all the great scientists of Northwestern University available, it was decided that this crime should be solved by science. The various professors got together for a conference, went into a huddle and became scientific detectives. They photographed the scene of the crime, searched thoroughly for clues, analyzed the soil and everything found nearby and adjourned to a further date for a further scientific conference. Meanwhile, an old-time Irish detective assigned to the case by the police department, using only practical common-sense methods gained by long experience, solved the crime, brought in the

> murderer and had a complete confession – all within forty-eight hours. He simply noted carefully the description of a watch missing from the victim's handbag; then found where a small boy purchased such a watch from a negro bootblack for fifty cents. The bootblack was the murderer.' (Dunlap 1931, p.324).

That is not to say that scientific methods are regarded as lacking application to the field of criminal investigation. In his article, Dunlap is quick to distinguish between the two domains, limiting his conception of 'science' to 'those various branches of scientific research that we hear so much about in connection with modern crime detection such as biology, pathology, toxicology, bacteriology, parasitology and the like' (Dunlap 1931, p.323). He does not reject out of hand the potential contribution that science can make: 'no sane person should care to under-estimate the value of any scientific means for solving crime problems. Science in all its branches should be called into play wherever there is possible use for it' (Dunlap 1931, p.325). Furthermore, he celebrates the achievements of figures such as Calvin Goddard, the director of the Scientific Crime Detection Laboratory at Northwestern University, viewed as a key centre for furthering cutting-edge scientific methods. Alongside Goddard, a number of key figures are also hailed for their work in developments in fields such as ballistics, microscopy, chemistry, and polygraphy.

Ultimately however, the contribution of science in the course of crime investigation is viewed as limited. Instead, Dunlap emphasises 'that immense body of crime investigating forces who have used practical common sense as their chief asset in the great bulk of the work of crime detection' (Dunlap 1931, p.326). Whilst investigators should welcome the input of established scientific methods where appropriate, 'science' as a whole, needs to know its place:

> 'Science should simply confine its efforts to the solution of all problems that call for special scientific treatment, and never undertake to steal the show, so to speak, by underrating the importance of practical common-sense methods in the general investigation of practical common-sense

> methods in the general investigation of nearly all crime cases...Science is not, by the wildest stretch of the imagination, a substitute for practical methods, as the public has often been led to believe; instead it is just a most valuable acquisition and potent aid which should go hand in hand with plain practical common sense and good judgement in a combined effort to cope with the difficult crime situation that confronts the law-enforcement agencies in every section of the land' (Dunlap 1931, pp.326-327).

Hence, in this interwar reflection on criminal detection, 'science' is kept distinct. It is seen to encompass a number of recognisable disciplines, each of which might yield insights that may be of relevance in only a relatively limited number of cases. However, as the example of the Northwestern campus murder is intended to show, the 'practical common-sense' mode of reasoning is, in this missive at least, viewed as superior to the consciously 'objective' non-experiential but empirical reasoning processes employed by the scientists. Thus here, scientific disciplines are, as a whole, viewed as providing mere technologies, sources of aid through which some useful and relevant information might be garnered that may help the detective in his quest to solve a particular crime. Whatever evidence may be revealed via these methods, however, is seen as only one piece in a bigger puzzle, the resolution of which is seen as being better suited to the methods of the 'old time successful crime investigator' (Dunlap 1931, p.327).

We cannot speculate further on the precise leanings of the author with regard to his philosophical views as to what constitutes 'scientific' reasoning, but it is clear that he sees no place for the conscious use of a 'scientific' mode of reasoning, (regardless of specific epistemological discussions), in the world of criminal detection. For the author, there is no substitute for 'the natural aptitude for crime investigation, the genius and skill for getting results displayed by experienced detectives of the so-called old school' (Dunlap 1931, p.323).

Hence, as the title of Dunlap's article makes clear, the reasoning processes of 'science' and those of 'practical common-sense' associated with 'old school' detectives are regarded as wholly distinct with the latter being usually preferred for the task in hand. These two forms of reasoning can thus be viewed as demarcating specific roles (scientist, detective) and practices (scientific inquiry vs criminal detection). Here, the reasoning employed by the detective is of the 'old school 'variety; it relies on skill, experience and intuition. 'Science', on the other hand, is seen as abstract, esoteric, but only of limited use to criminal detection, where an everyday sense of the nuances which constitute social life is essential to understand the doings of criminals.

What Dunlap's comments provide is a sense of the long, complex and ongoing debate about the extent to which 'science' can be employed in the context of criminal investigation. In the modern age, and in the light of the portrayal of forensic science in both the news media and in popular television programmes such as *CSI*, one may be forgiven for assuming that things have changed so much over the past seventy years or so, that science is now an inextricable and vital part of the detection process – indeed, science *is* the detection process. Given this portrayal, it is perhaps easy to assume that criminal detection occurs merely via the click of a computer mouse; if one is able to derive a DNA match with a suspect, then the difficult cognitive work of comprehending the evidence is done. All that is left to be achieved is to apprehend the suspect and bring them to justice.

This popular representation of forensic science does, however, raise a number of further questions. First, there is the question of the extent to which science *can* be held responsible for modern detection practices. This however, begs more questions: '*how* and *why* has science come to exert such an influence? What, if anything, has changed since the criticisms raised by Al Dunlap in the 1930s? In his critique Dunlap does emphasise what he sees as a certain over-estimation of the capabilities of crime laboratories of the time, as well as heightened expectations of the public in terms of what science can achieve. It is perhaps open to question whether either of those situations have changed to any great extent; yet if we return to the question of how and why science has come to be so influential, it leads us to consider precisely what it is that grants science the power to exert such a hold in this context. It leads one to consider what is behind this representation of forensic science. New technologies such as DNA profiling are, at least in the popular idiom, viewed as sophisticated systems involving complex science. However, this complexity is possibly taken as an indicator of an immutable product. DNA profiling may be viewed, in one way at least, as a scientific 'black box' (Latour and Woolgar 1979). It produces results which often play an instrumental role in securing justice, but considerations concerning the inner workings of these systems may be sidelined. A closer investigation of the inner workings of the system may risk stepping into realms of further complexity, but in doing so it may reveal a hitherto unforeseen set of circumstances behind this apparently 'objective' technology.

Another related issue that arises is that of *what*, or *who*, dictates the outcome of forensic decision-making via technologies such as forensic DNA profiling. Furthermore, a consideration of black-boxed technology raises the question of what kind of reasoning processes are buried within these systems. What is the nature of the calculative procedures which lead to the generation of evidence? Are such forms of reasoning comprehendable, and justifiable, in a legal context? What relationship do these forms have with the traditions of the adversarial system? Then there are issues of *accountability*, an area which touches on the notion of 'expert testimony'. The ways in which 'expert witnesses' may be defined and identified is one that continues to elicit interest within the field of legal scholarship, in which the issue of admissibility plays a notable part in such discussions. However, the topic of expertise has also aroused a certain amount of interest within sociological circles. Here, interest has centred around the issue of how social actors may come to recognise 'experts' and 'expertise'. Sociologists have been interested in the practices through which such definitions may be drawn, and some have focused on the way in which 'expertise' is produced in the courtroom. These kinds of studies hold interest for this discussion. For example, with regard to forensic scientific evidence, who can be identified as being responsible for the evidence which is produced? How may they justify their responsibility, and hence authority, for this evidence? And by what precise means do they comprehend and interpret evidence to support the kinds of conclusions they report in court?

This final question is of particular interest to this thesis. Previous studies have tended to highlight the way in which social framings, such as courtrooms, act as *settings* for the construction and bestowal of 'expertise', and the practices which are involved. This study takes a different approach, in that it focuses less on the specifics of settings, and more on the practices involved in creating 'expert' knowledge. The emphasis in this study is on how these practices relate to *intersubjective understandings* of what constitutes expert knowledge, and I aim to make visible the role of these understandings by examining certain technologies of reasoning. The kinds of technologies outlined in this thesis exert relatively little visibility in the courtroom; instead they are largely the preserve of forensic scientists responsible for interpreting evidence of use in criminal investigations. The thesis therefore aims to provide a view of forensic science which is not normally rendered visible in courtroom or laboratory studies. In doing so, it is partly my aim to demonstrate the importance of such practices which may lead to the eventual construction and presentation of expert evidence in court. I seek to show how these technologies themselves reveal, upon close investigation, a series of ambiguities and areas of contestation, and thus I aim to shed some light on an area previously not covered in the science studies literature.

My specific focus of study is a form of probability theory known as Bayes Theorem. This is finding increasing application in a variety of fields, and has elicited a significant degree of interest amongst many working in the area of forensic science in the UK and abroad. Bayes Theorem has informed the development of a number of technologies involved in the interpretation of forensic evidence. In the context of forensic science, Bayes Theorem is viewed as a logical probabilistic framework for combining beliefs about numerous pieces of evidence relevant to a case. It is also said to enable reasoners to provide a convenient and accurate means of updating measures of belief in the light of incoming information. In some cases, the claims made of Bayes Theorem are that it provides a wholesale system for the general interpretation of forensic evidence, in a manner which suggests attempts to position Bayes as some form of defining feature of a new disciplinary identity for forensic science. Hence, Bayes Theorem, and the forms of 'Bayesian reasoning' which arise from it, are currently attracting considerable interest, and controversy, within forensic scientific communities.

The broadening of the use of Bayes Theorem in forensic science suggests a possible change in attitudes toward the role of science in criminal investigation. The words of Al Dunlap suggest an opinion that science only had but a part to play in the criminal investigation of the time. Yet in modern times this conception appears to have become somewhat blurred with the introduction of DNA profiling, typically viewed by publics as a particularly powerful form of evidence. The potential contribution that such evidence may make to the outcome of a case may be far greater than other non-scientific forms of evidence, such as eyewitness testimony for example. However, Bayes Theorem can possibly be seen to blur these boundaries further. It is a distinctly mathematical construct, yet one that is said to measure 'subjective' probabilities of belief, although these must be expressed in numerical terms. Bayes has found use in assessing the significance of DNA profiles, but it has also been used as a basis for methods which seek to provide guidelines for reasoning about evidence as a whole. This latter development in particular shows how Bayes Theorem becomes at once both an abstract mathematical construct, but also a means of guiding the reasoning of individuals involved in the process of evidential interpretation. Bayes therefore is not just another scientific tool, but potentially becomes a way of reasoning through an entire investigation.

A study of the ways in which Bayes Theorem is used in forensic investigation therefore provides an interesting and relevant opportunity to study the reasoning processes involved in the apprehension and construction of forensic evidence. As I attempt to illustrate in the course of this thesis, Bayes Theorem should not be merely regarded as an abstract mathematical equation, transcendent of social behaviour. I must make clear however, that I accept that Bayes Theorem is derived from a formal axiomatic system, and that my focus here is on the ways the theorem is used and interpreted in forensic science. It acts, however, as an important locus around which social practices occur. Viewing Bayes Theorem, and 'Bayesian reasoning' in such a way, enables me to posit an important contrast; namely between the idealised accounts of forensic investigation which occur regularly in the literature, and forensic investigation as it is actually practiced. My work builds on previous studies in this regard (Williams 2007), by showing how even the supposedly cognitive processes of forensic investigation are dependent on a range of practices (Williams 2007). Through these practices, renditions of Bayes arise which both facilitate, and are constituted by, intersubjective understanding.

The thesis proceeds as follows. Chapter 2 presents a review of relevant literature pertaining to work carried out within, and related to the field of Science and Technology Studies (STS). Here I present an overview of certain STS positions, such as Actor-Network Theory (ANT) and ethnomethodological approaches, and discuss the influence of these approaches on sociological studies of the relationship between law and science, and forensic science. I also discuss the potential of using other concepts developed within STS, with particular focus granted to the notion of performativity. I discuss work which has sought to show how expertise is constructed in localised settings such as courtrooms, and the ways in which the credibility of certain forensic scientific techniques may be constructed. I also provide an overview of sociological studies that have been carried out on the topic of forensic DNA profiling. In Chapter 3, I discuss and defend my methodological approach. I describe the resources that were utilised in the course of my research, and I present the lines of inquiry that were followed during the course of this study.

In Chapter 4, I discuss the origins of Bayesian probability. Here I attempt to argue that the original aim of the development of Bayes Theorem was different to how it subsequently developed; hence I aim to demonstrate how Bayesianism is essentially a social construct. I show here, how a version of Bayesianism arose out of an attempt to define itself against other competing modes of probability.

Chapter 5 provides a history of attempts to build theoretical approaches to evidence, which have subsequently influenced developments in forensic science. Although the field known alternatively as 'evidence studies' or 'evidence scholarship' is still arguably developing as a discipline, I show how it can trace its roots back to the early part of the 20th century. I present an overview of significant developments in this area, beginning with the work of the legal scholar John Henry Wigmore, who developed a graphical system of evidential interpretation. I then discuss more recent developments, including the use of Bayesian methods, and describe the impact of this area of study on forensic science. I compare and contrast Wigmore's methods with newer initiatives, and focus upon recent attempts to combine graphical methods with probabilistic approaches.

In Chapter 6, I provide an overview of the scientific basis which has underpinned the development of new technologies related to DNA profiling. I describe how legal challenges to DNA profiling exposed sources of uncertainty and contestation within these scientific foundations. I also discuss the application of probabilistic theories and approaches such as Bayes Theorem to DNA profiling, and draw upon the literature to describe the related controversies that have arisen concerning the reporting and interpretation of DNA profile data.

These chapters are intended to provide sufficient background for what is to follow in subsequent chapters, in which I focus in more depth on applications of Bayes Theorem to forensic science. These chapters are case studies which I use to study in more detail the relationship between theoretical approaches to the assessment and interpretation of evidence, and attempts to apply these approaches in certain forensic scientific contexts. In Chapter 7 I focus upon the development of automated systems used to interpret complex DNA profiles. The resolution of DNA profiles is by no means unproblematic, particularly in cases involving partial profiles, mixed DNA profiles or those obtained by the Low Copy Number LCN technique. Hence automated systems for their interpretation using Bayes-derived algorithms have been developed, most notably by the UK Forensic Science Service (FSS). In this case study I explore various issues related to these technologies. I show how these technologies demonstrate a conflation of the subjective/objective work involved in interpreting profiles, and how the process of interpretation is itself a constructed phenomenon. I seek to achieve this by closely investigating the scientific basis of these technologies, as given in the relevant technical literature. Furthermore, I show how these technologies have contributed to a change in the nature of the organisation of police investigation of crimes, and how new practices, such as the use of media, have been used to project a certain image of these technologies. However, I also show how these Bayesian technologies are also sites of contestation: in recent times these technologies have come under renewed challenge from the legal realm. Via a history of this legal challenge, and recourse to qualitative interview material, I recall how this area of contestation has developed.

My second case study, which forms the topic of Chapter 8, widens the study of Bayes to focus upon the Case Assessment and Interpretation (CAI) model. This model has been developed in order to provide a general framework for the apprehension and comprehension of evidence in forensic casework. It uses principles based on the Bayesian mode of probability in an attempt to guide the reasoning processes of forensic scientists and police investigators. In this chapter I first chart the development of the theoretical principles of the model via an investigation of the relevant technical literature. With recourse to qualitative interview material I then attempt to depict the experience of using the CAI to reason in a 'Bayesian' manner in the context of forensic investigation. My intention here is to highlight practices which are not reflective of the theoretical literature, but which still are very much an inextricable part of what it means to perform 'Bayesian' reasoning in this particular context. I also include a discussion of the issues faced by the developers of the CAI in trying to facilitate acceptance of the model. In this way I show how this framework, based on supposedly 'objective' statistical and probabilistic principles, has not received the kind of acceptance one might expect from such a supposedly 'scientific' construct. Hence I show how a series of pragmatic concerns, in the widest sense of the term, constrict the meaningful use of the CAI.

In Chapter 9 I re-assess and evaluative the results of the two case studies. Here I focus on the implications the results of the two case studies hold for relevant STS approaches, and for certain concepts of pragmatic reasoning which have been cited by evidence scholars and forensic scientists as forming the basis of investigative reasoning. I argue that my research depicts a complex picture in relation to previous relevant work, but that ultimately, it points to a need to consider in more detail the role of human agency in facilitating intersubjective understandings within scientific networks. Whilst the creation of scientific networks may involve a relational character, where actors and objects are arranged in a particular fashion, distinctly human practices play a vital role in constructing these networks. Human agency should, therefore, not be under-emphasised in approaches which seek to explain the stabilisation of such networks, as 'expertise' has a heterogeneous and intersubjective aspect facilitated by human practices. I conclude this chapter by discussing the implications that my work has for the field of evidence scholarship. Finally, in Chapter 10, I conclude the thesis by briefly considering possible future directions for further research in this area, and how this work may fit into a wider context.

CHAPTER TWO - THEORETICAL BACKGROUND

2.1 Introduction

Technologies of forensic science arise in the intersection between the realms of science and law. This is a space in which the need to determine justice meets with the need to maintain scientific propriety, and therefore where two ostensibly well-entrenched epistemological traditions cross paths. Forensic technologies, and the actors and institutions concerned with them, are embedded in this contested terrain. The processes of negotiating such an environment can be a viewed as a complex sociomaterial achievement, and these processes also show how the boundaries between 'scientific' and 'non-scientific', 'natural' and 'social', may be fluid, contingent and constructed. This is a theme which permeates much of the literature in the field of Science and Technology Studies (STS), and a small but significant number of STS analyses have devoted attention to the construction of forensic evidence. In this chapter I draw upon this literature to describe both its theoretical approach in general, and also more specific applications to forensic science. In doing so, I seek to introduce a number of studies which have considered the emergence of new scientific objects as 'co-products' of both nature and a variety of social agents (Jasanoff 2004). They also show how the distinction between the 'natural' and the 'social' is often contestable, and how the origins of such objects are difficult to trace to a definitive extent, emerging as they do from a complex network of practices, institutions and artefacts. By drawing upon these studies I aim to illustrate how scientific objects can be seen to originate from, and in some cases perpetuate, a state of epistemological and ontological flux.

I first discuss Actor-Network Theory (ANT), along with a closely related approach which centres around the concept of performativity. I will use the latter approach as a means of introducing some of the research areas I seek to pursue in the course of this thesis in relation to Bayes Theorem. I then discuss the implications of previous sociological studies of forensic science for the evaluation of ANT-related approaches.

My intention here is to foreground the areas which the studies outlined seek to develop. From this I move on to discuss the problem of defining what may be recognised as 'scientific' or 'expert' knowledge in the forensic realm. This is an issue which has been of interest to both sociologists and legal scholars. Sociologists have sought to identify and characterise the ways in which 'expertise' is constructed and recognised, and I show how various authors have approached the issue. Some of the sociological literature has specifically focused on the way in which 'expertise' is constructed in legal settings, and in related areas such as forensic science. I discuss this literature, and I also briefly describe the general problem of defining 'admissible' scientific evidence.

Finally I conclude the review by indicating how the literature discussed leads me towards the area on which I will subsequently focus in more detail. This concerns the use of probabilistic reasoning in forensic science, and in particular the growing influence of Bayes Theorem amongst forensic practitioners. In the next chapter, I describe how the theoretical insights outlined in the first section have acted as a guide for inquiry into an area of forensic scientific practice that has hitherto received relatively little attention from STS, but which, as I hope to show in the course of this thesis, can be illuminated by social analysis.

2.2 Networks of Science

Modern scientific enterprise is an increasingly multi-faceted socio-organisational process. The manner in which 'society' may be seen to shape scientific knowledge and practice has been studied in various ways. For example, Merton (1973) saw communities of scientists as adhering to a shared set of ideals concerning the goals and methods of science (Merton 1973).[1] The so-called 'Strong Programme' of science studies has, *inter alia*, highlighted the role of individual and social interests in resolving scientific controversies (Collins and Pinch 1993). Other works have sought to describe in more detail the complexity of the linkages between science and the

[1] Namely communalism (common ownership of scientific discoveries), universalism (claims to truth are evaluated in accordance with universal and criteria, and not on the basis of race, class, nationality, religion or gender), disinterestedness (scientists are rewarded for acting in ways that outwardly appear to be selfless), and organised scepticism (all ideas must be tested and subject to rigorous collective scrutiny).

wider social realm, and have shown how scientific inquiry is constituted by a rich array of actors and material entities. In this latter regard, a number of studies have described the emergence of 'technoscientific' practices and objects with an emphasis on the organisational networks which underpin them. Such networks may represent the convergence of a heterogeneous combination of interests, actors, disciplines, organisations and technologies (Latour 1999). Whilst the objective products rely on such networks for their existence, they also exert certain effects in return. Accordingly scholars have sought to describe the conditions which have led actors and organisations to co-ordinate their activities in the pursuit of certain desired scientific outcomes, and there has been a corresponding focus on the epistemic changes brought about by the realisation of such projects (Nowotny et al. 2001).

The Actor-Network Theory (ANT) approach, as advocated by Latour, Law and others, has played a particularly prominent role in demonstrating how scientific objects emerge and become embedded in networks through which new knowledge may circulate, perpetuate and adhere (Latour 1987; Latour 1990; Law and Hassard 1999). Although the precise nature and aims of ANT have been subject to some deliberation, certain common features are apparent. ANT studies generally accord human and non-human actors symmetric ontological status prior to their positioning within networks, which consist of a heterogeneous assemblage of actors and physical objects. It is these networks, or *actor-networks* to which they are often referred, which are able to exhibit collective agency. The role that each component plays in sustaining the network is relational, and the strength of the actor-network is dependent on the extent to which information is able to flow through the network (Latour 1999; Law and Hassard 1999). According to ANT, the strength of the scientific claims emanating from a network will be a consequence of the strength of the links in a network, and in its scope: convincing scientific claims require strong networks which reflect powerful forces.

The network concept broadly advocated by ANT has spawned a relatively small number of studies of forensic science, but these have demonstrated how the networks which perpetuate forensic technologies may reflect the influence of certain wider forces. For example, Williams et al (2004) describe how the development of the National DNA Database (NDNAD) was dependent on a series of changes to the legislation determining the ways in which police were able to legitimately take, store and use DNA samples (Williams et al. 2004). Whilst the existence of the NDNAD was reliant on the commitment of a number of actors, such as the Forensic Science Service (FSS), and the Association of Chief Police Officers (ACPO), it can be seen as a particularly strong example of the way in which the policy of 'New Public Management' (NPM) has been put into practice, involving the introduction of measures designed to monitor and improve the quality of public sector services. Hence the direct influence of policy can be seen in the establishment and subsequent expansion of the NDNAD, and as a manifestation of the desire to standardise, measure and optimise the levels of forensic science service delivery within and across police forces (Williams et al. 2004, p.60).[2]

In another example, Aronson (2008) describes the dominant role played by the Federal Bureau of Investigation (FBI) in shaping the standardisation of DNA profiling in the US in the 1980s and 1990s. The history he describes contrasts with other ANT accounts in that rather than having to actively cultivate an interest in order to compel actors to join the network, the FBI began from an *a priori* position of strength (Aronson 2008, p.197). Hence it did not have to recruit allies in order to establish itself as an obligatory passage point in DNA profiling, and was able to decide which institutions and actors could participate in the process largely by fiat. Aronson identified a number of strategies which enabled the FBI to reinforce this position of *a priori* authority. The FBI used a different restriction enzyme in its DNA protocol than Cellmark and Lifecodes, the two primary commercial suppliers of forensic DNA profiling at the time. The discrepancy between restriction enzymes, used to cleave DNA into fragments prior to profiling, meant that DNA fragments of different sizes would be produced, leading to a lack of uniformity of DNA profiles, and causing validation problems (Aronson 2008, p.201). More notably still, the FBI recruited and trained individuals with previously little or no experience in molecular biological techniques to perform their protocols. It was through these individuals, working in the network of existing public crime laboratories, that the FBI optimised its own set of standards and protocols, to the exclusion of the commercial companies. Although the standard-setting process would eventually widen to include several other influential

[2] However, the extent to which the NDNAD has impacted upon detection levels has been questioned (McCartney 2006).

parties and institutions,[3] these debates ultimately centred around technology constructed by the FBI. Hence, Aronson argues that the FBI 'built a forensic DNA typing [sic] network by constructing the network and the human, material and social aspects of it at the same time' (Aronson 2008, p.213).

Government policy, and the actions of powerful agencies such as the FBI, is not the only driving forces that have been identified as playing a role in the development of forensic DNA profiling. Other studies have highlighted the potential role of commercial enterprise. Daemmrich (1998) has focused on the construction of 'convincing expert testimony' by American DNA typing firms (Daemmrich 1998). He argues that these companies employed a strategy of 'vertical integration' in order to stabilise a form of knowledge able to withstand the highly contested domain of judicial identity testing. Daemmrich charts how firms sought to control as many facets of the forensic DNA testing process as possible, from basic method development and related research through to the production of DNA probes and other such reagents. The same firms also introduced complex bureaucratic procedures, producing a 'chain of custody' capable of guaranteeing the integrity of DNA samples, and were instrumental in forming self-regulatory organisations through which scientific procedures were standardised and validated. Daemmrich describes how these companies even provide training schemes to enable future expert witnesses to hone their skills in conveying evidence in the notoriously adversarial arena of the courtroom. The analysis thus demonstrates how commercial firms have had to manage a diverse array of activities and products, both 'upstream' and 'downstream' of the act of DNA testing in order to maintain the appearance of a coherent, convincing and credible form of knowledge.

Daemmrich's study displays how organisations, given sufficient resources, are able to structure sites of production in order to provide a suitable supporting network for their epistemic claims. However, it may be possible that the underlying values and interests of organisations may penetrate deeper into the realm of science with the corollary of influencing the course of development of new scientific paradigms. For example,

[3] Which included the US National Institute of Standards and Technology, the National Academy of Sciences National Research Council, Congress, the legal system and several leading scientific journals (Aronson 2008, p.213).

economic interests may become realised through the advent of new fields such as bioinformatics. In his exploration of this field, Adrian Mackenzie (2003) has argued that, by it's very nature, bioinformatics represents a particularly germane means for economic and commercial concerns to dictate the progress of scientific advances. Mackenzie sees bioinformatics as 'an economic process of ordering certain abstract determinations of living bodies so that information can circulate (more) freely through them' (Mackenzie 2003). Attempts to maximise the potential of the new discovery science are constrained by commercial pressures. Here there may lie implications for forensic DNA profiling and databasing, in that commercial interests may come to influence the future development of related technologies equally as much as the interests of government or law enforcement agencies.

ANT has also highlighted the ways in which consensus may be reached on complex and problematic areas of science. The attempt to construct objective knowledge in this regard can be construed as an attempt to 'black-box' subjective judgement. Black-boxing can also be seen to represent a form of closure regarding the production of scientific objects when their status is unquestioningly accepted by an epistemic community (Latour 1999). Latour and Woolgar (1979) argue that the creation of black boxes is a key element of scientific activity. They view black-boxing as the activity of 'rendering items of knowledge distinct from the circumstances of their creation' (Latour and Woolgar 1979), p.259-260). In this way, scientific activity is seen to incorporate a process of establishing scientific facts as somehow transcendent of the practices that gave rise to them. Latour and Woolgar give the example of a mass spectrometer as a scientific black box (Latour and Woolgar 1979, p.150). The technique of mass spectrometry, which is commonly used in chemistry and biochemistry to analyse the molecular weight of chemical molecules, has been applied to these fields for several decades. Although modern mass spectrometry can itself be construed as a complex physical and chemical process, the results it produces are often accepted unquestioningly, particularly if they contribute toward the progress of a scientific discovery. For example, Latour and Woolgar show how mass spectrometric data was used in establishing the discovery of a hormone known as Thyrotropin Releasing Factor (TRF). The results of the mass spectrometer provided an important locus around which the existence of TRF could be discussed and agreed upon; however the scientific basis of mass spectrometry as a technique was never questioned during the course of the scientific work. To open the 'black box' of mass spectrometry would have been highly costly in terms of time and resources for the laboratory, whilst the pursuit of TRF was the priority. Black-boxing of scientific artefacts is, therefore, a necessary part of the process of scientific work, for it may be necessary to close down sources of contestation in order for the given scientific goals to be attained.

In a forensic context, Halfon (1998) provides a study of how technologies maintain their black-box status in order to maintain an appearance of objectivity and thus credibility in courtroom settings (Halfon 1998). Whether or not the integrity of scientific objects is challenged is possibly dependent on how well they relate to relevant domains of expertise. Halfon (1998) has investigated the issue with recourse to forensic DNA profiling (Halfon 1998). He argues that objects which do not fully align with the purview of expert communities go unchallenged, and that this indirect process of closure is never subsequently contested in the courtroom. It is only those elements which have stimulated controversy between experts (i.e. those elements which they feel suitably qualified to comment upon) that may form sites of contestation in the courts. Through interviews with four scientists who have been involved in DNA cases to a varying extent, Halfon demonstrates how areas of closure became mutually dependent on the delimitations of the areas of expertise. In each of these cases individuals were able to negotiate areas of contested knowledge by defining narrow areas of expertise which overlooked certain 'extraneous' concerns (Halfon 1998, p.817). This gradual process of constraining the focus of criticism also narrowed the number of relevant individuals involved, and thus facilitated closure of controversies, which became reduced to a few technical issues (Halfon 1998, p.822).

Linda Derksen, in her exploration of issues relating to the development and acceptance of DNA profiling, has also sought to demonstrate how supposedly 'objective' knowledge is achieved via a series of markedly social processes and interactions, which attempt to 'erase the actions of the representing subject from the representations made of the natural world' ((Derksen 2000), p.803). Derksen argues that measurement is the result of a series of negotiations between nature, and, equally importantly, people. It was from social practices that judgements arose regarding what might be considered 'normal', what was accurate for practical purposes, and when procedures could be judged to have been competently performed. Socially constructed judgements and evaluations were therefore crucial for establishing what constituted a sufficient degree of accuracy and competence for the task at hand. Derksen argues that a direct link existed between the 'personal judgements, evaluations and negotiations' involved in assessing measurement error of DNA profiles on the one hand, and the debate between a disparate collection of actors concerning the calculation of random match probabilities on the other. Legal acceptance of the accuracy and objectivity of forensic measurements resulted from a series of efforts to establish and represent 'consensus' amongst the relevant epistemic communities.

Thus human activity, be it in the form of organised action, can also be seen to play an important role in the construction of 'objective' credible knowledge. Cole (1998), in his study of the rise of fingerprint evidence, argues that a representation of credibility in this case came about via the construction, by fingerprint examiners themselves, of specific rules and practices (Cole 1998). Firstly, dactyloscopists have always insisted that fingerprints be examined only under the guidance and supervision of an expert, arguing that correct fingerprint interpretation requires the requisite 'expert' training and experience. Second, dactyloscopists organisations such as the International Association for Identification (IAI) formulated rules preventing individual examiners from testifying in contradiction to one another. Third, cases of error were attributed to individual examiners rather than questioning the technique itself; once again this position has largely been mediated by dactyloscopy organisations and unions. Cole argues that these strategies have served to convey an image of fingerprint analysis as a specialised science, requiring years of training in interpretation. Furthermore, the decision to close ranks can be taken to be a sign of an attempt to present a unified front, thus attempting to convey an image of a unified and thus credible discipline. The point that should be taken here is that, whilst inscriptions are involved in relaying knowledge, it also takes social action, in the form of organisation, to portray credibility.

One feature common to many ANT studies is an interest in the way social orderings are able to stabilise and adhere, and how certain social and natural distinctions, such as 'human/non-human', 'natural/non-natural' etc, are constructed. ANT takes a semiotic approach to studying how these constructions arise, showing how each element in a network is defined in relation to the other components of that network (Law and Hassard 1999, p.4). This process of 'material relationality' (Law 1999, p.4) is seen as a *performed* phenomenon, and thus networks are emergent and potentially fluid entities. Hence ANT is interested in *how* it is that relations are performed, for these practices are never pre-given. The practices through which entities are constructed in such a relational manner are referred to as instances of *performativity* (Law and Hassard 1999, p.4). The concept of performativity has itself developed into a fully fledged area of interest, particularly with regard to economic sociology (Callon 1998; Callon 2006). In what follows I introduce the performativity thesis in more detail.

2.3 Representation and Performativity

The concept of performativity in this context is rooted in a discussion by Pickering (1994), who argues for the existence of two different ontological approaches to science, namely separate *representational* and *performative* idioms (Pickering 1994). According to Pickering, the representational idiom has traditionally predominated in philosophy and science studies, running from logical empiricism to logical positivism, as well as the history of science and the sociology of scientific knowledge, even through attempts to understand science as a textual phenomenon (Pickering 1994, p.413). This idiom is seen to have been preoccupied with the production of representations of nature, including facts and theories. However, Pickering regards this idiom as betraying a dependence on a narrow version of human agency: 'all of the agency in this idiom, then, is the agency of knowledge's human producers, the scientists' (Pickering 1994, p.413). Whilst the representative idiom is seen to have served the disciplinary interests of some philosophers, sociologists and historians, Pickering argues that a preference for alternative approaches to human agency, impedes the possibility of interdisciplinary synthesis (Pickering 1994, p.414).[4]

[4] 'As far as I can make out, within the representationalist idiom it is very difficult to achieve any real synthesis between the science-studies disciplines (or between those disciplines and their parents). The problem is that the disciplines take up quite different perspectives on agency: philosophers reinforce their own disciplinary identity by trying to spell out superlocal characteristics of reason and so on; sociologists are likewise enamoured of local and situated interests, or whatever; and historians tend to oscillate between the two competing master-narratives. The best that one can do towards synthesis in this situation is, it would appear, to add up the rival stories, to run both at once in the approach that I call multidisciplinary eclecticism.' (Pickering 1994, p.414). It may be more appropriate to think of Pickering's ambitions regarding the performative idiom as *antidisicplinary*, as he mentions later in this article: 'the synthesis promoted by the performative idiom is an antidisciplinary one, in which history, philosophy, sociology, anthropology and whatever collapse into one another' (Pickering 1994, p.416).

Pickering continues to argue that the adherence to the representative idiom is not mandatory, and that a second option, the performative idiom, is available which may facilitate a truly interdisciplinary approach. The key feature in this case is the requirement to think 'symmetrically about agency' (Pickering 1994, p.414). To confer agency to material artefacts, Pickering argues, is to accept the true interdependency of every stratum of scientific culture: material, conceptual and social. Furthermore, an adherence to the performative idiom removes any *a priori* assumptions about the nature of human agency, which is 'not, as the representationalist idiom would have it, something reliably given in advance that can provide an enduring explanatory resource in the analysis of scientific knowledge production'. (Pickering 1994, p.415).

At the heart of Pickering's performative idiom is a radical ontological position. According to Pickering, there are no 'pure' objects which lend themselves easily to monodisciplinary study. At the heart of the performative idiom is a radically emergent ontology, in which the different strata of scientific culture are:

> 'continually and constitutively intertwined with the others in practice via processes of heterogeneous interactive stabilisation. The social then, for example, is never purely social; there is no purely social dynamics of scientific practice; and there is, therefore, no room in the world for a pure (e.g. Durkheimian) sociology of science. And the same goes for the material and conceptual.' (Pickering 1994, p.416).

The key consequence that the performative idiom holds for human agency is an apparent 'incommensurability of human powers' insofar as human agency is portrayed as indeterminate, and shaped by the social-material-concept networks in which individuals reside.

Pickering (1994) provides few further guides to what a performative position entails. However, the work of other authors, most notably Callon (1998, 2006), have advanced this thesis (Callon 1998; Callon 2006). The broad concept of performativity has since been applied in a wide range of studies, although the work of Callon and Mackenzie has focused primarily on the field of economics. In a challenge to economic sociology, Callon (1998) has argued that the discipline of economics '*performs* the economy' in that it creates the phenomena it describes. The thesis advanced is that economics, and economic sociology, is wrong to enrich the supposedly calculative, self-centred *homo oeconomicus*, for such an entity does *a priori* exist. Instead, the focus should be to understand *how* such an agent is produced.

In the later of the two studies referred to above, Callon (2006) attempts to advance a more rigorous understanding of the concept of performativity. Here he shows that the basis for the 'performativity thesis' originates in pragmatist positions concerning language (Callon 2006, p.8). For example, Charles Morris argued that the supposed three-fold division of language into syntax (relations between signs), semantics (relationships between signs and what they denote), and pragmatics (relations between signs and their use context), cannot be entirely dissociated from one another (Morris 1938). In commenting on this, Callon argues that, rather than being considered one component of linguistics alongside syntax and semantics, pragmatics actually encompasses the totality of linguistic phenomena. This argument is supported by the work of J.L. Austin, who proposed the distinction between constative utterances ('the cat is on the mat') and performative utterances ('I promise', 'I baptise you', I sentence you to ten years imprisonment' etc). Austin sought to criticise the notion that language is purely representative, and argued that because language is uttered, no statements escape constituting the context in which they function: 'there is no language, only acts of language' (Callon 2006, p.10).

Callon sees no reason why Austin's argument cannot be applied to scientific discourse. The consequence of this, he argues, is that all science is performative, as opposed to simply creating representations of reality (Callon 2006, p.10). Central to his argument in this case is the supposed necessity of singular existential statements (SES) in science. SESs are indexical statements which refer to specific, singular entities and events located in a particular temporal and spatial situation: 'At *such*-and-*such* a place, at *such*-and-*such* a time, *such*-and-*such* a thread can be observed to break when we apply *such*-and-*such* a force over x kilograms' (Callon 2006, p.11). Glossing over the controversies between SESs and universal statements (USs)[5], Callon argues that scientific theories and models require SESs. The *raison d'être* of science is its inductive force; there must be a link between the USs of science, as represented in

[5] Universal statements being of the type 'At all places in time and space, all threads will be observed to break when such-and-such a force over x kilograms is applied'.

theories and models, and SESs which describe the actualised instances that those theories and models seek to capture. Without the ability for USs to beget SESs, there is no science.

The existence of SESs is, of course, dependent on the material worlds they describe. Callon argues that this also implies that these selfsame material worlds have no meaning unless one has a means of describing them, and if they are without meaning they cannot exist as a scientific entity. An SES, therefore, 'is entangled with the device that produced what it describes; the device and the series of actions undertaken are shaped by the statement, and vice-versa' (Callon 2006, p.12). Callon likens SESs to instructions for a complex scientific instrument, which are required for one to understand what the instrument may do, and how it may be used in order to carry out those tasks. Clearly, instructions hold no meaning if they refer to no discernible material entity.

Borrowing a term from Deleuze and Guattari (1983), Callon refers to the relationships between statements and their worlds (sociomaterial complexes), as *agencements* (Deleuze and Guattari 1983). Callon deliberately chooses this term, ahead of 'arrangement' or 'assemblage', in order to emphasise the potential of these constructs to possess agency. '*Agencement* has the same root as agency: *agencements* are arrangements endowed with the capacity of acting in different ways depending on their configuration' (Callon 2006, p.13). However, whilst *agencements* may perform in different ways depending on their fine-grained composition, the basic nature of the relationship, namely the inseparability of statement to referent, remains the same. Because *agencements* encompass statements as well as their referents, there is no need for any external explanatory factor to account for their formation. The construction of its own meaning is a crucial part of an *agencement.*

Of the studies that have attempted to utilise Callon's concept of performativity, the work of Mackenzie et al is worthy of particular attention (Mackenzie 2003; Mackenzie and Millo 2003). Mackenzie (2003) and Mackenzie and Millo, (2003) has used and developed the performativity thesis in their studies of financial derivatives markets, and the use of the Black-Scholes-Merton equation. The latter is a widely used mathematical tool for calculating the price of options in derivative markets.

Mackenzie (2003) charts the development of the equation, arguing that the mathematical work involved in the development of the equation did not correspond to a logical process of rule-following, but was instead the result of a creative, contingent process of *bricolage*, and that it emerged despite certain theoretical disagreements between the developers of the theory (Mackenzie 2003).

In a related paper, Mackenzie and Millo (2003) show how the operation and application of Black-Scholes-Merton has been performative (Mackenzie and Millo 2003). They chart how, in an initial phase, the predictions of option prices made by the model differed sharply with empirical prices. However, as time wore on throughout the 1970s and 1980s, the equation became regarded as a reliable indicator of prices, so much so that it became increasingly integrated into the informational infrastructure of large trading centres such as the Chicago Board of Exchange (Mackenzie and Millo 2003, p.127). In particular, the equation provided a reliable basis for the calculation of 'implied volatility', namely the potential deviation of the value of an option. However, following the US financial crash of October 1987, this predictive ability disappeared, and any subsequent measurements of implied volatility had to be modulated by a daily 'skew' estimate. Mackenzie argues that this latter development, in which behaviours worked against the predictions of the theory, represents an instance of what he refers to as 'counterperformativity'.

Mackenzie also tackles the question of how *homo oeconomicus* can create markets even when such a rational agent wouldn't, in theory, do so. In answer to the question of how to theorise the articulation between performativity and markets in terms of networks, cultures, moral communities etc, Mackenzie suggests that the answer involves 'both impoverishing and enriching conventional economic views of the rational actor' (Mackenzie and Millo 2003, p.140). With regard to the latter, Mackenzie points to the continuing importance of relational values, e.g. respect and reputation, in playing an important role in maintaining the existence of these agencements.

The studies above show how the construction of a 'scientific' technology can be regarded as a socially constructed entity. If science can be construed in such a way, as a complex and wide-ranging interlinkage of sociomaterial performances, then this consequently raises questions concerning *boundaries* of science. How does society demarcate 'science' from 'non-science'? In the following paragraphs I consider this issue in more detail. A related subject concerns the demarcation of forms of 'expertise', and I subsequently discuss this issue.

2.4 Boundaries of Science and Expertise

The concept of 'boundary work' has found significant import in sociological studies of science, most notably in studies by Gieryn (Gieryn 1983; Gieryn 1995). Rather than treat the delimitation of science as an analytical problem for philosophical inquiry, Gieryn frames it as a practical issue faced by scientists themselves. 'Boundary work' is used to describe the practices employed by scientists to demarcate 'scientific' work from 'non-scientific', or 'pseudoscientific' pursuits (Gieryn 1983). In describing these practices, Gieryn shows how science can be construed as a flexibly defined and thus rather ambiguous concept. Using a number of historical case studies, Gieryn (1983) highlights a number of ways in which boundaries may be drawn. For example, he describes how John Tyndall, a Superintendent of the Royal Institution during the Victorian era, used public addresses and writings to advance the authority of science over the previously pre-eminent domains of religion and engineering (Gieryn 1983, pp.784-787). He also shows how anatomists in nineteenth century Edinburgh blocked the acceptance of phrenology as a science by denying practitioners university positions, access to lecture halls and membership of scientific fora (Gieryn 1983, p.789), and how the scientific establishment in the US in the 1980s were able to maintain open research in the face of government concerns over national security (Gieryn 1983, p.791). In each case, Gieryn argues that 'science' was no definite *a priori* entity; instead it had to be constructed, through a variety of strategies, in relation a non-scientific 'other'. Hence ''science is no single thing: characteristics attributed to science vary widely depending upon the specific intellectual or professional activity designated as "non-science", and upon particular goals of the boundary-work.' (Gieryn 1983, p.792).

The construction of boundaries that delineate what is, and what isn't, 'scientific', can be regarded as a social achievement. A crucial factor in producing this delineation is the ability of 'scientific' actors to project a suitably convincing image of 'credibility'.

In the case of forensic science, 'credibility' is equated with public displays of expertise. A concern with the possibility of scientific 'boundaries' reminds us that the bestowal of 'expert' authority is a particularly social phenomenon. This contrasts to the view of classical philosophy of science, which considers 'expert knowledge' as a strictly epistemic phenomenon, able to be unproblematically delineated with recourse to some form of suitable epistemological scrutiny (Selinger and Crease 2006). Both sociology, and more modern philosophical works, have challenged this classical view. The work of Polanyi, for example, (1967) brought to attention the role of inexpressible, or tacit, knowledge in developing beliefs (Polanyi 1967).

Whilst tacit knowledge may take several forms, a key form is embodied skill. Dreyfus (1986) draws upon Merleau-Ponty (1962) to argue that expert judgement and behaviour is essentially a matter of embodied performance (Dreyfus et al. 1986; Merleau-Ponty 2002). Dreyfus argues that all expertise is principally a matter of practical reasoning, 'knowing how' rather than 'knowing that', and proposes a five-stage model to demonstrate how an individual acquires a particular skill. This model effectively represents points on a continuum, beginning with a 'novice' stage in which the individual first learns a context-free set of rules for performing an action, progressing through stages in which the learner gains experience in actualised contexts of performance and gains an awareness of the plurality of nuanced challenges, culminating in the 'expert' stage in which responses to the subtleties of a particular situation become intuitive and virtually automatic.

Dreyfus' position has been criticised for providing an incomplete account of how one comes to recognise expertise in others, and how any claim to expertise involves some form of social demand (Selinger and Crease 2006). Whilst Dreyfus is viewed as presenting a distinctly *asocial* account of how expertise is acquired, he also argues that political, social and cultural movements may obscure the ability to recognise expertise in others. However, Selinger and Crease (2006) raise the issue that it is actually precisely these factors which lead one to recognise expertise in the first place. In ignoring the recognition issue, Dreyfus' position is seen as leading to a paradox. He argues that nonexperts are unable to know what to look for when evaluating skill, and that only experts can recognise other experts (Dreyfus et al 1986, p.201; Selinger and Crease 2006, p.231). However, he also argues that experts *do* behave in a similar way to 'ordinary' people; the mechanism for the acquisition of simple, everyday skills is exactly the same for more complicated and technical competencies. 'We can projectively identify with experts, and understand the kind of knowledge they use in their judgements.' Dreyfus needs this latter claim to be predominant, 'as otherwise nonexperts would lack any basis to recognise, accept and trust the kind of knowledge that experts possess.' (Selinger and Crease 2006, p.232).

Selinger and Crease, along with others, notably Fuller (2006), argue that expertise should be construed as a distinctly social phenomenon; expertise cannot be assigned to oneself, it must recognised and bestowed by society (Selinger and Crease 2006, p.229). Indeed, one of the main themes of sociological accounts of science concerns the conditions in which expertise may come to be constructed, bestowed and recognised. In a similar manner to Gieryn (1983) these kind of accounts question the notion that expertise can be demarcated unproblematically via epistemological means, and instead point to the conditions and processes through the 'expert' is demarcated from the 'non-expert'.

Whilst such studies may serve to describe the conditions in which expertise arises, Collins and Evans (2002) argue that the problematisation of expertise in such a way has considerable political implications. They view this challenging of boundaries as creating a 'problem of extension', over where to draw the line at who to include and exclude when allotting decision-making rights for making scientific and technical policy decisions (Collins and Evans 2002). They propose a new area of study – Studies of Expertise and Experience (SEE), encompassing a normative theory of expertise, in order to 'disentangle expertise from political rights in technical decision-making.' (Collins and Evans 2002, p.235). Suggesting new categories of expertise, they argue for a method of analysis able to identify groups of 'experts' in specific technical disputes that resists the temptation to classify expertise simply on the basis of formalised accreditation, yet they reject a position that is claimed to be commonly found in science studies literature, namely the conflation of the scientific and the public realms (Collins and Evans 2002, p.250). Instead, expertise is classed as 'interactional' (in which an individual is equipped with enough knowledge to comprehend the vocabulary used by a particular group of experts), and 'contributory' (whereby an individual possesses enough knowledge to make a tangible contribution to a particular disciplinary field). Using a series of case studies, they argue that their theory is able to identify when more or less public involvement is required to solve a specific dispute, depending upon the amount of interactional and/or contributory expertise groups may possess, regardless of formal credentials.

The extent to which individuals and groups possess these forms of expertise depends on whether they are able to communicate across disciplinary boundaries and resolve technical disputes. In addition to interactional and contributory expertise, 'referred expertise' may be possessed by an individual (Collins and Evans 2002, p.257). This is where one may have *experience of* contributory expertise, which enables them to engage with individuals despite the latter possessing a different form of contributory expertise. Such referred expertise may be required by those managing or overseeing large scientific projects involving multiple forms of expertise. Furthermore, translation between different forms of expertise may be necessary, and for that interactional expertise is required in order for different experts to communicate freely with one another (Collins and Evans 2002, p.258). Another crucial factor is the ability to discriminate between credible and non-credible instances of 'expertise'. Collins and Evans argue that this ability comes through one's existence as a social actor. Judgements about the credibility of an actor may be formed on the basis of a number of different criteria, such as whether the actor is allied with the 'correct' social networks, whether they can demonstrate the right experience to support the claim, whether they have been known to make credible or non-credible claims in the past , whether the claim is internally or externally consistent, or whether the claim appears to be self-serving (Collins and Evans 2002, p.258). Collins and Evans therefore argue that expertise is largely performed. Language plays a particularly important role, enabling actors to communicate across disciplinary boundaries and to convince others that they possess expertise.

Collins and Evans bring further attention to the manner in which expertise may be identified, but also how expert knowledge may be shared and appropriated. This is an issue which has also received attention in direct relation to forensic science. The studies of Doak and Assimakopoulos (2007a, 2007b) have focused on collaborative networks in forensic science laboratories in the Republic of Ireland. They argue that tacit knowledge plays a key role in the formation of such networks, and that the acquisition of this tacit knowledge is highly dependent on social relations and interactions. The development of expertise was found to rely on a face-to-face interactions; for example the initiation of a nascent digital evidence service in the FSL was found to have been dependent on the attendance of certain individuals at other laboratories where digital evidence techniques were well-established (Doak and Assimakopoulos 2007a). Attendance at forums and conferences was also viewed by scientists as a vital means of gaining tacit knowledge. Furthermore, face-to-face consultations over evidential interpretation, were found to be a habitualised feature of working life within the organisation. Aware of the fact that their knowledge claims could be scrutinised in the courts, forensic scientists routinely conferred with colleagues to check whether they had followed the correct procedures, and to discuss their personal judgements concerning evidence. Possibly due to a familiarisation with the adversarial nature of the courts system, scientists relished the opportunity to have their judgements rigorously challenged by colleagues (Doak and Assimakopoulos 2007a).

These forms of interaction were not the only practices through which knowledge transactions were seen to flow, and Doak and Assimakopoulos (2007b) found that the socialised nature of the organisation was vital in providing an environment in which knowledge could be transferred via a wide array of practices. For example, negotiation of the complex and potentially overwhelming series of codified standard operating procedures (SOPs) was often circumvented by simply asking a colleague for guidance. Experience was also viewed as an important personal asset, and more experienced scientists were viewed as a key resource by their more junior colleagues. Indeed, the experience of handling cases was regarded as eliciting competencies in a manner unable to be captured by SOPs alone. Scientists also readily recognised which of their individuals had particular capability in certain areas, and were often targeted for consultations over specific matters. Furthermore, social practices relating to the maintenance of personal relationships were viewed as vital conduits for relaying tacit knowledge.[6] Informal contexts, such as the meeting of acquaintances during coffee breaks, were found to provide a suitable setting where tacit knowledge might

[6] Doak and Assimakopoulos (2007b) do counter that social interactions may potentially lead to the formation of cliques, which was not necessarily as beneficial. The formation of a number of cliques

permeate casual conversation, and working friendships might actually amount to relationships of coaching and mentoring. The need to reciprocate instances of help was also regarded as highly important, as was the sharing of information and the need for open dialogue. Even personal demeanour was regarded as an important factor, as approachability was cited as another important factor in the effective communication and transfer of knowledge.

The preceding discussion has demonstrated the potential richness of social orderings of science and expertise. This in turn brings attention to the indeterminacy of such concepts, which has been emphasised by other authors. For example, Wynne (1998) challenges the tendency to treat 'science' or 'expertise' as an 'autonomous, objective entity which has authority independent of the institutional settings in which it is used.' (Wynne 1989, p.28), and legal cross-examination has often successfully challenged the immutability and universality of scientific knowledge. For example, legal scrutiny of certain scientific tests, such as the Greiss test for nitroglycerine (which was used in prosecution evidence in the original trial of the Birmingham pub bombers), has shown them to be 'restricted to more limited, special, situations or cases, as hidden conditions or assumptions have been exposed by critical examination.' (Wynne 1989, p.32). However, in such contestations over the status of 'scientific' or 'expert' knowledge, no one position can be taken to be absolute:

> '...studies of scientific controversies have shown how scientific knowledge taken as natural and universal by one school may be exposed as a tissue of selective observations based upon a limited set of localised technical practices and theoretical resources, and accepted inference bridges across gaps in evidence, while partly leaning for credibility upon commitments to adjacent bodies of knowledge which are similarly constructed.' (Wynne 1989, pp.33-34).

Confronted with such a complex network of 'combined social-cognitive commitments' (Wynne 1989, p.34), social processes may serve to limit the extent of these deconstructive tendencies. Some controversies may involve highly competing scientific cultures or disciplines. Hence the importance of legal settings, and the procedures through which expertise is selected, becomes a vital social factor in the

---

within an organisation was seen as potentially leading to social barriers which might impede the flow of information (Doak and Assimakopoulos 2007b).

constitution of expert authority. Scientific knowledge is 'intrinsically vulnerable to systematically applied scepticism' (Wynne 1989, p.38).

Jasanoff (1998) has also highlighted the issues apparent in the negotiation of the divide between law and science, in a way which builds on the work of Wynne (1989). Using examples of exchanges between expert witnesses and legal professionals, she describes the division in understandings and worldviews between the two parties. In the adversarial court system, the notion of 'credibility' is not decided against a set of pre-given criteria; instead it emerges from a heterogeneously expressed, overlapping set of understandings, which amount to 'the dense cross-hatchings of lay and expert, communal and esoteric, vulgar and initiated', and accordingly she considers the practices involved in recognising what constitutes 'scientific' or 'credible' testimony. Credibility is 'constituted in legal contexts over scientific evidence' (Jasanoff 1998, p.721). Jasanoff cites judges as examples of 'gatekeeper' figures, able to demarcate the distinction between experts and non-experts, and argues that their judgements may lead to the creation of recognisable hierarchies of expertise ('scientists' over 'technicians', 'treating physicians' over 'epidemiologists' etc); they may also reject the testimony of one expert in favour of another, appoint their own neutral experts, or implicitly incorporate their own understandings of science. Whatever practices they employ, judges 'not so much *find* as actively participate in *creating* the dividing line between appropriate and inappropriate offers of expertise' (Jasanoff 1998, p.722, original emphasis).

Furthermore, Jasanoff (1998) argues that the constitution of credibility in such a manner has important consequences for individuals' powers of perception. She argues that 'lay intuitions and perceptions of the world, founded upon direct, unmediated witnessing, continually bump up against professionally configured claims of 'virtual' or expert vision (Jasanoff 1998, p.731). Jasanoff describes how the presentations of 'expert' witnesses may involve the use of technical means of depicting evidence, followed by interpretation by a designated 'expert', and provides the example of how the prosecution in the O.J. Simpson murder trial insisted on the need to exclude videotape footage of LAPD investigators, arguing that such footage required specialist knowledge to make sense of it[7]. This argument was rejected by the judge, who favoured the defence argument that the nature of such footage did not require expert interpretation. In this way, the judge ruled what was amenable to 'lay' knowledge over 'expertise'. Jasanoff does not indicate that a simple dichotomous hierarchy holds, but rather that these ways of seeing overlap in complex ways and it requires some form of authority figure, namely the judge, to act as an arbiter (albeit in ways that are dependent on legal practices and rules of general application which 'shape the overall context in which experts testify and may deprive some would-be experts of the opportunity to participate'). (Jasanoff 1998, p.732). Importantly, however, the ultimate issue of who is bestowed with perceptual authority in individual cases cannot be explained with any systematic recourse to rule-following. Jasanoff's heterogeneous set of examples show how 'expertise- contrary to what the law may doctrinally suppose – is constituted or reconstituted to some extent within the framework of the trial itself.' (Jasanoff 1998, p.734).

The legal arena is certainly one in which the makeup of science may be seen to be particularly malleable. Edmond (2000) has examined the ways in which judges make decisions regarding the epistemological status of scientific evidence (Edmond 2000). Using examples taken from English and Australian courts, as well as the US, he argues that decisions concerning admissibility are not achieved on the basis of any internal standards locatable within the realm of science itself, and nor do they correspond directly to external tests which may be imposed on courts to assess admissibility. Instead, Edmond argues that judges interpretations of the admissibility of scientific evidence is extremely flexible and can be seen largely as a matter of personal fiat. Judges use a range of criteria, which may reflect their own subjective opinions on what constitutes norms such as 'scientific method'. However, Edmond contends that the relationship between science and law is never unproblematic, for certain other considerations have to be taken into account, such as the need to balance the fact-finding process with wider social values and conceptions of justice. Difficulties may also arise involving the weighing of expert testimony against potentially conflicting

7 The prosecution argued that the lay jurors watching the footage would be misled into thinking that the investigators appeared to be on top of each other, when in fact they had maintained a proper distance in accordance with procedures (Jasanoff 1998, p.727).

lay eyewitness testimony. Such factors may well play on the mind of judges when defining 'science' and determining admissibility (Edmond 2000, p.251).

Edmond (2000) brings to attention the fact that decisions regarding the scientificity of evidence do not occur in a vacuum, and shows how considerations of scientific evidence are intertwined with wider societal concerns. Saks and Faigman (2008) however highlight the extent to which ostensibly 'scientific' evidence has often managed to escape what may be construed as more rigorous scrutiny. They argue that the scientific status of a great many widely used forensic procedures is baseless (Saks and Faigman 2008). Forensic identification techniques, such as the analysis of handwriting, bitemarks, toolmarks, tyre prints, and notably fingerprints, are described by Saks and Faigman as 'nonscience forensic sciences' (Saks and Faigman 2008).[8] This term is also used by the authors to describe other areas such as fire investigation, gunshot residue analysis and aspects of forensic pathology. Saks and Faigman give a number of reasons to justify the use of this term. First, they claim that these techniques were largely developed in police environments, and escaped the purview of institutionalised science departments located within universities, where, the authors argue, they would have been scrutinised for their adherence to well-established scientific norms and standards. More fundamentally, Saks and Faigman attack the way in which these kind of forensic techniques are designed to attempt to identify and individualise offenders.[9] The kind of categoric claims that are made via application of these techniques is argued to be antithetical to conventional scientific practice, in which truth-claims are regarded as conditional.

The situation is seen to have been compounded by the apparent incompatibilities between legal culture and the scientific domain, and by the inappropriate use of admissibility criteria for scientific evidence. Saks and Faigman argue that admissibility standards of scientific evidence in US courts, supposedly introduced by the 1993 *Daubert* ruling have been subject to flexible interpretation by judges. *Daubert* obliged the courts to consider a number of factors in determining whether proffered evidence could be considered 'scientific' or not. These factors encompass a

[8] These methods have alternately been referred to as 'soft' forensic sciences by National Institute of Justice (2007) (McClure 2007, p.4).
[9] Referred to in a critique by Broeders as the 'positivity doctrine' (Broeders 2005, 2006).

number of epistemological and methodological criteria (e.g. whether the evidence adheres to the Popperian concept of falsifiability, whether 'the theory or technique has been subjected to peer review', etc). This represents a somewhat more rigorous view than its predecessor, the 1923 *Frye* ruling, which merely required that scientific evidence had to have gained 'general acceptance' in the 'particular field to which it belongs'.[10] However, Saks and Faigman also suggest that the *Daubert* ruling, which had been seen as providing a more universal test of scientific admissibility, has often been ignored in the case of nonscience forensic science. They describe a paradoxical situation in which evidence, if regarded as lacking scientific validity, may become admissible, as this lack of scientific validity means it is not considered relevant for *Daubert* analysis (Saks and Faigman 2008, p.163-164). In the US, it is still the province of the judge to determine what evidence may require a *Daubert* hearing; however this paradox brings to light the fact that it is up to the judge to determine what, if anything, is 'scientific'.

Saks & Faigman add support the view that the boundaries that determine what is, and what isn't, accepted as admissible scientific evidence in court are fluid and perhaps not reflective of a commitment to philosophical strictures, or to what conventionally may be considered as appropriate scientific criteria. In this regard, they indicate another form of boundary work in which notions of 'science' are dependent on institutional context. Along with Edmond, they also show how decisions regarding scientific admissibility are often a matter of personal fiat with regard to the judge, who may be seen to act as a pivotal 'gatekeeper' figure.

Sociological studies have therefore emphasised the complications apparent in the demarcation of 'scientific' and 'expert' knowledge, particularly in legal settings. A certain number of studies have sought to argue that the creation of expert knowledges involves a highly localised and contextualised set of performances. This work has emanated from the ethnomethodological school of science studies. In what follows I introduce some examples of this work which has explored the relationship between science and law.

---

[10] Leiter (1997) claims that *Frye actually* represents a far more realistic epistemology of accessibility than *Daubert*, taking into account as it does the epistemic limits of the courtroom setting. As well as potentially offering a false view of 'science', *Daubert* places unrealistic demands on the finders of fact.

2.5 Ethnomethodology and the Construction of Expert Testimony

The mode of inquiry referred to in broad terms as *ethnomethodology*, was introduced in the late 1950s, primarily through the work of Harold Garfinkel (Garfinkel 1967; Lynch 1993). With regard to the production of scientific knowledge, ethnomethodological studies have largely involved descriptions of the micro-contextual work involved in defining a series of boundaries relating to intellectual labour. Ethnomethodologists can be seen to have elaborated upon the original notion of boundary work by describing the contextualised practices at play in delineating 'scientific' from 'common sense' reasoning, and 'expert' from 'non-expert' identities.

Garfinkel claimed that he decided upon the term ethnomethodology whilst preparing reports for multidisciplinary study of jury deliberations at the University of Chicago (Lynch 1993, p.4). He became interested in how juror's pursued 'some kind of knowledge of the way in which the organised affairs of the society operated' (quoted in Lynch 1993, p.4). Garfinkel recounted how they appeared to be reflexively concerned with certain relevant social postures, and talked of 'wanting to be legal' and 'of being legal'. However, when pressed on what they meant by these terms, the jurors claimed how they could not define what it meant to be 'legal' without first *actually* being a lawyer. Using this example, Garfinkel claimed it showed an awareness of the need to fix the meaning of the kind of terms used in everyday conduct. The jurors conducted themselves as practical reasoners, with no credentials or professional expertise for collecting and assessing evidence, conveying an argument, or making judgements (Lynch 1993, p.4). The jurors discussions over how to comprehend, and effectively taxonomise, 'facts' 'reasons', 'evidence' etc represented questions of social scientific interest. However the commonsensical manner in which they went about deliberating over these concepts did not reflect the methods of the sciences. Garfinkel saw their methods for making sense, and defining such terms, as phenomena to be studied in their own right (Lynch 1993, p.4).

Thus the manner in which agreement emerges is itself taken to be a rich source of study. A heterogeneous series of practices and representations, carried out between actors, are involved in forming, co-ordinating and reinforcing intersubjective understandings. Ethnomethodology endorses a mode of inquiry that focuses on the construction of intersubjective understandings as they occur, and without recourse to preconceived, idealised notions as to precisely what these understandings constitute. Building on the example recounted by Garfinkel, Lynch (1993) views *reflexivity* as constituting a key feature of ethnomethodology. Added to this he cites the importance of *accountability*, namely the way in which ethnomethodology studies the everyday methods for rendering activities rational and reportable for all practical purposes. Social activities should be *orderly, observable, ordinary* and *oriented* (e.g. participants orient to the sense of one another's activities, and contribute to the temporal development of those activities. They should also be *rational* (they make sense to those who know how to produce and appreciate them), and *describable.*[11] Ethnomethodological studies have been applied to a variety of contexts, with a notable amount of attention paid to the manner in which credibility may be constructed in public settings such as public hearings and courtroom proceedings. In their novel study of the controversy over the Iran-Contra affair, Lynch and Bogen (1997) draw attention to the ways in which credibility was constructed in the context of the hearings which attempted to investigate the affair. In particular, they show how Lieutenant-Colonel Oliver North was able to subvert the question-and-answer format of the cross-examination process via the employment of various speech acts and performances which undermined the concept of the hearing as a 'truth-finding engine' (Lynch and Bogen 1997, p.122). Through strategies such as answering questions with political speeches, or by querying the questions of his interrogators, North was able to skilfully deconstruct the process of the hearing as it occurred, and maintain his own standing at the expense of the investigative process.

Whilst not concerned with science as such, Lynch and Bogen's study sheds light on the public hearing as an important arena for the construction of credibility. In the case of forensic science, it can be seen that the setting of the courtroom hearing functions as a key site in which credible 'expert' testimony is constructed. Lynch (2004) studied how terms such as 'scientist' or 'expert' are instances of 'membership categories' (Lynch 2004). Such categories may be used to classify individuals or objects, but

[11] Lynch also describes *indexicality*, but claims that it generally encompasses a way of thinking about the whole field of language use which ethnomethodologists investigate. As Lynch puts it, 'indexicality is a ticket that allows entry into the ethnomethodological theatre, and it is torn up as soon as one crosses the threshold.' (Lynch 1993, p.18). For a more detailed account of the origins of the concept of indexicality, see Lynch (1993), pp.18-22.

more importantly still, they may be used by an incumbent in reference to themselves and their activities (Lynch 2004, p.165). With regard to terms such as 'expert', it is obvious how incumbents may refer to them in order to support claims to credibility, particularly in certain contexts such as courtrooms. However, whilst membership categories may bestow high status to the incumbent, they are also prefigured by a series of rights, obligations, credentials and tests of worthiness (Lynch 2004, p.166). The crux of Lynch's description of how the instance of membership categories work centres on the essential role of settings, such as courtrooms, as sites where such identities are locally constructed and contested. Membership categories are assigned to individuals through a series of performances and interactions located within specific arenas. Lynch uses examples drawn from the UK case *R. v Deen*, in which the role of DNA evidence featured prominently. Here, he attempts to show precisely how the 'expert' identity of a Professor Peter Donnelly, a statistician specialising in issues relating to population genetics who was appearing as a defence witness was constructed and contested in and through a dialogue with the defence counsel. Lynch first shows how Donnelly's credentials were relayed in sequence, starting from his academic qualifications, charting his membership of learned societies, and his research achievements (Lynch 2004, pp.170-171). However, he also demonstrates the fluid nature of 'expert' identity via Donnelly's cross-examination in which prosecution challenged whether Donnelly's particular disciplinary type of 'expertise' was relevant to the issue, and questioned whether the abstract nature of statistics was to be accorded standing relative to the practical experience of the forensic scientists who carried out the DNA tests. Lynch's analysis highlights the ambiguities inherent in conceptions of 'expert' knowledge, and how it is ultimately dependent upon the decision of external actors such as judges and juries.

Lynch and McNally (2003), in their history of the appeal hearings of the *R. v Dennis John Adams* case, have shown how courts maintained a distinction between 'scientific' and 'common sense' reasoning (Lynch and McNally 2003). The prosecution case against Adams, who was convicted of rape, rested largely on DNA evidence, which supposedly indicated that the probability of the DNA matching a random member of the population was one in two hundred million. However, this was the only significant piece of evidence put forward by the prosecution, and was opposed by strong evidence for the defence. Adams' girlfriend had provided an alibi, and the victim had also failed to identify him in a police lineup. During the appeal hearings the defence used a novel strategy, which involved the explicit use of Bayes Theorem. This technique was employed in order to attempt to demonstrate that the prosecution probability calculation had failed to take into account every piece of relevant information. Bayes Theorem was used as it was viewed as a method of incorporating this information to produce a figure which the defence hoped to show that the prosecution had wildly overestimated their claim. Bayes Theorem was employed by the defence to test the juror's own subjective estimates of probability, and to show how, if all the relevant information pertaining to the case was taken into account, the random match probability estimate actually increased to a point well beyond reasonable doubt. The implementation of such a strategy required members of the jury to reach their figure in a manner redolent of a mathematical exercise, by working through the problem individually and making the calculations themselves.

The approach was rejected in the course of two appeal court hearings. In the first appeal, the court opined that, in using Bayes Theory in such a way, expert testimony had encroached upon the jury's role as the trier of fact. Whilst expert evidence may be used in an advisory capacity, the testimony produced in this case was seen as inappropriately pronouncing upon the issue of guilt or innocence (Lynch and McNally 2003, p.94). Furthermore, according to the court, the apparent objectivity of Bayes Theory masked the 'element of judgement on which it entirely depends' (1st Court of Appeal, *R. vs* [Dennis John] *Adams*, cited in Lynch and McNally 2003, p.94). The second appeal also rejected the defence case on similar grounds, and in it, the judges expressed misgivings about such an attempt to 'attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial' (2nd Court of Appeal, *R. v.* [Dennis John] *Adams*, cited in Lynch and McNally 2003, p.96).

In describing the case, Lynch and McNally argue that the courts produced their own definitions with regard to 'scientific' testimony. The employment of mathematical calculations was viewed as entirely suitable when used in conjunction with ostensibly 'scientific' techniques such as DNA profiling, but not in relation to other forms of evidence, upon which the Bayesian technique was thought to impinge. Furthermore, the decisions reinforced the norm associated with the collective reasoning of the jury, re-drawing the boundary of legitimacy for 'common sense' reasoning. The individualistic nature of the Bayesian technique was ruled as antithetical to the traditions of the jury system. Lynch and McNally quote the memorably expressed view of one judge who stated: 'consider your verdict amongst yourselves, all of you together and not with one huddled in a corner with his calculator'. (quoted in Lynch and McNally 2003, p.95).

Some other studies have studied on the construction of forensic science outside the courtroom. In his study of forensic crime scene examiners, Williams (2007) emphasises the role of bodily techniques, particularly visual perception, alongside technologies which make use and build upon those capacities to construct forensic artefacts. Williams (2007) criticises arguments taken from the forensic science literature, which adopt idealised depictions of 'science' to account for the character of criminal investigation. Instead he argues that greater insight can be gained by studying the actual practices, and accountable conduct, of crime scene examiners. Williams argues that certain 'incorrigibles' such as Locards principle ('every contact leaves a trace') do not have any clearly defined epistemological basis, but still function as an instrumental principle guiding the conduct of the crime scene examiners. The resulting artefacts, rendered via the embodied skill of the crime scene examiners, function as mediating objects which facilitate understanding to others and aid reconstructive reasoning. The ensuing success or failure of an investigation is then dependent upon a series of standardised products and procedures which lead to the construction of these artefacts.

2.6 Conclusion

Throughout this chapter I have drawn selectively from existing literature to outline several approaches to the analysis of forensic scientific techniques such as DNA profiling, the assembly and use of 'expert' testimony in the courtroom, and the production of forensic evidence at crime scenes. In doing so, I have also outlined the theoretical basis of these studies. These constructionist accounts highlight the manner in which evidence, and understandings of it, can be seen to be 'co-produced' (Jasanoff 2004) in that 'scientific' products arise in a manner interdependent with the social orderings and orientations – in this case largely legal ones - which create and recognise them. Approaches differ, however, on the precise ontological basis upon which such co-production occurs. ANT argues that scientific endeavour is dependent on the way in which its components are arranged and positioned. Be they human actors or inert objects, these components are accorded equal ontological status in networks. Whilst retaining broadly similar ontological assumptions, performativity studies emphasise the role of language in the array of interdependencies intertwined within the assemblages through which science, and indeed, all social action, are meant to proceed. Ethnomethodology follows a slightly different line, arguing that 'scientific' products, and understandings of them, emerge from a primordial milieu of uncertainty, and are formed and clarified via human practices and performances.

All of these accounts therefore posit a mutual dependence between 'scientific' objects, and the means through which they come to be recognised as authoritative. In doing so they provide frameworks for explaining how intersubjective understandings of science arise. Yet it must be questioned to what extent these approaches are able to provide a fully comprehensive picture of scientific labour. Many studies have focused on certain ontological questions, concerning what might exist to create these understandings. However, it is not clear how these studies define 'understanding' itself. Part of the concern here is whether these approaches put too much stock in emphasising certain ontological assumptions in the course of attempting to explain how 'scientific' entities are recognised, and less on what it means to reason in a 'scientific', or 'expert' way. A range of epistemic activities are involved in forensic science, such as interpretation, assessment, evaluation, justification etc., yet it is unclear to what extent STS approaches are able to recognise these different activities and account for them.

This is particularly evident in the case of ANT. For example, Latour (1990) argues that 'inscriptions' are the exclusive means by which knowledge is comprehended and disseminated amongst human actors:

> 'We are so used to this world of print and images, that we can hardly think of what it is to know something without indexes, bibliographies, dictionaries, papers with references, tables, columns, photographs, peaks, spots, bands' (Latour 1990, p.36).

Drawing upon the work of Piaget, who described an experiment involving children measuring volume through the use of differentially sized beakers, Latour contends that all human cognition is instrumentally mediated. Original thought only comes about via the manipulation and re-organisation of material resources and the related inscriptions:

> '...most of what we grant a priori to "higher cognitive functions" might be concrete tasks done with new calibrated, graduated, and written objects' (Latour 1990, p.51).

Latour posits a world where knowledge is projected onto an unreflective social actor via a series of inscriptions, 'immutable mobiles' able to be modified, recombined, reproduced, superimposed, integrated into written text, or used to convert three-dimensional space into quantitative data. There is little room in Latour's schema for independent human thought; from his account, the pursuer of scientific knowledge is portrayed as a blank slate, a 'Lockean tabula rasa in Foucauldian garb' (Benhabib 1994 quoted in Wight 1999, p.130). With scientific information conveyed in such a manner, there is no scope for a differentiated consideration of reasoning activities, for it appears that, according to Latour's schema, these do not really exist.

Ethnomethodology, on the other hand, could be seen to take a different approach to epistemological matters, via the aim to describe the practices, and hence processes, through which actors come to generate understandings of what is involved in recognising scientific activity as such, in and through the course of scientific activity itself. Approaches taken by ethnomethodology studies have, however, generally involved a distanced stance in relation to the individuals and practices being studied, either through direct observation or via the use of recordings or transcripts. Ethnomethodology appears to construe the reasoning process as being determined by a combination of language and performance. Whilst this has highlighted the socialised nature of the manner in which understanding is reached, it may be the case that such an approach does not alone capture the whole process. For example, there appears to be little scope within ethnomethodology to study the 'internalisation' of understandings, and how the process through which this may occur. This is important, as the manner in which collective understandings become translated into personalised ones may have consequences for repeated, or similar, future instances of inquiring activity. The question here is to what extent purely intersubjective understandings can exist, or whether hermeneutic gaps between collective knowledge and personal meanings linger in attempts to reach shared understandings. Intersubjective understandings may not arise immediately; they make take time, or involve processes of mediation. How do personalised interpretations therefore become translated and incorporated into collective knowledges?

CHAPTER THREE - METHODOLOGICAL ISSUES AND RESOURCES

3.1 Introduction

The theoretical issues pursued in the course of the last chapter are further explored in the course of this thesis, which focuses upon the increasing application of statistical and probabilistic methods in forensic science for the interpretation and evaluation of evidence. In particular, the mode of probability based around Bayes Theorem has aroused both interest and, as I will subsequently describe, controversy. Bayes Theorem is used to assess personal measures of belief under conditions of uncertain knowledge, via a probabilistic formula. It is used for conveying beliefs as quantitative, probabilistic depictions, and as a framework for updating these belief-measures in the light of incoming information. Bayes Theorem is already used widely to calculate the weight of evidence of DNA profiles, and is being applied to the analysis of other forms of evidence.

Bayes can be viewed as possessing 'objective' and 'subjective' elements. It can be considered an 'objective' form, in that it is a quantitative mathematical construct, but at the same time it also possesses a subjective facet, in that it measures personal estimates of probability. This tension within Bayes, as an 'objective' measure of subjective experience, raises issues for the way in which theory and practice in forensic science are construed. Whilst Bayes may provide some form of theoretical framework, operating it to assess subjective measures of belief is practice-led. The precise practices from where these belief-measures may arise, and the role of Bayes in shaping them, is investigated in this research. This combination of the objective and the subjective presents a challenge to existing STS positions in the unpacking of these two aspects. What relationship do 'subjective' and 'objective' modes of belief hold in the operation of Bayes Theorem? Furthermore, can a mathematical form which encompasses objective and subjective aspects be regarded as a technology? If so, precisely what kind of technology does it represent? How is it used to facilitate intersubjective understandings of evidence? Crucially, what does the use of this technology mean for the way in which forensic investigation is practiced? Does it exert any influence on these supposedly 'expert' behaviours?

Although previous studies have concentrated on practices relating to the genesis of evidence in a variety of settings, relatively little work has focused on the uses of this evidence, or the manner in which the practices of evidence interpretation are managed and organised. The way in which evidence informs the progress of criminal investigations is an area which demands further sociological attention. The process of evidence interpretation begins right from the initial apprehension of the scene, with items of evidential interest being singled out for further consideration. Forensic evidence is subject to a process of identification, recovery, and subjected to transfer procedures before being interpreted and evaluated by forensic scientists working alongside police officers. Their findings in turn can influence the considerations of investigating police officers, who may then use evidence to inform the identification and questioning of possible suspects. Where it is felt appropriate, this evidence will play a vital role in the decision to charge a suspect, and subsequently this evidence will be conveyed in court to advance a prosecution argument, where it is also open to scrutiny by the defence.

Evidence may take on a plurality of forms: it may take on an inert material form, or it may come in the form of eyewitness testimony, information about individual character, or psychological profiles. It may be derived from a variety of material sources, such as fingerprints, footprints, toolmarks etc.; it may be used in various ways, used either in a directly investigative or prosecutorial manner, or to perform an ancillary function, existing in a relationship which informs the validity and reliability of another piece of evidence. Hence a whole range of items and procedures may be utilised in the course of a criminal investigation. Evidence may exist in a complex, and potentially problematic, series of inter-relationships; pieces of evidence may contradict each other, or may be recovered in a manner which enables only vague or ambiguous information to be derived from it. Whatever the precise nature of the crime, or the investigation, evidence interpretation can potentially involve a complex series of ratiocinations, encompassing a high degree of interdependent elements.

The interpretation and evaluation of evidence raises important issues however, as they represent the key processes around which understandings of evidence arise. Although pieces of evidence may be viewed as 'boundary objects' (Star and Griessemer 1989), it is not always clear in those kinds of studies by precisely what process mutual understandings arise, nor is it always clear why or how particular objects become rendered as boundary objects over others. The work outlined in this study is intended to advance previous approaches by shedding more light on the complexity involved in forensic investigation of evidence away from the courtroom. My intention, however, is not to examine technologies of evidence *production*, but instead to investigate the technologies of reasoning which facilitate evidence *interpretation*, a phenomenon, which has, so far, received relatively little attention from an STS perspective. The practice of interpretation seems to be something of a problematic area in STS. The view of it seems to be strongly linked with a certain Wittgensteinian notion, in which interpretation can only occur through some form of collective action, and is exclusively intersubjective in nature. In this case however I did not intend to enter into this research with too much in the way of assumptions as to what particular processes of 'interpretation' could entail. Bayes was treated as a locus around which interpretation of evidence was organised and shaped; yet the particular epithet 'Bayes' was also viewed as potentially concealing a highly heterogeneous picture, in which highly personalised contributions, e.g. those based on subjective experience could continue to play a part. One aim was to consider precisely how these contributions are used, and how experiences might be re-formed, in the context of Bayesian reasoning.

Whilst this study has acknowledged certain STS approaches when considering the topic, I have sought not to be too heavily influenced when considering my own methodological approach. A study of Bayes presents certain challenges. As with a large amount of research which involves a qualitative component (Bryman 2004, p.289), the questions of interest have tended to emerge out of an engagement with the topic itself. Whilst some initial research questions were posed however, it wasn't until further research was carried out that the study of Bayes was found to be so rich and potentially complex. Hence the research agenda has evolved as the study has progressed.

It must also be made clear that this study follows certain STS approaches in taking a sceptical approach to formalised methodological forms. I accept the position, common to ANT, performative and ethnomethodological accounts, of the reluctance to pin faith on the ability of more formalised methodologies to garner insight in an area which has so far evaded detailed exploration. Instead of accepting the claims made by certain forms of social theory, I adopt a stance which anticipates the complexities presented by the topic of interest. Given the complications and the untidiness associated with this domain, forensic science demands an approach which reflects a 'sensibility to the messy practices of relationality and materiality of the world' (Law 2007). Hence this study exercises caution towards supporting any grand social theoretic postulates, and, like Lynch and Bogen (1997), leaves 'no room for the special epistemic privileges often assigned to the use of social-science methods and theories' (Lynch and Bogen 1997, p.265). This study is intended to be interdisciplinary in nature, with science being apprehended with a sceptical but informed gaze; in order to propagate such an interdisciplinarity it was deemed appropriate to extend this to the more all-encompassing claims of certain social scientific theories and their apparent 'craving for generality' (Wittgenstein quoted in Lynch and Bogen 1997, p.270)[12].

The approaches discussed in the early part of this chapter act as starting points for inquiry, rather than representing full guidelines that have shaped the design of this study. Nonetheless, the concept of performativity has provided one influence for this work, insofar that the application of a probabilistic concept such as Bayes to a practical field such as forensic science represents an apparently clear instance of the usefulness of this concept. In this regard, part of the rationale for this study is assess to what extent the concept of performativity can explain the way in which Bayes is utilised. At the same time however, this project has kept open the possibility that certain areas may require other conceptual tools to account for the observed results

[12] 'Our craving for generality has another main source: our preoccupation with the method of science. I mean the method of reducing the explanation of natural phenomena to the smallest possible number of primitive natural laws; and mathematics, of unifying the treatment of different topics by using a generalisation. Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics, and leads the philosopher into complete darkness. I want to say here that it can never be our job to reduce anything, or to explain anything. Philosophy really *is* purely 'descriptive' (Wittgenstein, quoted in Lynch and Bogen 1997, p.269-270).

3.2 Studying Bayes

Bayes has, so far, only been the subject of one published study by STS scholars (Lynch and McNally 2003). Although Lynch and McNally's article raises some extremely valuable insights, I was concerned that it did not capture the rich series of practices involved in using Bayes in the course of forensic scientific investigation away from the courtroom. Accordingly, this thesis should be understood as a preliminary step in inquiring about the relationship between theory and practice in forensic scientific investigation, rather than being concerned to analyse the presentation and interrogation of evidence in court. The initial stage of the work therefore involved a high degree of fact-finding and familiarisation with the basic concepts of Bayes and the way it was portrayed as being of use to forensic investigation. At this stage the technical literature was the major object of a familiarisation process. Making initial sense of the precise nature of the issues of interest became a complex task. This was aided however, by a small number of interviews with forensic scientists and other figures. It must be added here that none of the discussions directly led to respondents suggesting areas of study; most of the individuals had little knowledge of STS, and did not press me to pursue specific paths. I was thus able to maintain a balance, between being engaged, informed and interested in the work of forensic science, but remaining distanced and impartial. Over time, informed by both the technical literature and the discussions that I experienced and also witnessed, a tangible series of research questions began to emerge. In some ways the research followed a path typical of other qualitative studies, where interviews were carried out, the results reviewed, with the consequence of informing future interviews, and in doing so, helping guide the progress of the research. Given the lack of previous studies in this area, this was deemed an appropriate strategy to follow.

In addition, certain events took place during the period of my study which have also driven the direction of the research. In particular, the outcome of the Omagh bomb trial, and the publication of the Caddy report into Low Template Number profiling, have raised considerable controversy both within and outside the UK forensic scientific community, and continue to resonate in laboratories and courtrooms. I outline the details of such controversies in a later chapter, but suffice to say here that such developments emphasised both the manner in which Bayes could be construed as a controversial and contested technology, and also indicated that this research could address certain issues which underpin these more high-profile controversies.

Whilst STS has made a major contribution in its interrogation of the theory-practice divide, the fact remains that there is still a perceived distinction in forensic science between the two, especially with regard to evidence interpretation. That does not mean however, that forensic scientists are not keen to bridge that divide; one of the main aims of this thesis is to gain further understanding of how forensic scientists apply supposedly theoretical constructs such as Bayes Theorem in the course of their work. Therefore, the sociological problem at issue here is only partially one of how actors come to recognise evidence. What is of further interest is the way in which actors use specific theoretical constructs, and the uses made of these constructs in the practice of interpreting evidence.

What I aim to show is that, although theoretical constructs such as Bayes Theorem may play some form of guiding role in modern forensic investigation, the latter alone only play a partial role in actual interventions of evidence interpretation. An understanding of the role Bayes Theorem plays in forensic investigation is not complete without a consideration of the practices through which it is interpreted and applied. In doing so, I hope to make a notable contribution, by demonstrating the usefulness of a sociological perspective on some of the problems faced by scholars of evidence.

However, I also still ultimately aim to show that attempts to introduce a secure fundamental epistemological basis for forensic investigation give insufficient attention to the ways in which reasoning processes themselves are socially constructed. I emphasise the possibility that reasoning systems may evolve, or at least fluidly adapt. I discuss this further in my final remarks on Bayesianism and abduction. It is equally important to consider the possibility that not only do systems of reasoning evolve, but so too does the nature of the milieu in which the reasoning takes place – thus we need to consider whether the role of the forensic scientist is changing, how and why it may be changing, and how this might be accommodated within the current UK legal system.

3.3 Approach of Thesis

In order to begin to understand the place Bayes Theorem occupies in forensic science, and to start to address the possibility that the shaping of 'Bayesian reasoning' is a contingent phenomenon, it is first necessary to examine the way in which Bayes Theorem itself has evolved and changed over time. Whilst the equation which depicts Bayes Theorem has remained largely the same, ideas concerning the use of the Theorem, and what ontological assumptions it was meant to reflect, have been subject to a significant amount of interpretation. By understanding the changing interpretations of Bayes itself, it is possible to gain an idea how Bayes may be considered - not as a rigid mathematical construction, but as a more fluid form of technology, where the precise possibilities for actualisation and use may take a plurality of forms.

Hence I begin the study by focusing on the origins of Bayes theorem. Drawing upon both primary and secondary sources I show how the modern interpretation of Bayes differs from the version originally conceived by its eponymous progenitor. I do this in order to show how the context in which Bayes is used has affected representations of the theorem. I also show how Bayes has emerged out of a competing set of ideas about probability, and how they in turn have been affected by underlying ontological and metaphysical assumptions held by certain individuals. Both Bayesianism, and another mode of probability known as frequentism, emerged out of a debate which had, at its heart, issues relating to how the world was perceived to be ordered.

I introduce this work in order to show the usefulness of regarding Bayes, and probability as a whole, as technologies which are shaped as much by the contexts in which they are used, as much as they influence the behaviour of those who seek to use them. I carry this notion forward in Chapter 5. Here I draw upon a variety of primary and secondary sources to introduce attempts to adduce a greater degree of systematisation to the interpretation of evidence of relevance to criminal cases. Again my approach is historical, as I seek to trace the introduction of Bayes into legal and forensic scientific reasoning. In this chapter I therefore seek to show how Bayes, and the role of statistics in general, has come to be an important part of discussions in modern forensic science, and of a wider academic approaches to the issue of evidence.

I broadly use the same approach in Chapter 6, where I provide an introduction to the scientific issues that have been raised in the course of the utilisation of DNA profiling for forensic use. This serves a number of functions. First, this chapter provides an overview of the scientific basis of the production and interpretation of forensic DNA profiles. The latter has played a particularly important role in introducing the greater use of statistics and probability for the evaluation of forensic evidence, and this chapter also outlines the relationship between technology and statistics. More importantly however, it also shows how the scientific and statistical basis of DNA evidence has come to be comprehensively examined by the adversarial Anglo-American legal system. Hence I also explain, in detail, how legal deliberations have also shaped the development and use of statistical applications in forensic science.

I explore the way in which Bayes has been used in forensic science in more depth in Chapters 7 and 8. Here I present two case studies which comprehensively explore the use of Bayes in two separate contexts. The first concerns the development of automated systems for the interpretation of complex DNA profiles. This area was chosen since Bayes Theorem has formed a significant element in the generation of the algorithms which form the calculational basis of these systems. Thus, an opportunity presented itself to study the practices involved in the development and use of these systems. It was intended to investigate the extent to which the use of Bayes, and representations of it in scientific literature, were sufficient to project an image of the resultant technology as being ostensibly scientific. The aim of the case study then, was to highlight the amount of scientific labour involved in the creation of systems of evidence interpretation, involving as they do a series of mathematical and scientific technologies. Furthermore, this case study also sought to investigate the extent to which the deployment of Bayes could be seen to project an image of 'objectivity' to these technologies. Hence, the study sought to identify how areas of objectivity and subjectivity were defined and managed. Of particular interest in this case was the possibility that a number of areas still dependent on a relatively high degree of experiential judgement and assumption still existed, in a manner not alluded to in the scientific literature, or by forensic scientists themselves.

This case study focused on one relatively narrow area of forensic science. However, as recent cases have shown, these technologies have made an important contribution to criminal justice in the UK. Techniques such as Low Template Number DNA profiling, and Pendulum List Search, which utilise Bayesian reasoning as a calculational platform, have attracted considerable interest in the media, and have helped to reinforce an image of modern forensic science as an efficient 'hi-tech' mode of producing criminal justice. Such technologies have also made an important contribution to policing practices in facilitating the re-inquiry of so-called 'cold cases'. However, they have also come under increasing scrutiny in recent times, particularly with regard to LCN. The controversies which continue to dominate discussions in the UK forensic science community[13], have been brought to public attention via events such as the judicial criticism of LCN in the Omagh bomb trial. The study of Bayes in this context is therefore not a trivial matter, and a closer examination of the way in which Bayes has been used in the construction of these technologies brings to light the reasoning practices of the forensic scientists who developed them. An important point to consider is that these technologies have been developed by the FSS, now a Government-owned company who, despite a certain widening of the market, continue to dominate the provision of forensic science services in the UK. The work of their scientists is therefore highly influential with regard to both policing strategies and criminal justice, and hence their practices merit close study. Furthermore, the way in which the FSS communicates its work is also important. The growing marketisation of forensic science provision in the UK may play a role in the manner in which Bayes is depicted in FSS literature. Hence part of this case study has involved focusing on this relationship.

The second case study involved a more holistic look at the management of evidence, and centred on the Case Assessment and Interpretation (CAI) model, which was also originally developed on behalf of the FSS. The CAI was developed with the growing commercial nature of forensic science provision partly in mind, and was intended to consider how forensic scientific services could be delivered to police to ensure optimal cost-effectiveness. The CAI consists of a framework which draws heavily on Bayesian probability. It seeks to clarify and define a number of processes which the

[13] As discussed at a conference organised by the Forensic Science Society and attended by the author, 17 April 2008.

authors perceive as being instrumental to reasoning about evidence. A key aim of the CAI is to help guide forensic scientists in formulating rational propositions to test and explain the origin of evidence, which is viewed one of the central aspects of the Bayesian approach to forensic science. The CAI can thus be construed as a technology which provides a means for forensic scientists to clarify and organise their reasoning processes concerning criminal cases. Viewed in other way however, the CAI, and Bayesianism as a whole, can be perceived as making a significant contribution to the disciplinary identity for forensic science as a whole. The CAI framework is intended to be applied holistically to all forms of forensic evidence, and the language of the authors of the CAI, who suggest that Bayes is the '*only* logical way to reason about evidence', suggests a certain prescriptive and programmatic intent.

The CAI has featured prominently in forensic scientific literature, and the authors of the CAI continue to actively contribute to forensic science conferences, both in the UK and abroad. Their work has elicited a considerable amount of interest, and in some cases formal acclaim, but it also attracted criticism, and doubts have been expressed concerning the applicability of Bayesianism to casework. Hence the CAI provided a significant topic for the second case study. It also provides an interesting contrast to the first case study. Whilst the use of Bayes in automated DNA interpretation systems represents an application to a specific piece of evidence, the CAI represents a notable extension of Bayesian reasoning, being applied generically across a range of forms of evidence. Moreover, it became clear that the application of Bayes to the investigation of criminal cases as a whole presented a greater set of issues. With this came the realisation that, in the case of CAI, a diversity of practices could be seen as being linked to the actualisation of Bayes. A key difference in this second case study concerned the fact that a greater role was placed on human actors in performing Bayesian reasoning, in contrast to the first case study, where human actors merely designed the calculative processes. An opportunity arose therefore, to consider the role of embodiment in Bayesian reasoning. This, in turn, led to the consideration of the construction of different forms of Bayesian reasoning. This issue of differing 'forms of Bayes' and the role of human agency in constructing them, is a topic which I discuss further in Chapter 9.

3.4 Choice of Methods

This study used a mixture of methods. The consultation and analysis of a number of different documentary forms played an important role in this work. Although I outline below the various kinds of documents that were collected and consulted, I will briefly comment on the reasons why documentary analysis formed a prominent part of my approach, and then explain how this, in combination with other methods, comprised the methodological strategy for this study.

Firstly, the use of documentary analysis allowed the study to incorporate a historical dimension, the reasons for which I explain above. Given that a number of intertwining topics are discussed in this research (such as the origins and development of Bayesian theory, its use in judicial and forensic settings, the evolution of studies concerning evidence interpretation, the permeation of statistical concepts into forensic science, the development of DNA profiling and resultant challenges etc), it was necessary to provide overviews of a number of these issues in order to provide the appropriate historical background for understanding how Bayes has come to be applied to forensic science, and to highlight the issues such a development presents. This often involved the consideration of relatively long timespans; for example, the history of Bayes Theorem dates from 1763, and attempts to provide a systematic basis for evidence interpretation date from around the early 1900s. Hence documentary material was used to develop my accounts. Although secondary sources were used in order to verify some of the arguments, primary sources, in the form of relevant academic articles, are readily available.

Documentary material also played an important role in the cases studies. The kind of material consulted in these studies involved technical documents, most notably articles from scientific journals and related textbooks, to understand how Bayes is understood and applied by forensic scientists to produce technologies of reasoning better suited to the challenges presented in the course of forensic casework. Together these provided the most convenient and readily accessible guide to the scientific basis of the current work which forms the subject of the two case studies. A small number of other documents were also consulted in the course of the case studies. These involved official documents provided by police forces, court reports and news media. These documents provided some details of the impact of Bayesian forensic technologies on a series of activities, such as policing strategies, judicial decision-making and media reception.

The approach to the collection and analysis of documentary material broadly followed the model which is suggested by Altheide (2004). Preliminary research questions were considered which were based on an initial familiarisation with the area of study. This informed the creation of more specific areas for research, out of which arose the two case studies pursued in the thesis. These particular case studies were regarded as representing a pair of instances which were not only considered to possess research potential, but which had also attracted a significant degree of attention from within and outside the UK forensic science community. Hence both these case studies were regarded as not just major scientific issues, but distinctly social ones as well. Furthermore, the two studies were also regarded as sufficiently distinct, yet also displaying enough commonalities to facilitate a fruitful exercise in comparing and contrasting two instances of the application of Bayesian reasoning in evidence interpretation.

The documents consulted in the case studies may be regarded as largely official, or at least public, documents, and thus an awareness was constantly maintained that these were constructed texts (Abraham 1994). That is not to say however, that the information found within them has been uncritically absorbed, and a suitably sceptical stance was maintained. On the contrary, one of the key aims of the case studies was to critically examine the kinds of claims made in these texts, with a view to comparing the depictions of Bayesianism with 'real-world' experience. Here however, is where the documentary approach was found to possess limitations. Much of this technical material provided a theoretical account of Bayesianism, with little attention paid to specific instances of the issues encountered when utilising this technology in the course of actual casework, a crucial area of interest. Furthermore, it was only possible to gain a minimal understanding of the latter through some of the comments made in other literature such as judicial and news media reports. Hence other methods were considered necessary in order to gain a broader understanding of the issues involved. Originally, these were sought in order to gain a fuller idea of the process of using Bayes in forensic investigation. As the research progressed however, it also became more apparent, via media reports and other sources, that certain questions were being raised about the effectiveness and suitability of using such a technology in forensic science. Literature that addressed the actual experience of using Bayes in casework was not available in the public domain, and it was therefore decided to pursue this via a small number of semi-structured interviews.

The use of semi-structured interviews was deemed appropriate for a number of reasons. As discussed in further detail below, given the access issues, a full participant observation study was ruled out at an early stage. As the study was concerned with self-assignations of Bayes, it was necessary to allow interviewees to talk about how they themselves viewed Bayes, and hence the interview technique provided highly insightful. It enabled interviewees to talk directly about some of the processes related to criminal justice. This in turn enabled me to gain an insight into a wide variety of procedures and issues at various stages of the criminal justice process, and they made a significant contribution to the evolution of more specific research questions, in a manner typically encountered in the utilisation of qualitative methods (Silverman 1993, Foster 1995, Davies 1999). It is unlikely however, that I would have been able to gain access to such a variety of stages if a full participant observation approach had been adopted, but the interviews added a great deal of colour to the research. The capacity to gain such a broad understanding was considered necessary, given the position forensic science occupies between the realms of science and law. Efforts to incorporate Bayes have occurred both at the level of the courtroom and of the forensic science laboratory (of which I provide a further overview in Chapter 4), and any understanding of how Bayes is applied at the forensic scientific level must also consider possible impacts at other stages of the criminal justice process. I provide further details of this in chapter 6.

An ethnographic approach was considered less appropriate for this study, as there was less emphasis on the construction of a material forensic technology, and more on how a forensic technology was shaped, and how conceptions of a pre-existing term such as 'Bayes' were interpreted and represented. A small amount of participant observation was conducted via attendance at conferences organised by professional associations. Through these, it was possible to observe the deliberation of issues relating to Bayesian interpretation of evidence. Being a technology of reasoning however, it was decided that an investigation of the forms it took would need to occur with the cooperation of reasoners, whenever possible, in order to delineate how it might be enacted. Few *a priori* assumptions were made concerning how Bayes might come to be rendered as a tangible reasoning technology, and the possibility was left open from the start that in different contexts, different forms of Bayes might arise by different means. This precluded an especially rigid approach to methodology in general, but it also indicated there was more to be studied than by observation alone.

Even if an ethnographic technique had been adopted, it is likely that it would have run into the same problems experienced by Innes (2003), such as an unwillingness for investigators have 'on the record' comments attributed to them. Innes also notes how some of the most frank and most useful comments came in relatively less institutionalised contexts (Innes 2003, p.287). With the interviews that I conducted occurring away from similarly institutionalised and less pressurised environments, a similar level of relaxed honesty was encountered during the course of this work, and I also paid particular attention to honing the kind of techniques for good practice in interviewing as suggested by the likes of Kvale (1996) and Charmaz (2002).

Innes (2003) also highlights the difficulties experienced in gaining access to police organisations in order to perform ethnographic analysis of detectives at work (Innes 2003, p.284). This study had originally been envisaged as a comparative study involving two police organisations. Although Innes was granted access to one organisation however, he reports difficulty in gaining access to a second force; approaches to several other police organisations for access were rejected with the consequence that the focus of the study had to be adjusted in order to compensate for this problem (Innes 2003, p.284).

His experience of the difficulties in gaining access to law enforcement agencies involved in sensitive casework influenced the course of this study. For example, attempts were made to contact key actors within the Forensic Science Service (FSS) to interview them about their work on Automated DNA Evidence Interpretation Systems, yet no replies were forthcoming and hence no direct access could be gained. This is certainly in line with an established FSS reputation for longstanding concerns with security and confidentiality, along with a more recent concern with commercial confidentiality. In this case, the access issues experienced in this study were of a distinctly non-public nature (Lofland and Lofland 1995), than the more 'public', but perhaps no less challenging groups as encountered in other studies (Willis1977, Giulianotti 1995, Hobbs 1993, O'Reilly 2000). Although some studies on the police have shown it is possible to gain access on a covert manner to a closed public organisation (Rubinstein 1973, Punch 1979, Holdaway 1983), this was simply not an option for this study, given time and other constraints.

As I describe in Chapter 7, questions have been raised concerning the relative lack of information that has been released by the FSS with regard to the scientific technologies used by that organisation to interpret DNA profiles. In my research, one interviewee openly criticised the FSS for failing to provide suitable levels of information relevant to cases. The kinds of technologies that have been produced by the FSS were identified as an important topic of study in the course of this research, yet it was decided that given the lack of response from the organisation, any direct engagement with the actors would have proved to be extremely problematic.

3.5 Specific Methods and Materials

3.5.1 Interviews

This research did not involve a direct consideration of crime. Instead, the focus of the project largely centred around the actions of forensic scientists and the technologies they utilised in their reasoning processes. Nonetheless the nature of the research did mean that interviews often involved discussions of a relatively sensitive nature. The interviews often covered criminal cases, which in some instances were still ongoing. These instances did provide highly relevant information, and where relevant parts of these discussions are included in the ensuing case studies. In most cases however, interviewees were able to generalise to the extent that the precise details, e.g. names, locations and other such identifiable specifics were omitted without compromising the usefulness of the information which was divulged.

During the course of the research however it was possible to gain access to a number of key figures involved in the areas of forensic science of interest to this study. A total of 12 individuals were interviewed during the course of this work, with 1 Interviewee (Interviewee 5), being interviewed on two separate occasions. There was a certain amount of overlap of topics, such that discussions with some interviewees often resulted in topics relevant to both case studies being addressed. Of these interviews, all but 2 of the interviews took place directly in person, with the results being recorded and transcripted where the data was felt to be relevant to the study. The remaining 2 interviews involved individuals who were based in the USA, and took place by telephone, with notes being recorded.

Consultation of much of the literature provided a somewhat idealised view of Bayes which contrasted with some of the accounts given by interviewees. A more serious issue which arose from the research, and which acted as a justification for the interview approach, was the fact that the degree of awareness of Bayes varied greatly amongst individuals consulted. Whilst all the interviewees were aware of the term 'Bayes', the extent of the knowledge varied from wide-ranging experience of attempting to apply it to forensic science, to hardly further knowledge beyond the name itself. The lower limit of knowledge was often found with more junior practitioners, as those involved in the provision of forensic scientific training exhibited a relatively strong awareness of the potential uses of Bayes in forensic science. Furthermore, it was those practitioners who were largely involved in the collection of evidence (e.g. crime scene examiners), who demonstrated the least awareness of Bayes. This is perhaps not surprising as it is these individuals who are perhaps least likely to have been trained about Bayes; although some interviewees felt that a knowledge of Bayes might help scene examiners in knowing 'where to look' for relevant evidence in the context of an investigation, this opinion was not reciprocated.

3.5.2 Profiles of Interviewees

a. Interviewee 1: Had recently completed a PhD in the US on the application of Bayesian Networks to DNA profiling (parental analysis). At the time of the interview this interviewee had embarked on postdoctoral work at a leading UK university, which concerned the further development of Bayes Networks in order to apply them to DNA profiling, and had already published in the scientific literature at the time of the interview.

b. Interviewee 2: This interviewee holds two PhD qualifications in Mathematics and Computer Science, as well as a medical degree. He is based in the US where he is the CEO of a company which develops computer programs for the derivation of DNA profiles from mixed samples. He has also published work on this subject and submitted conference papers.

c. Interviewee 3: Ran a forensic science consultancy in the UK, generally specialising in defence work. He had a scientific background, and had previously been head of a police lab in a large British city. His current work had led him to testify in high-profile criminal court cases in the UK, and in one particular instance he had featured prominently in criticising Low Template Number (LCN) DNA. This interviewee also has a slightly more critical view of the application of Bayes to forensic science, and has written articles pursuing this line.

d. Interviewee 4. A professor of genetics in the US, he also runs a consultancy which advises defence briefs over issues associated with DNA evidence. He had also testified in a high-profile court case in the UK, where he had been critical of LCN evidence.

e. Interviewee 5: Had worked for the FSS for over thirty years before moving on to run his own consultancy business. He has been involved with the genesis of the CAI from it's onset, and continues to teach the application of Bayesianism to undergraduate students. He has published a number of articles and book chapters on the CAI and on the application of Bayesianism to forensic science as a whole.

f. Interviewee 6: Professor of Statistics at a British university, he has published several books and articles on the application of statistics and probability to forensic science. He has been conducting research on the development and application of Bayesian Networks to forensic scientific problems, from issues surrounding DNA evidence to their application to a whole criminal investigation.

g. and h. Interviewees 6A and 6B: Postdoctoral students of Interviewee 6.

i Interviewee 7: Worked in police before leaving to organise and teach forensic science courses to undergraduate and postgraduate students.

j. Interviewee 8: A Senior Lecturer of forensic science at a British university. Worked for the FSS for eighteen years, leaving as a senior scientific officer to become part of an FSS management team. After writing a widely circulated report, he helped found a new operational forensic scientific support resource for the police, and took part in around 230 reviews of murder investigations. Until he retired in 2005, he was a member of a number of ACPO committees and served as an advisor to the UK Parliamentary Select Committee Inquiry into forensic science. He has also acted as a consultant on a number of TV and film projects concerning forensic science.

l. Interviewee 9: Works as a researcher at the National Policing Improvement Agency (NPIA).

In most of the interviews, inquiries were pursued which attempted to cover both case studies. The lines of questioning are detailed in the more detailed consideration of the case studies below.

3.6 Case Study 1: Automated DNA Interpretation Systems

3.6.1 Documentary Examination

Various forms of literature were used in the course of this case study:

Scientific and technical literature

This literature encompassed books and chapters, articles published in academic journals, and conference papers and presentations. These were referred to in order to gain an understanding of a number of issues. First, to gain an understanding of the kind of theoretical approaches that have been applied to the design of probabilistic systems for DNA evidence interpretation. Second, to gain an understanding of the kind of claims made regarding these technologies. Third, to gain an understanding of the scientific concepts underpinning these systems. Journal articles were taken from a number of disciplinary areas, encompassing forensic science and legal medicine, statistics and probability, and genetics

Other articles and papers, often published in forensic science journals, or presented at conferences for forensic scientists, sought to comment upon these techniques, and aimed to highlight specific scientific and legal issues. Examples of such papers include Budowle et al (2001) and Butler (2006) These articles were consulted in order to gain an understanding of the controversies raised by DNA evidence interpretation technologies amongst the relevant scientific communities and individuals.

Although an awareness of the constructed nature of such documents was maintained, this material provided a wealth of information, and together constituted an important resource for understanding the fundamentals of Bayesian theory and how it has, (and is intended to be) used in forensic science.

Reports and Reviews

Prior to, and during the course of this study, a number of reports, commissioned either by government or private bodies, were published which included discussions of issues relating to the technologies of interest in this study. These also highlighted a number of related scientific, legal and ethical controversies. For example, Williams et al (2004), Genewatch (2005), and the Nuffield Council of Bioethics (2007) included discussions of a number of topics related to the technologies involved in the case study, amongst other considerations of a range of topics related to the forensic use of genetic material and bioinformation in general (Williams et al 2004; Genewatch 2006; Nuffield Council on Bioethics 2007). The report on the Omagh judgment (Weir 2007) featured a concerted discussion of LCN DNA (Caddy et al 2008). The Caddy (2008) report was dedicated to the issue of the scientific and legal validity of LCN DNA profiling (Caddy et al 2008). As I discuss in one of the case study chapters, the Caddy Report itself was the subject of a significant amount of controversy, and I include two responses to that report amongst this collection of data (Jamieson and Bader 2008, Gilder et al 2008).

Many of these kind of reports were more critical about the introduction of Bayesian technologies into forensic science, in particular the application to DNA profiling. Whilst these must also be regarded as constructed documents, this material added to the sense in which Bayesian technologies are contested sites, within the domains of ethics, law, and, importantly, science. Whilst these documents alone did not provide a comprehensive indication of these issues, they highlight a series of areas of dissensus which informed the study, and helped to guide the formation of questions for the semi-structured interviews.

Court Reports

Reports from UK Appeal Courts are freely available, and a number of these were consulted in the study, in order to assess the extent to which techniques such as LCN were utilised as evidence in court, the way in which they are used to incriminate suspects. They were also consulted in order to help gauge the extent to which technologies such as LCN feature in court cases, and to ascertain how such evidence is received in court.

Police and Government Literature

Material published by police and government bodies provided information about the role of DNA evidence interpretation technologies in the course of certain police operations, most notably cold case review operations such as Operation Advance and Operation Phoenix. This literature provided information about the precise aims, objectives and scope of the operations and how DNA interpretation technologies were to assist in them. The literature also provided some statistical information concerning: the number of re-opened cases, the number of cases in which LCN could be applied, number of arrests made, convictions secured etc. This literature helped to provide an understanding of how new Bayesian technologies have influenced a change in certain police practices with regard to serious crime, a development which has been influenced in part by central government policy.

Literature from Forensic Science Providers

Material taken from providers of forensic science products and services, such as the UK Forensic Science Service (FSS), was also consulted to assess the contribution made by DNA evidence interpretation technologies to the resolution of criminal investigations. This material took the form of website information, and 'fact sheets' which provided summaries of the technologies and their use in investigations.

News media

News media sources were also consulted as a further source of information, which included: further details of the circumstances of criminal investigations and their subsequent success/failure; notification of relevant published reports and their subsequent reception by relevant public figures; opinions expressed by public figures such as judges, police officers and politicians etc; actions and responses of governments and other bodies such as the Crown Prosecution Service, Association of Chief Police Officers (ACPO) etc.

e-Symposia

Prior to, and during this study a number of symposia have been held by the Forensic Institute, which have been conveyed via the Internet. E-symposia have been held annually since 2005 on the theme of Human Identification, and considerable focus has been given to the type of theoretical and technological approaches to DNA interpretation featured in this case study. The e-symposia have consisted of a series of presentations on legal, scientific and technological developments in this field, and also panel discussions on related issues. Participants have included a number of leading figures, many of whom have been directly involved in the development of Bayesian algorithms and subsequent interpretation technologies, and who have also published extensively on these areas. As well as being relayed live via the internet, the conference proceedings are also available in archived form for access to subscribers. Hence these have also been used in this study, and have been found to be another valuable resource for gaining an understanding of the scientific basis of the technologies, as well as the nature of the discussions and debates surrounding them.

Other Websites

The internet was also accessed on occasions, in order to search for other relevant material, using internet search engines. This returned a relatively limited amount of suitable material, although it did result in the discovery of online discussions of some aspects of the featured technologies. However, more notably this did result in the discovery that the FSS had employed a public relations firm to promote a particular product via news media. This is discussed in the case study.

Secondary Material

A number of transcripts which formed part of an early study of the history of DNA profiling. These also provided access to some of the thoughts of key actors concerning the use of statistics in forensic science, and have been incorporated into this research where it was deemed appropriate

<u>3.6.2 Fieldwork</u>

This comprised of:

1) semi-structured interviews,
2) attendance at conferences for forensic scientists

1) Semi-structured Interviews

The type of enquiry pursued in the course of these interviews can be differentiated into the following lines of questioning:

- Questions concerning the precise nature of the work involved in the development of Bayesian algorithms used in the systems for interpreting DNA profiles, and the subsequent process surrounding the implementation of these systems in casework. Hence themes in this area included: the lived experience of constructing the algorithms and finding solutions to design flaws etc; the nature of the skills, expertise and background knowledge required in this area; the type of actors involved in the creation of the systems and the forms of communication between them; how the input of these actors affected the subsequent design of the systems; challenges involved in using these systems in theoretical examples and making the subsequent transition to actual casework etc.

- Questions concerning the scientific and legal validity of such systems, including themes such as: identification of the nature of the scientific and legal issues associated with these technologies, including the recovery of evidence, and the generation and interpretation of resulting DNA profiles; identification of the stances associated with discussions over these issues, and how these positions were justified; the reception of these technologies by jurisdictions outside the UK; the experience of the use of these technologies to supply evidence in court cases; means of validating the systems etc.

- To what extent these technologies embodied the tenets of Bayesianism: how the data was shaped and interpreted to fit with the formulae; identification of the kind of approaches used to adapt Bayesianism to the problem of DNA evidence interpretation.

2) Conference Attendance

In addition to the interviews, it was also possible to access more candid discussions of the issues via attendance and involvement with a small number of conferences and seminars. These events functioned as important spaces where scientific issues were debated and discussed. In some cases the technologies under study came under a strong degree of scrutiny. Although it could not be said that technology was 'deconstructed' in the same way as Jasanoff and Lynch report the locale of the courtroom, these discussions served to open up many of the issues which were circumscribed in the published material. They also allowed an insight to be gained of the most up-to-date developments, as opposed to the literature in which a certain backlog in publishing was assumed in some cases. Two conferences were attended in person, plus a day-long seminar given at a university. In addition to this, four web-based seminars were accessed.

In addition to the interviews, I attended two conferences organised by the Forensic Science Society. These were held at British Universities over the course of 1-2 days each, with the first taking place in June 2007, and the second in April 2008. These were organised by the Forensic Science Society, a leading professional body which has members in over 60 countries (FSS 2008). These conferences attracted leading figures in the fields of DNA evidence and evidence interpretation, and the second conference placed particular focus on the issues surrounding LCN analysis.

The visits to these conferences had the following purposes: to help understand how those in the UK police and forensic science communities viewed DNA Evidence Interpretation Technologies in general; how these communities identified and framed the scientific and legal controversies associated with these technologies, but also how they identified related advantages; to keep updated with new scientific developments within the field.

3.7 Case Study 2: The Case Assessment and Interpretation (CAI) Model

3.7.1 Documentary Examination

Academic Articles

A series of articles have been published in forensic science journals which have specifically outlined the principles and key features of the CAI. These articles were used in addition to book chapters which also described features of the CAI. In addition, literature was also consulted which discussed the use of Bayes in forensic science and the law on a more general basis. These articles were located across a range of disciplinary areas, encompassing not only forensic science, but also legal studies, evidence scholarship, and statistics and probability.

3.7.2 Fieldwork

Fieldwork was undertaken which comprised of:

a) Semi-structured interviews

b) Attendance at a one-day seminar for undergraduate forensic science students at a British university, which focused on the use of Bayesian reasoning in criminal investigation. This seminar was a compulsory part of their degree course.

c) Attendance at conferences which focused on the application of probabilistic approaches such as Bayes, held in April 2008 (see previous section for details).

Interviews

All interviews were carried out in person. The following individuals were all interviewed once, with the exception of Interviewee 5, who was interviewed twice, and was also present for part of the interview with interviewee 7. Given this interviewees close relationship with the CAI, it was felt appropriate to interview him more than once in order to help gain greater insight into the CAI. The length of the interviews varied from approximately 45 minutes to approximately 2.5 hours.

During the course of these interviews, inquiry focused on the following themes:

- The relationship between forensic scientists and the police, and the relationship between forensic science providers and the police as separate organisations: determination of the possible changing nature of the type of products and services offered to police forces; expectations of forensic science on the part of the police; discussion of whether the roles of forensic scientists is changing, and whether the scope of 'forensic investigation' is widening;

- Attempts to apply the CAI to casework: the identification of specific issues experienced when attempting to apply the framework to casework; identification of the kind of cases to which the CAI has been applied; discussion of whether certain kinds of cases suit the use of the CAI in comparison to others; assessments of the suitability and potential of the CAI for casework use; questions concerning the existence and availability of data sources for use in CAI-based inquiries;

- Questions concerning the origins of the CAI: Key drivers toward the initial development of the CAI; the nature of theoretical work which influenced and shaped the development of the CAI; the individuals involved in the creation of the CAI, and the decision-making procedures that were employed;

- The manner in which the CAI has evolved and developed: how, and why, the CAI has evolved in a particular direction; how actual casework has influenced the design of the CAI; issues specific to particular forensic sub-disciplines, e.g. DNA, fingerprints, handwriting analysis etc; how operational forensic scientists have contributed to the development process etc

- The reception that the CAI has experienced: the nature of the response from operational forensic scientists, police officers, the FSS and the forensic science community as a whole; criticisms of the CAI and attempts to apply Bayesianism to forensic investigation in general; to what extent the CAI and Bayesianism is actually being used in the course of casework.

- The nature of the relationship between academic/theoretical research conducted in the domain of forensic science, and forensic scientific practice in the course of criminal investigations.

- Translating and conveying probabilistic data to lay audiences: Discussions of the problems associated with converting numerical data into a form more readily comprehendable to jurors, advocates judges etc.

Field Visit

In addition to the above, a two-day visit was undertaken to a large and prominent police force based in a vast metropolitan area. During this visit I was able to engage with a number of members of staff from all levels of the hierarchy. This ranged from custody officers, scientific support staff, procurers of forensic science, to the head of forensic sciences. The field work included visits to police stations and to forensic laboratories, as well as the headquarters of the force itself.

CHAPTER FOUR - THE ORIGINS OF BAYESIANISM

4.1 Introduction: Bayesianism Today

It is an appropriate starting point to this study to consider not only what is meant by 'Bayesianism' in modern forensic science, but also to understand how such a form of reasoning has to come to take its current shape, which is the aim of this chapter, and Chapter 5. In what follows I provide a brief historical overview which seeks to trace the origins and development of Bayes Theorem, from the original publication of the work by its eponymous progenitor, to the current interpretation of Bayesianism as it is regarded today. My aim here is to question the extent to which the development of such a mathematical form can be attributed to a steady and iterative logical process. Hence I consider the manner in which the original publication was received in its original epoch and by its contemporaries, and how it has been subsequently interpreted by a succession of figures associated with the field of statistics and probability.

In the first part of this section I provide a brief history of the development of Bayesian theory. In the course of this I seek to demonstrate how modern Bayesianism has arisen out of two approaches to probabilism that are often portrayed as existing in direct opposition to each other: *inverse probability*, which attempts to assign measures of belief to the probability of causes of observed events, and *frequentism*, which concerns itself with drawing of conclusions based only on the observations that can be garnered from data generated in the course of experiments. I will argue that the modern interpretation of 'Bayesianism' (sometimes referred to as 'neo-Bayesianism') owes as much to individuals traditionally conceived of as opponents of the position, as it does to its ostensive acolytes. Bayesianism can be seen to have arisen from an ostensible dialectic, between inverse probabilistic and frequentist positions. However, I question the basis of this apparent dialectic, and I demonstrate that the ways in which modern Bayesianism is conceived of, and practiced, reflect the way in which it has transcended the conventional dichotomy between the two opposing probabilistic positions. Finally I conclude by briefly considering how this transcendence of the inverse probabilistic/frequentist dichotomy may have served to bring into being a powerful new dynamic, one which facilitates intersubjective agreement, albeit one in which agreement is reinforced in a more 'objective' manner.

4.2 'Essay Toward Solving A Problem In The Doctrine Of Chances'[1]

Today, Bayes Theorem is generally accepted as taking the following form:

Posterior Probability (Probability of a Hypothesis Given Evidence)
= **Prior Probability** (Probability of Hypothesis) x Probability of Evidence Given Hypothesis

Or, in mathematical terms

$$P(H|E) = P(H) \times P(E|H)$$

The name 'Bayes Theorem' can trace its origins back to a paper published in 1764, entitled 'An Essay Towards Solving a Problem in the Doctrine of Chances' which was published in the *Philosophical Transactions of the Royal Society of London*, posthumously attributed to the Reverend Thomas Bayes. In this paper, the specific 'problem' is given as thus:

'*Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named." (Bayes 1764, p.4).

[1] Questions have been raised concerning whether Bayes himself actually wrote the article. Stigler (1999) cites a passage in a volume entitled *Observations of Man*, in which the author, David Hartley, talks of an 'ingenious Friend (sic)' who has worked on the same problem that Bayes had addressed some fifteen years earlier. Stigler has argued that, whilst there is no evidence to suggest that Thomas Bayes was acquainted with Hartley, the latter was a friend of Thomas Saunderson, the fourth Lucasian Professor of Mathematics at Cambridge University. Stigler thus puts forward the possibility that it may have actually been Saunderson who was responsible for the article. It has even been suggested that a widely-published portrait of the Reverend Bayes is not actually that of himself.

The article describes a scenario in which a ball, *W*, is rolled across a flat and square table, in such a way that the final resting place is represented by the horizontal co-ordinate θ. A second ball, *O*, is repeatedly rolled across the table *n* times. A 'success' is recorded whenever *O* comes to rest to the left of *W*. The main proposition that Bayes proves in this case is that, if one assumes a uniform distribution for the prior distribution of the probability of success (*p*), then it is possible to obtain an expression that the true probability, θ, lies between two values *a* and *b*, given the number of observed successes x. What is being investigated therefore, is a *probability distribution*, where *a* and *b* denote bounds of a probability distribution between which θ may lie (Stigler 1982, p.251). In mathematical terms the aim is to find the following:

$P(a<\theta<b|X=x)$, where X is a binomially distributed variable, and x the observed instances of it (e.g. in the above example, whenever a success is recorded with *O*).

At the time of publication, the probability distribution of this unobservable 'true' probability, on the basis of the observable outcomes of trials $P(\theta|x)$, was known as the *inverse probability*, in contrast to *direct probability,* which concerned the probability of directly observable events (the latter denoted mathematically as $P(x|\theta)$. Inverse probability generally involved the estimation of an unobservable parameter from observed data, and was commonly associated with fields such as astronomy or biology, where the direct, non-mediated access to objects of interest was impossible. Inverse probability was also often perceived as implying, at least with its critics, an interest in causality. A controversial issue with regard to inverse probability centred around the concern with prior probability distributions. In the billiard table example, a uniform distribution is assumed, recognising the fact that on an even table, there is an equal chance of the ball coming to rest at a particular position than any other. In other circumstances however, where less is known about the nature of a particular scenario, the setting of a uniform prior distribution, (or, for that matter, any other kind of prior probability distribution), can be seen to amount to an ontological claim about the world. This was viewed by many as reflecting an unsustainable metaphysical aspect, inappropriate for supposedly rational philosophising, and formed partial motivation for a series of criticisms, from the mid-nineteenth century onwards. These acted as a partial catalyst for the development of the frequentist interpretation of probability. In what follows I provide a brief history of these responses and their impact upon probabilistic thinking.

4.3 Critiques of Inverse Probability

Although inverse probability became the subject of a considerable degree of interest from mathematicians from the late eighteenth century onwards, and throughout the nineteenth century, Thomas Bayes was a rather marginalised figure during this period. Instead, figures such as Pierre-Simon de Laplace, more often associated with 'classical' probability[2],came to be attributed in pushing forward inverse probabilistic theory, in works such as his *Memoire sur la Probabilite des Causes par les Evenements*, published in 1774. Like Bayes, Laplace argues that any prior probability distribution for an unknown parameter must be uniform. However, Laplace also makes clear that any posterior probability distribution (the perceived distribution in the light of new information) must be proportional to a factor which is now taken to be the 'likelihood' of the data. Despite this innovation, Laplace maintained a principle of 'indifference' which prevented one from specifying a prior distribution. Furthermore, although the concept of 'inverse probability' is commonly attributed to Laplace, he does not appear to refer to it directly as such (Fienberg 2003). It also seems apparent that Laplace produced his work independently of any knowledge of the Bayes article (Stigler 1999).

The eighteenth century probabilists were not generally inclined to draw a sharp distinction between states of mind and states of the world (Daston 1994, p.333). Furthermore, the dichotomy of 'objectivity' and 'subjectivity' was seen at the time as an archaic idiom, confined to largely obscure works on metaphysics and logic, and of little relevance to mathematical probability

[2] Laplace defined 'probability' as follows: 'The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.' (Laplace 1820)

(Daston 1994, p.333). Hence the sense of probability as a personal measure of belief intermingled relatively comfortably with the idea of probabilities being derived from observed frequencies (Hacking 1975). However, from the mid-nineteenth century onwards, a number of mathematicians and philosophers sought to critically respond to the work of figures such as Laplace, who came to be seen as promoting a dangerously erroneous form of judgementalism. For example, certain French positivist authors, including Auguste Comte and Destutt de Tracy, denounced the application of inverse probability to areas such as judicial decision-making and scientific inquiry (Porter 1986, p.84). In Comte's case, this reflected his fundamentalist beliefs about the place and role of mathematics viz. society (Porter 1986, p.155). He believed that mathematics had no place in social science, his own positivist philosophy dictating that the sciences were irreducible, and hence that each science required its own distinctive method: 'As for any application of number and of a mathematical law to sociological problems, if such a method is inadmissible in biology, it must be yet more decisively so here...' (Comte 1855 [1974], p.492)

Comte went even further in denouncing the whole project of mathematical probability, or, as he referred to it, a' fanciful mathematical theory of chances' (Comte 1855 [1974], p.492). He saw the work of figures such as Laplace as counter-productive to the development of truly rational mathematical approaches:

> 'It is impossible to conceive of a more irrational conception than that which takes for its basis or for its operative method a supposed mathematical theory, in which, signs being taken for ideas, we subject numerical probability to calculation, which amounts to the same thing as offering own ignorance as the natural measure of the degree of probability of our various opinions. While true mathematical theories have made great progress, for a century past, this absurd doctrine has undergone no improvement, except in some matters of abstract calculation which it has given rise to. It still abides in the midst of its circle of original errors, while mankind are learning, more and more, that the strongest proof of the reality of speculation in any science whatever is the fruitfulness of the conceptions belonging to it.' (Comte 1855 [1974], p.493).

Comte saw the work of figures like Laplace, with their emphasis on subjective beliefs, as hopelessly solipsistic and metaphysical, and acted as a barrier to the need to investigate phenomena at close hand, and in detail, before making any conjectures about them:.

> 'It is with a feeling of shame...that I should have to announce at this time of day that we must study simpler phenomena before proceeding to the more complex; and that we should acquaint ourselves with the agent of any phenomenon, and with the medium or circumstances, before we proceed to analyze it.' (Comte 1855 [1974], p.493).

This kind of classical probability then, was seen as a barrier to the development of truly rational methods, like Comte's positivist philosophy that would supposedly liberate mankind from metaphysical folly. However, despite his vehement opposition to probability, Comte was cited as a key influence by the Belgian Adolph Quetelet, who developed a social theory which revolved around the conception of the 'average man', formed from the collation of statistical details from collections of individuals (Porter 1986, p.156).[3] Thus much of the adverse response to the Laplacean model of probability may have been stimulated by attempts by certain mathematicians to apply statistics to the study of social problems. What may be seen in their responses is a desire to move away from abstract mathematical discussions and towards a more practical, and readily applicable version of probability. The French economist and mathematician Antoine A. Cournot did not view subjective probability as invalid *per se*, but denied the existence of it possessing any meaningful use. Maintaining a distinction between objective and subjective probabilities, Cournot claimed the supremacy of the former, which gave the 'measure of the actual possibility of things', whilst subjective probabilities related 'partly to our knowledge and partly to our ignorance' (quoted in Porter 1986, p.84). Subjective probabilities were inherently fallacious in that they varied 'from one intelligence to another, according to their capacities and the data with which they are provided' (quoted in Daston

[3] And who would perform statistical research linking crime to a range of social factors.

1994, p.336). However, Cournot did not wish to see the practice of statistics brought down to a too prosaic level, and thought that rigorously developed theory would allow observers to transcend mere 'compilations of facts and figures' (Daston 1994, p.336). Furthermore, there appeared a desire amongst others to define and clarify a number of related terms in their bid to develop a more usable science. For example, the mathematician Denis Poisson appears to have been instrumental in clarifying the distinction between the 'chance', say, of obtaining a head or a tail from a coin toss, and the 'probability' of one's beliefs about obtaining a certain result (Daston 1994).

Thus whilst the intentions of these critics may have varied, they all promoted in their critiques an alternative conception of probability that emphasised the observation of sequences of real events over any epistemic notion of probability. Porter (1986) and Daston (1994) have both drawn attention to the possibility that these authors promulgated a revitalised distinction between objectivity and subjectivity, and that they shared a conception of 'objective probability' which can be seen to represent the origins of the frequentist position (Porter 1986, Daston 1994). Of the British authors of the time, the polymath Robert Leslie Ellis pursued the critique of the existing probabilistic order in a particularly enthusiastic manner. Having initiated a public debate involving, amongst others, the astronomer John Herschel (Porter 1986, p.79), Ellis concluded that the form of probability as practiced by Laplace and others amounted to nothing more than a 'sensational philosophy', and in response, proposed that probability statements be only made about series of events rather than measures of belief, the first time that any such argument had been made (Porter 1986, p.80). Ellis argued for an empiricist form of probability, one in which 'probability must be associated with observation, not ignorance, of phenomena and allied with notions of order and statistical regularity, not chance'. (quoted in Porter 1986, p.80).

The interpretation of probability put forward by Ellis, which emphasises the observation of repeatable instances of the same phenomenon, and the measurement and analysis of their frequencies, is now referred to as *frequentism*. Although this term would not be coined for some time to come, the roots of frequentism can be detected in a number of critiques of inverse probability put forward by other thinkers throughout the nineteenth century, who viewed the latter as involving questionable assumptions.

For example, the mathematician George Boole viewed the conversion of direct to inverse probability as involving two arbitrary parameters, first the *a priori* probability that a fixed cause exists, and to the probability that this unspecified cause would suffice to produce the observed effect (Porter 1986, p.80 verbatim). Boole came to subscribe to an empiricist approach to probability, stating that 'logic and probability set before us, what, in the two domains of demonstrative and of probable knowledge, are the essential standards of truth and correctness – standards not derived from without, but deeply founded in the constitution of the human faculties' (quoted in Porter 1986, p.81). However, he combined this with a certain realist construal of the regularity of nature:

> 'the rules which we employ…in the other statistical applications of the theory of probabilities, are altogether independent of the *mental* phenomena of expectation. They are founded upon the assumption that the future will bear a resemblance to the past, that under the same circumstances the same event will tend to recur with a definite numerical frequency…' (quoted in Porter 1986, p.81).

Boole concluded that one could determine probabilities given strong or perfect knowledge of a particular situation, but even this knowledge needed to be based on repeated instances of success or failure.

Whilst others promulgated an early form of frequentism in their responses to Laplacean probability, a certain heterogeneity can still be detected in their attitudes. Like Ellis, John Stuart Mill began as another vehement critic of the probability of causes, although he refined his view somewhat following the intervention of John Herschel (Daston 1994, p.337). However, with regard to Mill's conception of probability, Daston (1994) argues that Mill's view of 'scientific' probability was not the same as those of Poisson or Cournot. Whereas the views of the latter were of an ontological bent, emphasising states of the world not necessarily experienced by human observers, Mill, it is argued, based his notion of probability on personal knowledge:

> 'We must remember that the probability of an event is not a quality of the event itself, but a mere name for the degree of ground which we, or some one else, have for expecting it'. (quoted in Daston 1994, p.337).

The German Jacob Friedrich Fries, a Kantian, was responsible for introducing frequentism to his native country (Porter 1986, p.85). Taking a distinctly nomological bent, Fries criticised the inverse probabilists for ignoring the faculties necessary for rational judgement. He saw the object of probability as the rescue of general laws of nature from the obfuscation of contingency. If this were achieved, it would render the arbitrary assumptions of inverse probability erroneous and redundant (Porter 1986, p.85). John Venn developed a particularly rigorous treatment of frequentism, arguing that quantitative belief could not be justified with respect to individual nonrepeatable events, but, like Fries, argued that true probability rested on a postulate of ultimate statistical regularity (Porter 1986, p.85).

Thus this period saw the genesis of approaches which can be seen to form the foundations of the frequentist position. However, it would be some time before frequentism would become accepted as a viable and applicable scientific approach. This occurred in the early part of the twentieth century, driven by the work of Ronald Aylmer (R.A.) Fisher in response to the statistical hegemony of the day, which at the time was represented by the work of Karl Pearson at University College London.

4.4 Karl Pearson and R.A. Fisher

Although Karl Pearson is primarily associated with a number of other statistical innovations, such as the chi-squared test, linear regression, and the classification of probability distributions, Pearson also subscribed to, and promoted, a form of inverse probability. However, although Pearson emphasised the role of experience in determining *a priori* personal probabilities, Harold Jeffreys (later to play an important role in the renaissance of inverse probability) claimed that a strong frequentist thread ran through Pearson's work:

> 'the anomalous feature of his work is that although he always maintained the principle of inverse probability...he seldom used it in actual applications, and usually presented his results in a form that *appears to identify a probability with a frequency.*' (Jeffreys 1939, p.383, quoted in Fienberg 2003, p.7).

At the beginning of the twentieth century Karl Pearson was a venerated figure within the field of statistics, and came to found the first ever academic statistics department, at University College London. However, in papers published in 1912 and 1915, Pearsons' work came under the first series of attacks from R.A. Fisher, then still an undergraduate at Cambridge (Fisher 1912, 1915; Howie 2002). In his critique of Pearson's chi-squared test, Fisher proposed an alternative method, suggesting an 'absolute criterion' based on what may now be conceived of as frequentist principles – 'a direct and invariant procedure for maximising a function proportional to the chance of a given set of observations occurring' (Fisher 1915). However, Pearson's response was to accuse Fisher of mis-applying inverse probability, and of falling prey to the Principle of Insufficient Reason (the idea that, in the absence of any useful information, equal probabilities must be assigned to all possible hypotheses). Following his academic feud with Pearson, Fisher rejected an offer of a position in the latter's statistical Laboratory at UCL and instead took up a position at a small agricultural research facility at Rothamstead, Hertfordshire. Here he developed and refined certain statistical techniques, most notably confidence intervals and significance testing.

Fisher devoted a great deal of his career attempting to overcome what he saw as the inherently arbitrary nature of inverse probability, taking his cue from the proto-frequentist critics of the 19th century (Zabell 1989). Fisher attempted to make his first break with inverse probability via his formulation of the maximum likelihood estimate of a parameter, which he delineated from the concept of maximum posterior probability (where probability distributions are adapted to reflect new information). However, rather like his forebears Fisher appears to have had a rather idiosyncratic idea of what 'inverse probability' actually meant. Edwards (1997) claims that Fisher employs two contradictory meanings of the term in his 1912 paper (Edwards 1997). Fisher does state in this paper that an estimated probability, p, of an observed random variable corresponds to the function of an underlying causal parameter θ, in line with a generalised definition of inverse probability at the time. However, he also states:

> 'P is a relative probability only, suitable to compare point with point, but incapable of being interpreted as a probability distribution, or giving any estimate of absolute probability.' (Fisher 1912, Edwards 1997,)

Here Fisher appears to be insinuating that the measure of p, obtained from direct observation, is but an estimate based on the observers individual experience, and that this experience alone is insufficient to enable any underlying causal parameter (if one can be construed to exist) to be calculated. Furthermore, according to Fisher, inverse probability is incapable of reflecting any continuous probability distribution from which the true value of θ may be drawn, given the lack of information concerning the causal parameter. Fisher's rejection of the possibility of the inverse probabilistic method to reflect continuous probability distributions reflects his failure to accept the setting of uniform prior probabilities. He opposed this assumption as many before him had, as he saw it as dangerously arbitrary and unable to be verified on a strictly empirical basis.

Somewhat ironically however, Fisher had been accused by Pearson and his colleagues of employing such an assumption in his early work. In their response to Fisher's criticisms of their work on chi-squared, Pearson and his colleagues interpreted Fisher's use of the phrase 'most likely value' as the value obtained from maximising a posterior probability distribution from an assumed uniform prior (Soper et al 1917). Although they did not mention the term outright, Fisher accused them of assuming that he himself had been influenced by Bayes Theorem (Fisher 1921). Part of Fisher's work had focused on the correlation coefficient, the means of determining the strength of the relationship between two random variables. He had shown how the 'most likely' value of the correlation of a sample was slightly smaller than that of the population from which they were derived. As he states in Fisher (1921):

> 'This conclusion was adversely criticised in *Biometrika*, apparently on the incorrect assumption that I had deduced it from Bayes theorem. (Fisher 1921, p.207).

Fisher vehemently denied this assumption, and rejected the concept of inverse probability outright:

> 'As a matter of fact, as I pointed out in 1912 (Fisher 1912) the optimum is obtained by a criterion which is absolutely independent of any assumption respecting the *a priori* probability of any particular value. It is therefore the correct value to use when we wish for the best value *for the given data*, unbiased by any *a priori* presuppositions.' (Fisher 1921, quoted in Edwards 1997, p.180).

In his later work Fisher's approach to inverse probability appears to have changed slightly, to a form which appears to be closer to the Bayesian reasoning used today. In his 1922 paper 'On the Mathematical Foundations of Theoretical Statistics', Fisher states the principle of inverse probability as resting on:

> ''If the same observed result A might be the consequence of one or the other of two hypothetical conditions X and Y, it is assumed that the probabilities of X and Y are in the same ratio as the probabilities of A occurring on the two assumptions 'X is true', 'Y is true'. (Fisher 1922, quoted in Aldrich 1997, p.164).

Fisher still had his reservations about the tenability of inverse probability which centred on the assumption of a uniform prior distribution. His objections reflected his frequentist beliefs:

> 'Inverse probability amounts to assuming that…it is known that our universe had been selected at random from an infinite population in which X was true in one half, Y in the other.' (quoted in Aldrich 1997 p.169)

The formula above also reflects another of Fishers intentions. In the two papers published in 1921 and 1922, Fisher sought to strike out against inverse probability and formalised the concept of 'likelihood'. Fisher still felt that inverse probability made too many presumptions on the basis of the data that it used:

> 'if the population of interest is itself drawn from a known super-population, we can deduce using perfectly direct methods the probability of a given population and hence of the sample. But if we do not know the function specifying the super-population, we are hardly justified in simply taking it to be constant.' (quoted in Howie 2002, p.61).

Fisher regarded the shortage of empirical warrant as a major shortcoming of inverse probability, and argued that Bayes himself had considered this a problem. Furthermore, he saw the 1763 article as a resolutely frequentist document, a logical relationship between direct frequencies in populations. He accused the inverse probabilists of the day of mis-interpreting the paper:

> 'Bayes attempted to find, by observing a sample, the actual probability that the population value lay in a given range. In the present instance the complete solution of this problem would be to find the probability integral of the distribution of [an underlying parameter] θ. Such a problem is indeterminate without knowing the statistical mechanism under which different values of θ come into existence; it cannot be solved from the data supplied by a sample, or any number of samples, of the population. What we can find from a sample is the *likelihood* of any particular value of θ, if we define the likelihood as a quantity proportional to the probability that, from a population having the particular value, a sample having the observed value of *θ*, should be obtained. (Fisher 1921, p.24)

Here, Fisher argues that Bayes emphasised the fact that a series of observations were being made, which constituted a sample in a larger, and possibly infinite, set of observations. In order to determine θ, a probability distribution curve would need to be plotted, the function of which would represent another function, that of the 'statistical mechanism' which would have to be known to the observer. In the absence of this knowledge, the value of θ derived from observations could only be regarded as a likely estimate. The 'likelihood' would be the measure derivable from the probability of observing a particular value of θ given that this is the true value. Fisher essentially argues therefore, that Bayes had not after all appealed to inverse probabilistic assumptions, but emphasised observation instead, rendering him an empiricist and a proto-frequentist.

Fisher continued to define likelihood as a notion distinct to the concept of probability. In his 1925 publication, *Statistical Methods for Research Workers*, a volume that came to be used widely by a variety of scientific researchers, Fisher writes:

> 'The deduction of inferences respecting samples, from assumptions respecting the populations from which they are drawn, shows us the position in statistics of the theory of probability...this is not to say that we cannot draw, from knowledge of a sample, inferences respecting the population from which the sample was drawn, but that the mathematical concept of probability is inadequate to express our mental confidence or diffidence in making such inferences and that the mathematical quantity [likelihood] which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability' (Fisher [1925]1932, pp.9-11, quoted in Aldrich 2003, pp.80-81).

Rather than a measure of probability *per se*, likelihood is therefore viewed as an entirely separate calculational measure.

In outlining his conception of likelihood, he used the phrase 'Bayesian' in a pejorative fashion to define any approach different to his own, and rejected the ability of any such approach to deliver probability distributions for unknown parameters (Fienberg 2003). The use by Fisher of the term 'Bayesian', albeit in such a belittling way was somewhat striking in that, until then, Bayes' name had remained largely at the margins of statistics and probability, eclipsed by the supposedly more prestigious figures like Laplace and Pearson. More intriguingly however, Fisher's concept of likelihood bears striking parallels with the likelihood function central to modern interpretations of Bayesianism ('Neo-Bayesianism'): the probability of observed data given a particular choice of parameter (or hypothesis). This despite Fisher's overriding philosophy of probability being normally viewed as frequentist, a position now normally strongly contrasting with Bayesianism. Fisher defined probability in terms of frequencies of theoretically infinite populations, and argued that many inferences concerning parameters, such as linkage values in genetic problems, were simply inexpressible in terms of probabilities, given the way in which inverse probabilities were used at the time (Howie 2002).

To summarise this section, I have sought to describe how the intervention of R.A. Fisher, via his work and also with his debates with Karl Pearson, laid many of the foundations for what is now regarded as frequentist probability. Particularly in the early part of his career, Fisher was opposed to inverse probability and sought to develop statistical methods which overcame its apparent shortcomings. Most notable amongst these is the formulation of his concept of likelihood. What can also be seen through the discussions between the Fisher and Pearson however, is the inconsistency of the use of terms such as 'inverse probability' and 'Bayesianism'. Much of the criticism from Pearson about Fisher's work is based on the latter supposedly mis-interpreting what inverse probability was meant to represent; and even continuing to rely on inverse probabilistic assumptions in his work. On the other hand, other figures, such as Harold Jeffreys, accused Pearson himself of essentially using frequentism in his own work. As I have described however, in his derivation of likelihood, Fisher actually claims the original work of Thomas Bayes to be frequentist in orientation, and misunderstood by avowed inverse probabilists. Yet this did not stop Fisher from using the term 'Bayesian' as a pejorative term on a par with inverse probability. This is possibly all the more ironic given the manner in which the work of figures such as Fisher, and subsequently Jerzy Newman and Egon Pearson (Karl's son), represented attempts to semantically fix terms such as 'significance', 'confidence' and 'likelihood' itself.

The work of Fisher and his colleagues proved to be enormously influential, and was instrumental in delivering frequentism to a position of pre-eminence. By the 1930s, inverse probability, despite being the subject of so much interest in the previous century, had fallen considerably out of favour. Fisher's own earlier work can be seen to have contributed to this decline. *Statistical Methods for Research Workers*, which espoused his earlier frequentist views, had sold widely in the US and UK. Furthermore, Karl Pearson's son Egon, and his collaborator Jerzy Neyman, had developed their own frequentist approach. Thus by 1930 frequentism had become the dominant paradigm in statistics, with both Fisherian and Neyman-Pearsonian ideas being used in a wide variety of scientific fields, albeit in an often erroneous chimeric form (Gigerenzer et al 1989).

The methods of Fisher, Neyman and Egon Pearson were embraced as suitably powerful and original by researchers, but more importantly the semantic fixity of such terms as 'confidence', 'significance' and 'likelihood' in the hands of these figures must have also lent considerable authority to their ideas. Indeed, the fixing of terms such as these could be seen to be one of the main contributions of statistics and probability during the early twentieth century. However, this was generally a tendency of frequentist thought, and the work of Bayes during this period appears to have been subject to a fair degree of interpretation. What may also help explain the lack of movement in probability theory circles was the strong and widely accepted distinction that had arisen between 'statistics' and 'probability'. By now the former term had come to be referred to the treatment of large quantities of data, whereas the latter was taken as an epistemic concept associated with inference (Howie 2002, pp.186-187). Given the fact that 'statistical' concerns were by now closely tied to practical applications, (and hence proving exceedingly useful), it is perhaps unsurprising that there was little theoretical activity to reconcile the two elements of the dichotomy.

4.5 The Revival of Inverse Probability and the Rise of Neo-Bayesianism

Through the work of figures such as Harold Jeffreys however, inverse probability would come to be revived. His volume *Theory of Probability*, published in 1939, outlined a theory of inductive inference based on the principle of inverse probability and using Bayes theorem (Aldrich 2003). Whilst Jeffreys largely respected Fisher's views and his standing in the field (Howie 2002), he sought to provide a more systematic form of inference in contrast the relatively *ad hoc* style employed by Fisher. It is, however, interesting to focus upon how each conceived of likelihood. Aldrich (2003) distinguishes Jeffreys' idea of likelihood from Fisher's by pointing out that the former derived likelihood from a family of conditional distributions where the conditioning variable is a parameter, whilst the latter involves a family of unconditional distributions merely indexed by a parameter (Aldrich 2003). Hence, in the case of Jeffrey's likelihood, the parameter is viewed as *causative*, whilst in Fisher's case, it is merely *correlative*. Fisher also claims that likelihood is 'appropriate for measuring our order of preference among different possible populations', whilst it seems Jeffreys would have argued that posterior probability to be the most appropriate measure (Aldrich 2003).

It would take some time before Jeffreys' work was used by others seeking to develop inverse probabilistic methods further. However, the onset of the Second World War provided an impetus for further research, and Jeffreys ideas did influence the work of the celebrated computer scientist Alan Turing (Fienberg 2003)[4]. This in turn impacted upon the work of Denis Lindley and I.J. Good, who would later make significant contributions to the growth of what has been coined the neo-Bayesian movement (Fienberg 2003). However, neo-Bayesianism is regarded as gaining further impetus via the instrumental

[4] The development of computers during this time appears to have been instrumental in facilitating the increased study of neo-Bayesian methods (Interviewee 2).

interventions of Leonard J. Savage at the University of Chicago, and, Howard Raiffa and Robert Schlaifer at Harvard Business School.

What is notable about the work of Savage, and of Raiffa and Schlaifer, is that, despite their status as the progenitors of neo-Bayesianism, little recourse is actually made to the original Bayes article in their key works. L.J. Savage's influential 1954 volume *The Foundations of Statistics*, often cited as a key neo-Bayesian text, makes only one reference to the term 'Bayesian' in the entire text, and does not appear to avowedly subscribe to any form of Bayesian philosophy (Fienberg 2003). Furthermore, in the preface to the 1971 second edition of his volume, Savage admits to his ignorance of the likelihood principle during the writing of the original volume. Perhaps rather controversially, Savage cites Neyman and Pearson as leading the way towards Neo-Bayesianism: '...personalistic statistics appears as a natural late development of the Neyman-Pearson ideas', and claims that the likelihood principle, 'a corollary of Bayes theorem', is a consequence of the analysis of admissibility[5] as carried out by the frequentists (Savage 1972, p.iv).

Savage's retrospective acknowledgement of the distinctly non-Bayesian approach of Neyman and Egon Pearson is somewhat at odds with the approach taken by Howard Raiffa and Robert Schlaifer, who were researching the problems faced by businessmen making decisions under conditions of uncertainty. They had attempted to apply frequentist concepts to personal decision-making, but apparently found difficulties in translating these ideas. Commenting on Schlaiffer (who had no mathematical background), Raiffa made the point that his colleague saw the work of the frequentists as incommensurate with their intentions:

> "he read Fisher, Neyman and Pearson...and he concluded that standard statistical pedagogy did not address the main problem of a businessman – how to make decisions under uncertainty. Not knowing anything about the subjective/objective philosophical

[5] Admissibility: a particular decision function (ie a function chosen in an attempt to optimise a random variable x), which produces an expected loss less than or equal to any other value.

divide, he threw away the books and invented Bayesian decision theory from scratch.' (quoted in Fienberg 2003, p.17).

Here it seems that, whilst Fisherian and Neyman-Pearsonian methods served the purposes of certain types of scientific inquiry, Raiffa's remark about Schlaiffer appears to indicate the limitations of frequentism outside of scientific contexts, (e.g. laboratory situations where experiments can be controlled etc.) There appears to be recognition of the fact that frequentist methods were limited in the guidance they could provide in other areas of decision-making. Or, frequentism was found to be worthwhile for scientific *reporting*, but not necessarily for inference in conditions of uncertainty. However, as Raiffa's remark about Schlaifer 'inventing Bayesian decision theory from scratch' indicates, the rejection of frequentism did not come about from a certain logical inevitability of the superiority of the Bayesian method. In fact, the *ad hoc* nature of Schlaifer's approach indicates a certain disregard for any kind of frequentist/inverse probability dichotomy. Instead, their approach represents a combination of the two. Schlaifer, along with Raiffa, who *had* been trained in classical statistics, derived Bayesian definitions for previously frequentist concepts such as sufficiency, and, more crucially, adopted Fisher's definition of likelihood (Fienberg 2003).[6] There is also an explicit acceptance of the ability of frequentist and Bayesian methods to co-exist:

> 'the so-called 'Bayesian' principles underlying the methods of analysis presented in this book are in no sense in conflict with the principles underlying the traditional decision theory of Neyman and Pearson' (Raiffa and Schlaifer, quoted in Aldrich 2002, p.86).

Thus despite the comments by Raiffa about his colleague above, it appears to be the case, that, like Savage, they eventually recognised the contribution of frequentism during the later stages of their work.

---

[6] Fienberg (2003) actually claims that Raiffa and Schlaifer adopted a version of likelihood used by both Fisher and Jeffreys. However, this of course runs counter to Aldrich's claim of a distinction between the two, which I accept.

The leading Anglo-American statistician I.J. Good was the first to explicitly use the term 'neo-Bayesianism', defining it as such:

> 'By a neo-Bayesian or a neo-Bayes/Laplace philosophy we mean one that makes use of inverse probabilities, with or without utilities, but without necessarily using Bayes' postulate of equiprobable or uniform initial distributions, and without explicit emphasis on the use of probability judgements in the form of inequalities.' (quoted in Fienberg 2003, p.18).

What is notable here is the rejection of the postulate of uniform prior distributions, which seems to have been a firm tenet of the original 1763 article. This is a key move which helps define the neo-Bayesian approach, enabling neo-Bayesians to widen the scope of their analysis to personalised, subjective degrees of confidence, and to allow the justification for determining prior probability distributions via a range of methods.
The fact that Good makes explicit reference to 'Bayesianism' is significant nonetheless, as is the Fisherian tendency to use 'Bayesian' only sparingly, and then normally as a generally derisive term, had endured until the 1950s/60s (Fienberg 2003, p.18). Even in Good's comment it is clear that, despite the use of 'Bayes' in their new philosophy, there is a considerable move to distance the new approach from the concepts outlined in the 1763 article. The rejection of the uniform prior assumption appears as an acknowledgement of the criticisms ranged against Bayes and Laplace from the litany of critics, from the likes of Cournot et al in the nineteenth century to Fisher in the early twentieth.

The permeation of Neo-Bayesianism into the wider scientific milieu appears to have been stimulated by a number of developments. In addition to the wartime efforts of their UK counterparts, US statisticians had made important contributions to weapons development, including the atomic bomb (Interviewee 2 2007). Another notable development also came about via a military application. In 1966, a B-52 bomber carrying four hydrogen bombs collided mid-air with a refuelling aircraft above the Mediterranean Sea, nearby the coast of Spain. Three of the devices were recovered on land near the

Spanish village of Palomares. The fourth however was declared missing, and a search team, led by Dr John Craven, was charged with recovering the device. In the course of this search they devised a system in which the area of investigation was divided into a grid. This effectively acted as a probability distribution, and the team went about searching for the device in an order where the highest probability were searched first. As each area was searched, the probabilities in each grid area were revised using Bayes Theorem. If an area did not show signs of containing the device, the probability that this area contained the bomb were lowered at the expense of other areas. As well as providing a means of guiding the search party around the search site, it also enabled the team to estimate the length of time they would need to search, which also helped with the planning of the mission. The success of this method led to Bayesian search theory being used on other occasions, most notably for the recovery of the submarine *USS Scorpion*, which went missing in 1968, and has the method has since been incorporated into search and rescue planning software by the US Coast Guard (de Groot 2004).

The approach found even more widespread application, with statisticians working for NBC in the USA developing a series of sophisticated Bayesian methods to enable the results of elections to be called (Fienberg 2003). Since then, Bayesian inference has also played an instrumental role in the field of Artificial Intelligence, and is now commonly used as the algorithmic basis of internet search engines such as Google (Economist 2006). The apparent success of Bayesian inference in this context has stimulated a great deal of academic interest over the issue of whether the human mind can be likened to operate in a Bayesian way, with some studies arguing for a positive correlation (Griffiths and Tenenbaum 2006; Xu and Tenenbaum 2007; Gopnik et al 2007). Bayesian Theory has also enjoyed a heightened level of popularity in a number of academic disciplines, most notably economics, but also archaeology (Scales and Snieder 1997), medicine (Lucas 2001), veterinary science (Clough et al 2003). Some Bayesians have even gone as far as stating that Bayesian inference goes far beyond the uses of conventional statistical methods, enabling reasoners to determine the relationship between correlation and causation, such as that between smoking and lung cancer (Economist 2000).

4.6 Conclusion

This brief historical survey has sought to describe the development of Bayesian approaches to probability, beginning with the publishing of the Reverend Thomas Bayes' original article on the subject in 1793. What this account demonstrates is an enduring degree of uncertainty concerning the aims and intentions of Bayes. Some histories of probability align Bayes' work with the inverse probability movement which was in existence at approximately time, and most notably associated with Laplace. It was the latter who bore the brunt of much criticism from empiricist critics in the nineteenth century, who themselves lay the seeds for the development of frequentist methods by R.A. Fisher and others. Bayes' work was largely marginalised until it received mention by Fisher, who alternated between using his name as a synonym for inverse probability and claiming it as a frequentist innovation. Hence it can be seen that a fair degree of uncertainty exists regarding what the work of Bayes was meant to truly represent. What is more clear however, is that Fisher's conception of likelihood, and other frequentist assumptions, were incorporated into the work of 'Neo-Bayesians'. This new usage of Bayes therefore, appears to encompass insights from both the frequentist and inverse probabilist traditions.

The origins of Bayes are therefore shrouded in a considerable amount of interpretive and semantic ambiguity. The modern interpretation of Bayes owes a considerable amount to the work of figures who are either associated with a frequentist, or at least proto-frequentist position. In modern statistical and forensic scientific discourse the term 'frequentism' is however used to describe a mode of probability entirely separate, and often opposed to, the modern Bayesian approach.

It is this approach which has aroused a great deal of recent interest in forensic scientific circles, namely for the apparent potential it apparently holds for the interpretation and evaluation of forensic evidence. Bayesianism is seen in some respects at least as a suitably more 'scientific' method, in contrast to the reliance on experience and intuition, which is viewed by supporters of Bayesianism as a flawed and inappropriate means of performing forensic scientific investigation. Furthermore, in forensic scientific circles, Bayes is seen as a superior form of probability to frequentism, most notably for the apparent ability to function as a holistic means of assessing various pieces of evidence relevant to an investigation, in contrast to the apparent limitations of frequentism, which can only address individual evidential problems. Yet as this chapter has demonstrated, modern Bayesianism itself rests on foundations which themselves demonstrate signs of subjectivity and ambiguity.

In this way, modern forms of Bayes represent a combination of both 'objective' and 'subjective' ideas concerning probability. The rejection of the uniform prior probability assumption opened the way for the possibility of using Bayes Theorem in a way that allowed prior assumptions to be based on various forms of data, including datasets of observed variables.

The combination of what are sometimes taken to be incommensurable approaches - inverse probabilism and frequentism - has notable implications. In the form of neo-Bayesianism, the two have been combined to form an extremely powerful model of 'rational' decision-making, based on the combination of two deterministic modes of thought pertaining to each position. Neo-Bayesianism borrows from frequentism the obsession with patterns and sequences, and, from inverse probabilism, it borrows the emphasis on ascertaining underlying causal factors influencing observed events. A model of a 'rational decisionmaker' is created, albeit one that has to contend with an ontology in which an underlying 'cause' that can be ascertained from a world in which nature is seen to adhere to ordered patterns of sequences. It is a model which serves to combine versions of both the 'objective' and the 'subjective'. A dynamic of 'objective subjectivity' is constructed, whereby cohorts of decisionmakers come to agree on their beliefs, supposedly free of personal bias, and instead come to accept their way of reaching a decision as being 'objective'. This may be seen to act, therefore, as a powerful social instrument for facilitating and reinforcing intersubjective agreement.

In the next chapter I chart the history of Bayes Theorem in relation to its application to the study of legal evidence, and show how it has come to be a highly influential part of modern evidence studies.

CHAPTER FIVE - THE DEVELOPMENT OF EVIDENCE SCHOLARSHIP

5.1 Introduction

In the previous chapter I provided a historical account of the uses of, and attitudes toward, Bayes Theorem in a variety of contexts. In this chapter I supplement this with an overview of the development of other related efforts to provide systematic methods for the interpretation of evidence - in relation for courtroom trial procedures in particular and forensic science in general. Modern approaches differ from earlier ones in that they involve the use of probabilistic methods, most notably the version of Bayes Theorem described in the previous chapter. However, an examination of their historical trajectory demonstrates how each attempt has contributed to the evolution of a field of study which has been referred to by terms such as 'evidence scholarship' or 'evidence studies'. Insights from this field have exerted a major influence on thinking within forensic scientific circles regarding evidence interpretation and evaluation. In what follows, I map the development of this area by drawing upon examples of work which have sought to address issues of evidence interpretation in a more systematic manner. Examples, drawn from a number of different disciplines, including law, philosophy, statistics and forensic science, will show how the study of evidence has evolved to encompass Bayesian forms of reasoning.

I begin by describing the work of the legal scholar John Henry Wigmore, who, in the early part of the twentieth century, outlined what is generally regarded as the first holistic and systematic framework for the interpretation of evidence in the course of criminal trials. Whilst now regarded as highly influential by scholars working in the field of evidence interpretation, Wigmore's work was largely marginalised for a considerable part of the twentieth century. In the 1950s and 1960s however, the possible use of statistics and probability aroused both interest and controversy within the academic study of law in the USA, and I describe the debates which arose within the academic literature at the time. I also compare the reception of probability in other fields with that found in forensic scientific circles. I show how the work of the Americans Paul Kirk and Charles Kingston proved to be influential in introducing a greater consideration of probabilistic concepts in forensic science. This was also facilitated in the UK by figures such as Ian Evett, who, working in conjunction with the Bayesian Denis Lindley, helped to introduce Bayes into some of the working practices of the UK Forensic Science Service (FSS). I discuss both of these developments, and then show how Bayesianism in UK forensic science has evolved further. I show how this process of evolution has not only been influenced by the work of figures such as Kirk, Kingston, Evett and Lindley, but also by the revival of interest in the work of Wigmore himself. This rediscovery has contributed to what has been referred to as 'The New Evidence Scholarship' (Lempert 1986), which in turn has also helped facilitate the emergence of new technologies from bodies such as the FSS.

5.2 Beginnings: John Henry Wigmore

In June 1913, John Henry Wigmore published 'The Problem of Proof', an article in which he set out a programme for studying legal evidence presented in court (Wigmore 1913). Wigmore viewed his programme as focusing on two distinct areas: *the issue of proof in the general sense*, 'the part concerned with the ratiocinative process of contentious persuasion' (Wigmore 1913, p.77), and the other that of *admissibility*, 'the procedural rules devised by law, and based on litigious experience and tradition, to guard the tribunal (particularly the jury), against erroneous persuasion' (Wigmore 1913, p.77). In his opinion, studies of law had exclusively focused on the latter, in a manner that he felt to be detrimental to the progress of law. For Wigmore, notions of proof 'in the general sense' would remain invariant as the foundation for law, in contrast to admissibility, which he saw as 'merely a preliminary aid to the main activity' of proof, namely 'the persuasion of the tribunal's mind to a correct conclusion by safe materials.' (Wigmore 1913, p.78).

Wigmore also saw the principles of proof as an invariant, objective set of postulates which concerted scientific inquiry would help to reveal in the course of time. In his writings he anticipates that:

> 'The judicial rules of Admissibility are destined to lessen in relative importance during the next generation or later. Proof will assume the important place, and we must therefore prepare

ourselves for this shifting of emphasis.' (Wigmore 1913, p.78).

Here, Wigmore appears to predict a move away from a concern with tradition in legal procedure, to one more in thrall to the influence of rational, and more avowedly 'scientific' principles. However, he also argued that legal professionals required suitably timely instruction in a 'science of proof', otherwise he feared a repeat of the experience of Continental Europe of the 1800s, where one system had been abandoned without a clear definition of a suitable replacement:

> 'For centuries, lawyers and judges had evidenced and proved by the artificial numerical system[1]; they had no training in any other, - no understanding of the living process of belief; in consequence, when 'legal proof' was abolished by fiat and the so-called 'free proof' –namely, no system at all, was substituted, they were unready, and judicial trials have been carried on for a century past by uncomprehended, unguided, and therefore unsafe mental processes.' (Wigmore 1913, p.78).

Whilst the previous numerical system may be seen by modern standards to be lacking in rationality, Wigmore's point is that it was replaced by a subsequent system which contained no equivalent system for guiding the process of proof. Although Wigmore saw the nature of legal procedure as open to reform, he saw more work needed before it could be considered as a suitably rational method of proving.

In approaching the issue of proof, Wigmore stated his aims as:

> 'To perform the logical (or psychological) process of a conscious juxtaposition of detailed ideas, for the purpose of producing rationally a single final idea. Hence, to the extent that the mind is unable to juxtapose consciously a larger number of ideas, each coherent group of detailed constituent ideas must be reduced in consciousness to a single idea; until the mind can consciously juxtapose them with due attention to each, so as to produce its single final idea.' (Wigmore 1913, p.80).

Wigmore was concerned that the large number of pieces of evidence, testimony, arguments etc that make up most criminal cases meant that the average finder of fact

[1]The numerical system to which Wigmore refers to above involved the method of weighing evidence based either on the number of witnesses who testified in a particular fashion, or by representing witness testimony by other numerical means, such as allocating fractions to testimony depending on the nature of the issues which were the subject of such testimony.

could struggle to comprehend all the elements relevant to it: 'Many data, perhaps multifarious, are thrust upon us as tending to produce belief or disbelief' (Wigmore 1913, p.79). Furthermore, he saw the need to avoid adducing disproportionate weight to individual pieces of evidence by privileging them over other evidence:

> 'our object is (in part) to avoid being misled...through attending only to some fragments of the mass of data. We must assume that a conclusion reached upon such a fragment only will be more or less untrustworthy.' (Wigmore 1913, p.79).

Wigmore emphasised what he saw as the inherent fallibility of human belief, distinctly separate from 'external reality, or actual fact' (Wigmore 1913, p.79). Belief about an apparent fact, being a purely mental phenomenon, is dependent on 'how fully the data for the fact have entered into the formation of our belief' (Wigmore 1913, p.79). However, a key issue for Wigmore with regard to the apprehension of facts is temporality; one can only apprehend multiple facts one after another:

> 'But those data have entered into the formation of our belief at *successive times*; hence a danger of omission or of interior attention. "Knowledge in the highest perfection would consist in the *simultaneous* possession of facts. To comprehend a science perfectly, we should have every fact present with every other fact. We are logically weak and imperfect in respect of the fact that we are obliged to think of one thing after another"[2]. And in the court room or office the multitude of evidential facts are originally apprehended one after another. Hence, the final problem is to coordinate them. Logic ignores time; but the mind is more or less conditioned by it.' (Wigmore 1913, pp.79-80, original emphasis).

Hence one of the main problems that Wigmore perceived with regard to the process of proving concerned the formulation of an appropriate method of depicting the numerous pieces of evidence relevant to a case. Allowing the apprehenders of fact to view the evidence in such a holistic way would facilitate a more logical approach to the consideration of evidence, and overcome the privileging of certain pieces of evidence due to the conditioning effect of temporal succession. Wigmore saw the latter as exacerbated in the context of judicial procedure: 'So many interruptions and distractions occur, both to the lawyer in the preparation to the jurors in the trial, that

---

[2] Wigmore does not cite his source for this quotation in his article.

facts cannot be properly co-ordinated on their first apprehension' (Wigmore 1913, p.80).

However, there were a number of issues with regard to evidence that Wigmore had to take into account. For example, he was conscious of the possible uncertainties concerning the reliability of certain forms of testimony. Hence he sought to devise a system which enabled each piece of evidence relevant to a case to remain prominent, but which simultaneously allowed a comprehensive overview of all the facts pertaining to the case. For any such systemic representation of evidence, Wigmore specified a number of conditions:

- it must contain a logical typology of evidential forms
- it must be possible to use this typology to depict and arrange all the evidential data in a given case
- it must be able to depict the relation of an evidential datum to each and all of the others
- it must be able to show the distinction between a 'fact' as *alleged* and a fact as *believed/disbelieved*
- it must present the evidence in a simultaneous manner
- it must be generally be usable by practitioners of law
- the scheme should *not* attempt to show what one's beliefs *ought* to be; 'it can hope to show only what our belief actually *is*, and *how* we have actually reached it' (Wigmore 1913, p.82).

This final point is of particular interest, as a subsequent discussion by Wigmore demonstrates that in some ways he is ahead of his time with regard to his approach to evidence. Wigmore is quick to point out that any such scheme cannot tell one what should be believed:

> 'We know only that our mind, reflecting upon five evidential data [A,B,C,D,E], *does* come to the conclusion X, or Not-X, as the case may be. All that the scheme can do for us is to make plain the entirety and details of our actual mental process. *It cannot reveal laws which should be consciously obeyed in that process.*' (Wigmore 1913, p.82, emphasis added).

Perhaps more notably still, Wigmore is keen to argue over the existence of the limits to current logic in terms of the kind of conclusions, and hence general laws, that can be induced from evidence. In doing so he also distinguishes his aims from other contemporary scientific approaches to evidence:

> 'Much indeed has been done that is theoretically presentable in judicial trials. Much indeed has been done that is theoretically applicable to circumstantial evidence; e.g. the method of differences, in inductive logic, may enable us, with the help of a chemist, to say whether a stain was produced by a specific liquid. But these methods must be pursued by a comparison of observed or experimental instances, newly obtained for the very case in hand, and usually numerous, hence they are impracticable for the vast mass of judicial data.' (Wigmore 1913, pp.82-83).

What Wigmore appears to be arguing here is that each criminal case is unique, and, given this fact. The rules of scientific induction, which he sees as central to scientific analysis, are inapplicable. Because each criminal case is unique, at least to a certain extent, it is impossible to make any comparisons across cases. Hence any system used to study the means of proof by evidence, has to be able to reflect the nuances of each case.

Wigmore argued that the logical and psychological developments of the time had not yet extended sufficiently in scope to consider the net effect of a mass of mixed data bearing on a single fact, i.e. problems concerning how the possibility of contradictory evidence affected the comprehension of a particular fact to be proved. Wigmore's stance toward developments in scientific disciplines, and over their potential to inform a science of proof, was therefore somewhat unclear[3]. What Wigmore sought was significantly different of the aims and objectives informed by the logic and psychology of his day. Although those disciplines made some contribution to the study of legal proof, Wigmore viewed them as betraying a certain lack of practical application with regard to the concerted study of judicial evidence considered as a whole. Furthermore, Wigmore accepted that even a consciously 'scientific' approach could not provide definitive answers with regard to questions of guilt or innocence. The kind of system he proposed was not intended to enable one to calculate such answers, or to attempt to

[3] Wigmore did not advocate methods such as the conscious use of psychological methods to help assess the reliability of witnesses for example. However, in this case his objections appear to be based to a great extent on pragmatic grounds, namely that complex psychological testing to determine the degree of reliability of a set of witnesses would unnecessarily delay the progress of a trial (Wigmore 1913).

uncover general laws regarding evidence and criminality, but to guide the reasoning processes of the finders of fact and enable sources of error or bias to be weeded out:

> 'Perhaps we cannot explain *why* we reach that result, but we know at least that we *do* reach it. And thus step by step we set down the separate units of actual belief, - connecting, subsuming, and generalising, until the subfinal grouping is reached; then dwelling in consciousness on that; until at least a belief (or disbelief) on the final fact evolves into our consciousness.' (Wigmore 1913, p.83).

What Wigmore appeared to have in mind was not the production of a 'science' of evidence as such, but to merely the consideration of the development of technologies which might help judicial reasoners to overcome the cognitive problems associated with cases, such as bias, oversight or the unbalanced privileging of certain pieces of evidence against others. He also sought to highlight the types of interdependencies between pieces of evidence which could be identified, such as the ability of one form of evidence to influence the perceived level of accuracy or reliability of another piece of evidence.

Wigmore's method involves the use of a graphical system to depict supposedly how advocates reason through a case. In doing so Wigmore aimed to provide a suitably holistic and descriptive representation of how the reasoning processes in a case may proceed. The Wigmore method also enables one to map the reasoning processes of evidence interpretation from both a prosecution and a defence viewpoint.

An example of Wigmore's chart system is provided in Figure 5.1. Evidence is first classified into a number of different forms: testimonial, circumstantial, explanatory (used either to explain circumstantial evidence in a prosecution context, or to discredit circumstantial evidence, when used in accordance with a defence position), and corroborative, which involves different usages depending on whether it is being used in conjunction with circumstantial or testimonial evidence. Corroborative evidence for the former could involve the divulgence of extra information which could restrict the number of possible explanations surrounding circumstantial evidence. For example, in the case of the discovery of a knife possibly implicating a defendant in a murder case, the evidence that no third party was seen near the knife might add weight to prosecution circumstantial evidence concerning a defendant's guilt. In a similar prosecutorial vein, corroborative evidence for testimonial evidence might seek to add further trustworthiness to testimonial evidence, for example by emphasising the supposed neutrality of any eyewitnesses etc). Corresponding defence evidence might, for example, advance evidence which could question the trustworthiness of the testimonial evidence.

Figure 5.1 displays many of the features typical of a Wigmore chart. The case under consideration concerns the murder of a farmhand in Massachusetts, which was brought to court in 1901. The accused, who was subsequently found guilty, worked alongside the deceased on the same farm. The headless body of the victim had been found hidden in a sack in an unused well four to five hundred feet away from the horse barn of one Keith, the landowner. The prosecution alleged that the defendant killed the victim in the barn, using some kind of tool located in the building.

Circle 60 indicates a proposition relating to circumstantial evidence. The dot inside the circle indicates that this is believed by the prosecution to be true. In this instance the proposition is that clothing found on the remains of the body had mud from the barn. The adjoining square 61 indicates testimony from a witness which claimed a match with mud from the barn. Arrowhead 62 indicates testimonial evidence from the defendant that this mud hadn't been identified.

Figure 5.2 depicts the symbols used in Wigmore's method. Different sets of symbols are used for different forms of evidence, depending on whether the evidence is testimonial, circumstantial, explanatory or corroborative in nature, and whether it is being used in an affirmatory or negatory manner. The positioning of symbols in relation to each other bears upon their affirmatory relationship, with a single straight line being placed between two evidential facts to indicate the supposed relationship with each other. Any supposed fact seemingly acting to prove the existence of another fact is placed below the latter. A supposed explanatory or corroborative fact, which may strengthen or lessen the evidential force of another, is placed, respectively, to the left or right of that fact. The relevant symbol for a fact observed by the court is placed under the relevant fact.

Wigmore also proposed a system for indicating the possible probative force of evidence, using various types of labelled arrows. Affirmatory evidence is depicted by adding a single arrowhead. Supposed negatory evidence is represented by a single arrowhead plus the addition of a small circle (see Figure 5.2). Multiple arrowheads may be used to show where particularly strong credit may be accrued to certain pieces of evidence. For pieces of evidence which elicit a strong degree of belief, solid dots may be added to broadly represent the degree of belief in the alleged fact. Where there may be reason to doubt evidence, a question mark may be added to the relevant symbol, or to an arrow to indicate doubts over the probative force. Measures of disbelief in an evidential fact are represented by hollow circles.

Figure 5.1. An Example of a Wigmore Chart (reproduced from Wigmore 1913, p.93)

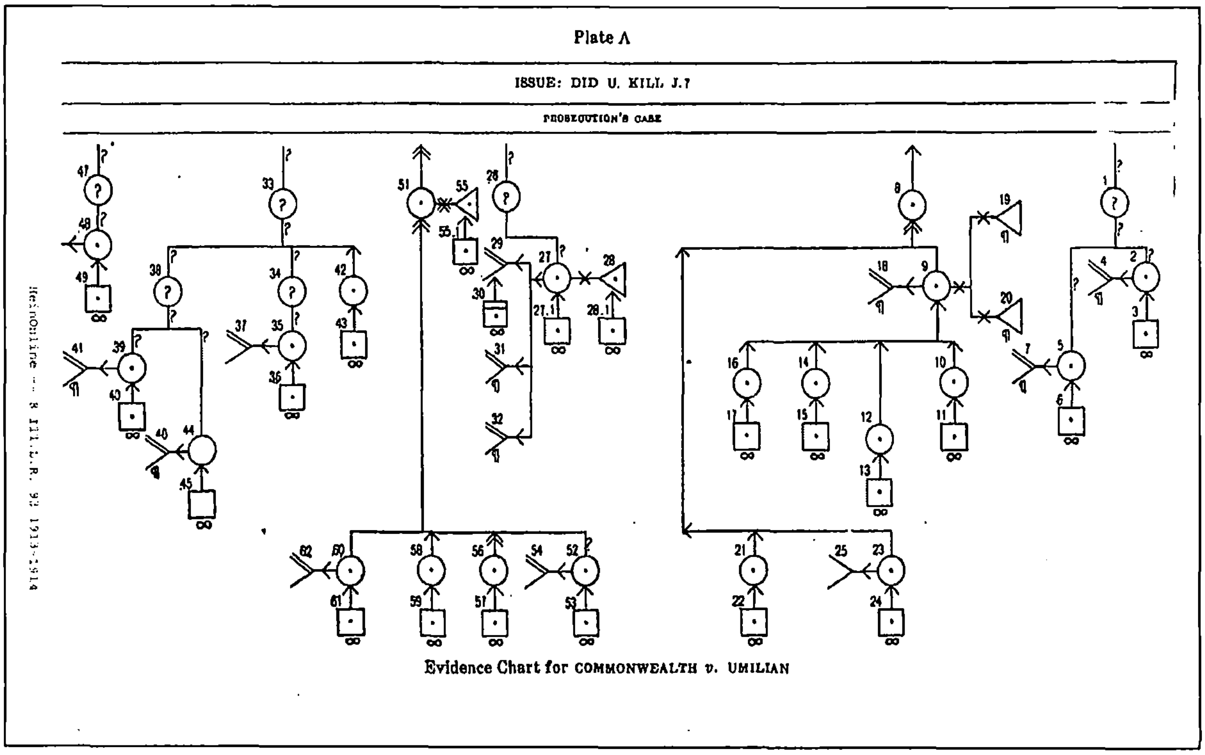
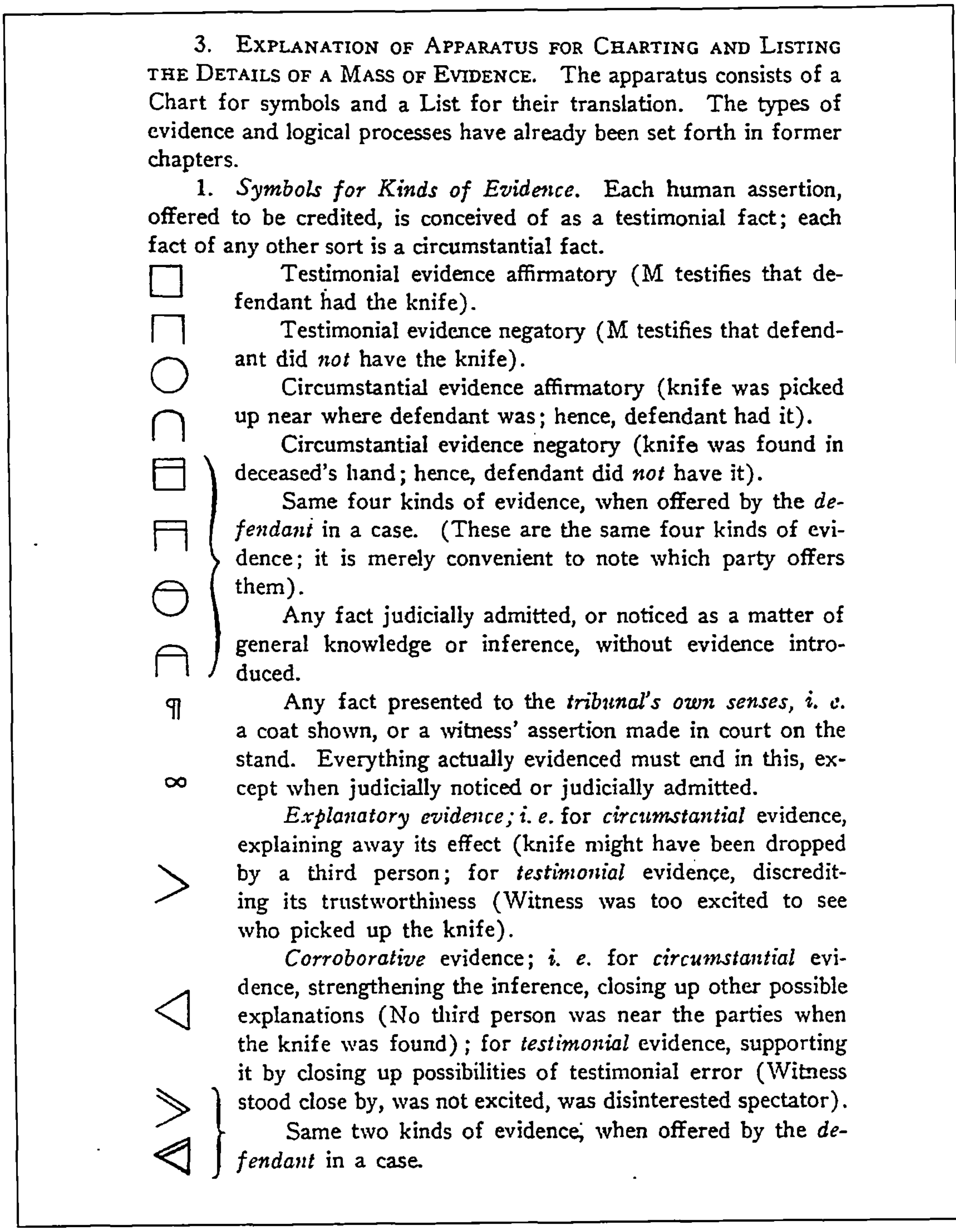


Figure 5.2a. Reproduction from Wigmore (1913) showing key for symbols used in chart system (Reproduced from Wigmore 1913, p.84)



3. EXPLANATION OF APPARATUS FOR CHARTING AND LISTING THE DETAILS OF A MASS OF EVIDENCE. The apparatus consists of a Chart for symbols and a List for their translation. The types of evidence and logical processes have already been set forth in former chapters.

1. *Symbols for Kinds of Evidence.* Each human assertion, offered to be credited, is conceived of as a testimonial fact; each fact of any other sort is a circumstantial fact.

Testimonial evidence affirmatory (M testifies that defendant had the knife).

Testimonial evidence negatory (M testifies that defendant did *not* have the knife).

Circumstantial evidence affirmatory (knife was picked up near where defendant was; hence, defendant had it).

Circumstantial evidence negatory (knife was found in deceased's hand; hence, defendant did *not* have it).

Same four kinds of evidence, when offered by the *defendant* in a case. (These are the same four kinds of evidence; it is merely convenient to note which party offers them).

Any fact judicially admitted, or noticed as a matter of general knowledge or inference, without evidence introduced.

Any fact presented to the *tribunal's own senses, i. e.* a coat shown, or a witness' assertion made in court on the stand. Everything actually evidenced must end in this, except when judicially noticed or judicially admitted.

Explanatory evidence; i. e. for *circumstantial* evidence, explaining away its effect (knife might have been dropped by a third person; for *testimonial* evidence, discrediting its trustworthiness (Witness was too excited to see who picked up the knife).

Corroborative evidence; *i. e.* for *circumstantial* evidence, strengthening the inference, closing up other possible explanations (No third person was near the parties when the knife was found); for *testimonial* evidence, supporting it by closing up possibilities of testimonial error (Witness stood close by, was not excited, was disinterested spectator).

Same two kinds of evidence, when offered by the *defendant* in a case.

Figure 5.2b Reproduction from Wigmore (1913) showing key for symbols used in chart system (Reproduced from Wigmore 1913, p.85)

2. *Relation of Individual Pieces of Evidence, shown by position of Symbols.*

A supposed fact tending to prove the existence of another fact is placed *below* it.

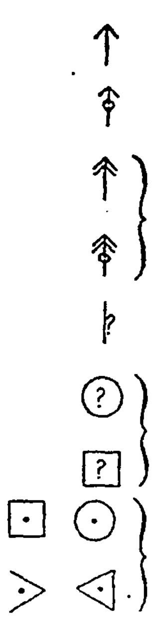
A supposed explanatory or corroborative fact, tending to lessen or to strengthen the force of fact thus proved, is placed to *left* or *right* of it, respectively.

A single *straight* line (continued at a right angle, if necessary) indicates the supposed relation of one fact to another.

The symbol for a fact observed by the tribunal or judicially admitted or noticed (¶, ∞) is placed directly *below* the fact so learned.

3. *Probative Effect of an Evidential Fact.*

When a fact is offered or conceived as evidencing, explaining, or corroborating, it is noted by the appropriate symbol with a connecting line. But thus far it is merely *offered.* We do not yet know whether we believe it to be a fact, nor what probative force we are willing to give it, if a fact. As soon as our mind has come to the necessary *conclusion* on the subject, we symbolize as follows:



(1) *Provisional credit* given to *affimatory* evidence, testimonial or circumstantial, is shown by adding an arrow-head.

Provisional credit given to *negatory* evidence, testimonial or circumstantial, is shown by adding an arrow-head above a small cipher.

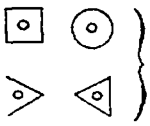
Particularly *strong credit* given to those kinds of evidence respectively is shown by doubling the arrow-head; this is usually applicable where several testimonies or circumstances concur upon the same fact.

(2) A small interrogation mark, placed alongside the connecting line, signifies *doubt* as to the probative effect of the evidence.

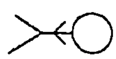
Similarly, for each kind of symbol, a small interrogation mark within it signifies a mental balance, an uncertainty; the alleged fact may or may not be a fact.

(3) A dot within the symbol of any kind of *alleged fact* signifies that we now *believe* it to *be* a fact. Particularly strong belief may be signified by two dots; thus.

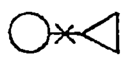
Figure 5.2c Reproduction from Wigmore (1913) showing key for symbols used in chart system (Reproduced from Wigmore 1913, p.86)



A small cipher within the symbol of any kind of alleged fact signifies that we now *disbelieve* it to be a fact. Particularly strong disbelief may be signified by two such ciphers; thus .

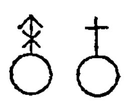


(4) If a single supposed *explanatory* fact does, in our estimation after weighing it, detract from the force of the desired inference (in case of a witness, if it discredits his assertion), we signify this by an arrow-head pointing to the left, placed half way across the horizontal connecting line.



If a single *corroborative* fact is given effect in our estimation, we signify this by a short Roman letter X, placed across the connecting line.

Doubling the mark indicates particular strength in the effect, *i. e.* , or .



Ultimately, when determining the total effect, in our estimation of *all* explanatory and corroborative facts upon the *net probative value* of the specific fact explained or corroborated, we place a short horizontal mark or small X, respectively, upon the upright connecting line of the latter fact.

Thus, for *net probative value,* several grades of probative effect may be symbolized: signifies that the inference is a weak one; signifies that it has no force at all; signifies that it is a strong one; signifies that it is conclusive. When the supposed inference is a *negatory* one, the same symbols are used, with the addition of the negatory symbol, *i. e.* (Witness asserts that defendant had *not* a knife in his hand; witness's credit is supported by the fact that he is a friend of the deceased).

4. *Numbering the Symbols.*

Each symbol receives a number, placed at the upper left outside margin. These numbers are then placed in the Evidence List; they are written down consecutively, and opposite each one in the list is written a brief note of the evidential fact represented by it.

The List is thus the translation of the Chart.

With regard to ancillary evidence, the presence of an arrowhead pointing left on a link between explanatory evidence and testimony represents a degree of detraction in the apparent force of the evidence (e.g. it may represent evidence that might discredit the assertions of a witness). A single corroborative fact is signified by placing an 'X' across the line. As before, in both these cases double symbols may be used to indicate a particularly significant degree of strength. In instances where they may be numerous instances of explanatory and/or corroborative facts bearing upon a particular piece of evidence, a single 'X' or horizontal line, is respectively placed on the connecting line of the latter fact. A piece of evidence may thus be seen to have net probative value, which is symbolised by either the presence of a single horizontal line to indicate weak probative force, whilst a double line is used to indicate no net probative force. An 'X' is used to indicate strong net probative force, whilst a double 'X' indicates conclusive force. For negatory probative force, the same symbols are used, with the addition of a circle to indicate negation.

Wigmore's chart method was unique. It represented one of the first, and certainly the most enduring attempt to address the issue of the rationality of courtroom testimony. What made his method particularly innovative was the emphasis on addressing the problem of courtroom proof in a holistic way. The chart method not only emphasises the fact that a multiplicity of pieces of evidence and testimony may be involved in a trial, but also that interdependencies may exist between them, in terms of how they influence each other's perceived worth. Hence another important feature of the chart method is the means by which it allows a reasoner to estimate the possible inferential strength of evidence. Wigmore's work is also notable in the use of a visual system through which evidence is depicted, allowing the reasoner an overview of the inter-relationships apparent in a case.

Despite this, Wigmore's work had little impact in either academic or practitioner contexts at the time of its initial publication. The precise reasons for this are largely a matter of conjecture; however they may be largely related to practical concerns. The construction of Wigmore charts, particularly via the technology available at the time, would have been a highly time-consuming and labour-intensive affair. A more serious defect however, is the fact that Wigmore charts only allow one to interpret evidence *after* the case. This in itself would have limited the chart method to be considered largely as a pedagogic aid only, and as a means of demonstrating to students the possible sources of irrationality present in a case. The chart method, in the manner that Wigmore describes, does not possess any predictive power, which might have accorded it more usefulness, and was mainly confined to use as a teaching aid in his lectures to students at Northwestern University.

5.3 Statistics in the Courtroom

The subject of evidence continued to be addressed by American legal scholars nonetheless, notably by figures such as Charles Tilford McCormick and John MacArthur Maguire, who both wrote influential textbooks (Maguire 1947, McCormick 1955). Discussions on matters of evidence also continued in the American academic literature, such as the issue of re-defining key evidential expressions such as 'reasonable doubt', 'clear and convincing evidence' and 'preponderance of evidence', in terms that could convey the degrees of belief (McBaine 1944). Later articles addressed the issue of the extent to which the study of evidence could be modelled on the methods of science. Criticisms were made about the lack of scientific import into evidential matters, for example Cleary (1952) advocated greater consideration of the insights of psychology in enabling an improved understanding of the perception and behaviour of witnesses brought to give evidence in court (Cleary 1952). The call for external influences to help aid the problem of evidence also included considerations of the potential of probability theory and statistics. For example, Ball (1961) advocated a 'change in attitude toward judicial truth, a change which channels intuition into the realisation of what can be accomplished with the aid of modern statistical tools' (Ball 1961, p.830).

Discussions concerning the use of statistics in the American courtroom continued throughout the 1960s in the wake of one particularly controversial instance. The case *People vs. Collins* involved the trial of an inter-racial couple for a robbery in California of an elderly woman. The victim claimed that her assailant had been a young Caucasian woman with blond hair tied in a ponytail, who escaped in a yellow car driven by an African-American man who wore a beard and moustache. The prosecution case rested on the argument that the chances of selecting any couple possessing these characteristics would be one in twelve million, a figure that was obtained by the use of the product rule. The California Supreme Court ruled against this prosecution argument, viewing the use of probability theory in such a manner as erroneous. Aside from the issue concerning the reliability of eyewitness testimony, objections were made concerning the application of the product rule in such a way. The assumptions of independence for each characteristic were seen as patently false; for example, the probability of a man possessing a beard *and* a moustache was not seen as encompassing two strictly independent features.

Around the same time, a parallel discourse was emerging in forensic science. In 1963, Paul Kirk, a leading figure in the US forensic science community, published a short article, 'the ontogeny of criminalistics', which effectively marked an attempt to define the scientific essence of forensic science:

> 'The real aim of all forensic science is to establish individuality, or to approach it as closely as the present state of the science allows. *Criminalistics is the science of individualism.*' (Kirk 1963, p.236, original emphasis).

Kirk's statement came at a time when there was also an increased interest in the possibility of the use of statistical methods in forensic science. In 1964, a special session on statistics was held at a meeting of the American Academy of Forensic Sciences, which reflected the apparently 'growing awareness of the usefulness of statistical methods' (Kingston 1965a, p.79). Charles Kingston, a student and collaborator of Kirk, published two articles concerning 'Applications of Probability Theory to Criminalistics' (1965a, 1965b), which discussed and put forward a number of probabilistic models for the assessment of Partial Transfer Evidence (PTE), including an explicitly Bayesian model. Kingston (1966) also addressed the issue of statistical dependence, which had been ignored by the court in *People vs. Collins* (Kingston 1966).

A further notable intervention in this context came via Kaplan (1968) who sought to apply the insights of decision-making theory which had been developed in the field of business studies. As described in the previous chapter, the challenge of business decision-making under conditions of uncertainty was approached via the application of a form of Bayesian reasoning. In advancing a Bayesian model, Kaplan sought to address the 'incredibly unsophisticated' use of statistics in court, as represented by *People vs. Collins*. In the article Kaplan draws a distinction between 'objective' and 'probabilistic' theories of probability, equating respectively to frequentist and subjective modes, in which probability in the latter corresponds to measures of personal belief (Kaplan 1968, p.1066-1067)[4]. Kaplan viewed objective probability as 'not very helpful for application to the complex factual decisions that are the grist of our legal system', given that trials were singular events:

> 'Given a typical contested trial, for instance, it is meaningless to speak of the probability of the defendant's guilt in terms of the number of times he would be guilty in an infinite number of exactly similar cases because, first, there are not even two exactly similar cases, and, second, even if there were many identical cases the court must reach a verdict, not a ratio, in the case at the bar.' (Kaplan 1968, p.1066).

The trial scenario is instead viewed as more amenable to a personalistic mode of probability which informs decision theory. In its most elementary form, the decisionmaker, faced with a range of options, anticipates the consequence of each option by estimating the probabilities of outcomes relating to each option. The expected value of each option is determined by multiplying the probability of each outcome by the value accorded by each outcome. For example, in a simple gambling game where there is a 1 in 2 chance of receiving £1000 (say on calling the outcome of a coin toss), the expected value would be £500. In a game where there is a 1 in 3 chance of winning £1000 (where one has to guess the location of a ball hidden under one of three upturned cups), the expected value would be £333.33. The concept of expected value however is not able to explain decisionmaking alone, for it does not account for the differences in how individuals may personally perceive and value rewards, which may vary due to individual circumstance. This is explained by expected utility, which shows how decisions may be reached on options that have less expected value. For example, a decision to insure a property may have lower expected

[4] 'For many years most mathematicians and statisticians defined the probability of an event's occurring in a given trial as the proportion of times an event would occur as the number of exactly similar trials approached infinity. We shall call this 'objective' probability'.Professor Savage, in a small number of postulates, described the personalistic theory of probability. This theory postulates that it is meaningful to ask of someone "If I gave you a reward for guessing correctly, would you choose to guess that *A* is true of that *B* is true?" If the subject chooses *A*, then we may say for him *A* is more probable than *B*. If he is indifferent to whether he chooses one or the other, we may say that as far as he is concerned they

value than a decision not to insure (i.e. it will cost an individual to pay a premium for an event with relatively low probability, such as fire, burglary etc), but the possible consequences of not insuring, if such a disastrous event were to occur, are normally considered as too damaging for this option to be taken (Kaplan 1968, p.1069).

Although expected utility is considered a largely personal measure, in a business context it nevertheless usually equates to financial loss or gain. In a criminal trial scenario, expected utility is less easy to tangibly define and may vary according to a number of factors. For example, expected utilities may vary according to different forms of miscarriages of justice; the decision to convict an innocent suspect may be considered more harmful than the acquittal of a guilty one, hence the use of the 'reasonable doubt' standard in the prosecution of criminal trials. The disutility of a miscarriage may also vary according to the individual, with a wrongful guilty verdict more damaging in the case of someone considered of good moral character or high social standing. It may vary depending on the link between the nature of the crime and the mandatory punishments. The correct decision in a capital murder case may carry greater utility than in a case involving a minor transgression which carries a far less severe sentence. Furthermore, expected utility may vary according to the nature of the crime and other circumstances. For example, it may be considered less serious to society to acquit an embezzler who may not be placed in a position of trust in the future, than a sex offender who may repeatedly continue to transgress (Kaplan 1968, pp.1073-1074).

Under this schema the possible variance of expected utility may condition the personal assessments of jurors considering the probability of guilt or innocence. At the beginning of a trial, a juror could be asked to guess a personally selected probability of guilt or innocence of the defendant. This *prior probability* could be an entirely subjective guess, but as the trial progresses, the measure of probability of guilt or innocence may vary as pieces of evidence are sequentially put forward. The probability estimate following the presentation of all the evidence may well have changed to a posterior probability of a different value. Such a process is redolent of Bayes Theorem, and the latter is put forward by Kaplan as a suitable means of

are equally probable. Professor Savage also demonstrated that this personalistic probability can easily be expressed in terms of quantitative probabilities.' (Kaplan 1968, pp.1066-1067).

continually formulating updated posterior probabilities of guilt or innocence, and thus allowing beliefs to be updated in the light of new information. The hypothetical question of what might be considered a suitable starting prior probability is kept open, with Kaplan offering possible alternatives such as 1 in 2 (equivalent to an even balance between guilt or innocence), or 1 in 200 million for an American trial (corresponding to the population of the USA).

Kaplan's article is best regarded as an exploration of the possibilities of using Bayesian decision theory in a trial context, rather than advancing it as a suggested solution to the challenges of judicial factfinding. In conclusion Kaplan admits that the judicial process is ultimately better suited to the human brain, rather than the technology available at the time of writing, arguing that the deconstruction of all potential variables relevant to a case into probabilities and sub-probabilities would create an impossibly vast computational task. This leads Kaplan, in his conclusion, to ruminate about the possible differences between human reasoning and digital computation:

> 'Even though a judgement about a compound probability such as guilt must rationally be based on conclusions about a vast number of subprobabilities, the human mind, at least consciously, does not work that way. Perhaps a more accurate ultimate probability would be reached by asking a factfinder for its judgement about all subprobabilities, but no one has yet demonstrated that this is true. It may well be that, because of processes not yet understood, the human brain behaves more like an analogue than a digital computer, or that for some other reason the brain automatically corrects for misjudgements about some probabilities, either by making compensating errors in others or by reaching a final result by a mathematically incorrect means that intuitively compensates for the previous error.' (Kaplan 1968, p.1091).

The insights of both Kingston and Kaplan came together in an influential article by Finkelstein and Fairley (1970). Also responding to the criticisms over *People vs. Collins*, they published an article proposing an approach based on Bayesian probability theory (Finkelstein and Fairley 1970). The authors accepted the dangers inherent in the misuse of probability and statistics, and argued that it was both wrong and futile to attempt to attribute uniqueness to evidential data (Finkelstein and Fairley 1970, p.496). However, in presenting a Bayesian framework, they argued instead that it was possible to use probability theory to establish the significance, and probative force, of seemingly infrequently occurring evidence.

Finkelstein and Fairley argue that in many cases, identificatory evidence can not be claimed to be unique, merely rare (Finkelstein and Fairley 1970, p.516). Such evidence however, may still possess significant probative value. In order to assess and express this probative value however, a suitable statistical method is required. Finkelstein and Fairley cite Bayes theorem as an appropriate means for translating frequencies of rare events into a probability assessment by combining them with prior probabilities. In their paper, they show how Bayes Theorem may be used by introducing the example of a defendant being accused of using a knife in an alleged murder, based on an incriminating print found on the weapon

Their stated aim is to find P(G|H) in terms of P(H|G) and P(H|NG), where G is the event the defendant used the knife, and H is the event that a palm print similar to the defendant has been found ('Probability of given event H', in terms of 'probability of event H given guilt of defendant' (G), and 'probability of event H given the defendant is not guilty (NG)') , Therefore they wish to find the probability that the defendant used the knife, taking into account the chances that the defendant, or someone else left the palm print (Finkelstein and Fairley 1970, p.498). This is depicted as:

$$P(G|H) = \frac{P(G\&H)}{P(H).}$$

They show how this is actually an instance of:

$$P(G|H) = \frac{P(G)P(H|G)}{P(G)P(H|G) + P(NG)P(H|NG),}$$

which corresponds to Bayes Theorem. According to Finkelstein and Fairley, an application of Bayes Theorem would 'start with the probability that the defendant used the knife P(G) and that our views are modified or weighted by the two probabilities associated with the print, P(H|G), and P(H|NG). Our final estimate of the chance defendant used the knife is our initial or 'prior' view as modified by the statistical evidence.' (Finkelstein and Fairley 1970, p.499-500) Hence they envisage the use of Bayes as involving the formulation of a prior probability (P(G)), which is then modulated by subsequent evidence. Whilst P(H|G) would be assumed to be 1, P(H|NG) would depend on the frequency of characteristics in a suspect population, in this case statistics concerning the frequency of observing a certain type of palm print amongst a population of individuals. With P(H|G) set at 1, the values of P(G) and P(H|NG) play a crucial role in determining measures of the posterior probability of guilt P(G|H). The values for P(H|NG) relate to the frequency of a particular palm print occurring in a population, i.e. so if 50% of a population of individuals (all assumed not guilty) were found to deposit the same palm print on a knife, the posterior probability P(G|H), would be low (0.019) if the prior probability of guilt, P(G), was also low. A rarer frequency of a particular palm print occurring in a population will however result in a low value of P(G) increasing significantly to a high posterior probability of guilt; note a value of P(H|NG)=0.01, with P(G)=0.01 results in P(G|H)=0.909.

Table 5.1. Posterior Probability Table showing the effect of variance of the Frequency of Evidential Characteristics in a Population (Reproduced from Finkelstein and Fairley 1970, p.500).

	Posterior Probability P(G|H)				
Frequency of Characteristics(H|NG)	***Prior Probability P(G)***				
	0.01	***0.1***	***0.25***	***0.5***	***0.75***
0.5	0.019	0.181	0.4	0.666	0.857
0.25	0.038	0.307	0.571	0.8	0.923
0.1	0.091	0.526	0.769	0.909	0.967
0.01	0.502	0.917	0.97	0.99	0.997
0.001	0.909	0.991	0.997	0.999	0.999

The results presented in this case are theoretical and intended to show how the occurrence of evidence in a population may serve to act as an aid to guiding considerations of guilt or innocence if a Bayesian framework is used. Finkelstein and Fairley see the use of Bayes for identification evidence as a 'modest use which merely eliminates an unwarranted distinction between the force of statistical and other types of identification evidence' (Finkelstein and Fairley 1970, p.502). They also argue however, for the possibility of a 'stronger, more explicit use of the theorem', in which an expert witness could suggest a range of hypothetical prior probabilities, and specifying the posterior probability associated with each. For example, each juror would then choose the prior estimate that most closely reflected their subjective view of the evidence. The provision of precise statistical evidence would then be introduced to show how the resultant posterior probabilities would be generated. To illustrate their point, Finkelstein and Fairley cite the example of a case involving a lawyer accused of altering a document for his own gain. Eleven defects in the typewritten letters on the document were alleged to correspond with defects found on the defendant's machine. A professor of mathematics testified for the prosecution, and argued that, via application of the product rule, the joint probability of all the defects occurring from a random typewriter was one in four billion (Finkelstein and Fairley 1970, p.501). This argument was thrown out on the grounds that it was not based on observed data, and that such an abstract theory could not supply sufficient proof. Finkelstein and Fairley argue that this evidence, if incorporated into a Bayesian framework, could have been rendered acceptable. If each juror had adopted a prior probability estimate, then each piece of probabilistic data pertaining to the occurrence of each defect could then have been advanced on an iterative basis to obtain a posterior probability of belief. The authors argue that such an approach enables the data to inter-relate to the subjective beliefs of each juror, and is therefore more faithful to the traditions of the jury system:

> 'The jury's function is not to compare a defendant with a person selected randomly but to weigh the probability of defendant's guilt against the probability that *anyone* else was

responsible. Bayes' theorem translates [a statistic] into a probability statement which describes the probative force of that statistic.' (Finkelstein and Fairley 1970, p.502).

Interestingly however, Finkelstein and Fairley continue by arguing that personalistic, or subjective probabilities, in the legal context, 'are likely to be interpreted as expressing a frequency, just as "the chances of heads is one-half" expresses a frequency' (Finkelstein and Fairley 1970, p.504). Their view differs from previous authors in that they accept a possible role for statistics derived in a frequentist mode in a judicial context. They view such statistics as merely providing supporting information to inform factfinders in the course of Bayesian deliberations. Yet they go further in this combination of Bayesian and frequentist thinking, in their discussion of the 'reasonable doubt' standard:

> 'When we say that a defendant is guilty beyond a reasonable doubt, we mean that the evidence has brought us to a state of belief such that if everyone were convicted when we had such a belief the decisions would rarely be wrong. The 'beyond reasonable doubt' standard thus groups together cases which are similar but because the degree of belief in guilt has passed a certain mark...Thus, although it would it will usually be artificial to imagine a repetition of similar cases, one can nonetheless interpret subjective probability of guilt as the relative frequency of guilt over cases judged to be similar by the degree of belief they engender. The statement thus that 'there is a fifty per cent chance that defendant is guilty' thus means that if a jury convicted whenever the evidence generated a similar degree of belief in guilt, the verdicts in this group would tend to be right about half the time' (Finkelstein and Fairley 1970, p.504).

In criticising *People vs. Collins*, Finkelstein and Fairley show the use of statistics to be primitive and flawed; they do not however, reject the use of such statistics outright. Instead, they argue that the formulae used to generate the probative weight of the evidence was erroneous. They show instead how the incorporation into a Bayesian formula enables such data to be considered sequentially, avoiding certain assumptions made in that case. However, their arguments elsewhere belie a conflation of previous views concerning the status of criminal trials viz quantitative methods. They emphasise the use of frequentist statistics where appropriate, but in the course of informing a Bayesian schema which they view as suitable for jurors to reach conclusions on the ultimate issue given appropriate guidance. Their views on criminal trials however, appear to indicate some belief in a certain role for considering trials in a frequentist manner, in contrast to other authors who saw such proceedings as too individually nuanced to reflect such an outlook. Finkelstein and Fairley advance an approach that is both frequentist and Bayesian: it allows for frequentist data to inform decisionmaking, but only if subsumed into a Bayesian form of information processing. Yet their emphasis on 'reasonable doubt' seems to show they accept that cases do exert some form of commonality in terms of final outcomes. Hence it is unclear what form of probability is ultimately privileged.

Regardless of their ambiguous ontological outlook, Finkelstein and Fairley's paper sparked a celebrated debate with opposition led by Tribe (1971). Tribe's article criticised the use of Bayes in legal proceedings in a number of ways. First, the imperative to formulate prior odds was viewed as inviting arbitrariness. The need to combine probability estimates about each piece of evidence presented in the course of a trial was viewed as problematic in the case of 'fuzzier', less readily quantifiable items of evidence, and that the use of Bayes could possibly privilege quantitative data in the minds of jurors. Tribe also argued that the use of an explicitly mathematical technique could lead to the disproportionate consideration of certain questions and issues over others; issues pertaining to identity and occurrence would receive greater attention at the expense of questions of 'volition, knowledge and intent' (Tribe 1971, p.1366). The potentially high amount of information that might need to be considered in a case could lead to a need to use unwieldy mathematical constructions. More seriously, Tribe viewed the adoption of Bayesian principles as risking negative changes in reasoning behaviour, and that they threatened fundamental tenets such as the presumption of innocence and the concept of reasonable doubt. He argued that, in the process of employing probabilistic reasoning, a juror would have to give some weight to the fact that the accused had been chosen to stand trial, something which was wholly inappropriate under the presumption of innocence (Tribe 1971). This, he argued, might dangerously skew the opinion of jurors against the accused. Furthermore, the proposal to reach verdicts in terms of the probability of guilt or innocence (as well as the suggested use of expected utility theory to derive the suitable sentence) was seen to override the notion of reasonable doubt. Even if the ultimate decision was made on the basis of probabilistic reasoning, Tribe argued that the point at which the probability of guilt was sufficient to merit punishment such as a custodial sentence was essentially arbitrary, and hence inappropriate for use in court

5.4 The 'New Evidence Scholarship'

The Finkelstein & Fairley debate with Tribe helped to revive interest in the study of evidence, which by the 1970s had been questioned as an intellectually viable field. Despite the interventions of the authors discussed above, a great deal of what passed for evidence studies at this time concerned itself largely with rules and procedure for reporting evidence. In his overview of the field during this period, Twining (1984) opines thus:

> '...a stocktaking of our heritage of specialised scholarship and theorising about evidence suggests the following: a clear and straightforward theory of proof integrated with a rather simplistic normative truth theory of adjudication. A collection of relatively sophisticated concepts and distinctions in respect of such matters as rationality, relevance, standards of proof and presumptions. A broad consensus among specialists that any deviation from a presumption in favour of freedom of proof requires justification; a lack of consensus about what constitute good justifications, with the result that there is a rich, but confusing body of literature on the rationales and practical utility of those particular derivations that make up the surviving laws of evidence' (Twining 1984, p.273).

Twining criticised the field for the supposedly disproportionate emphasis on normative debate. He saw this as a result of adherence to the 'Rationalist tradition', which encouraged the postulation of '*one* law of evidence, *one* type of process, *one* kind of inquiry, and a single purpose and end to litigation.' (Twining 1984, p.273, original emphasis). This kind of argument had eclipsed empirical studies, with the result that little attention had been paid to a number of issues, including understanding of the psychological processes of proving, the actual operation and impact of the rules of evidence, the probative value of particular types of evidence in respect of certain types of hypotheses to be tested, or about how information was actually processed and used at different stages of litigation. Furthermore, Twining argued that this tradition of evidence scholarship exhibited contextual bias, in that it favoured attention to the progress of evidence in contested jury trials, 'rather than on what happens to information at every stage in a great variety of different processes and arenas' (Twining 1984, p.273).

From the late 1970s onwards however, a number of scholars began to reconsider the nature of evidence, and of proof itself, as a subject of fruitful study. The phrase 'New Evidence Scholarship' would come to be coined to describe this renewed academic approach to evidentiary matters, which shifted the agenda from a concern with the articulation of rules, to a more explicitly epistemological emphasis on the process of proof, which appropriated insights from fields including psychology, philosophy, and probability (Cohen 1977, Schafer 1976, Lempert 1986, p.439). Related studies on the nature of inference considered jurisprudential problems, and in doing so set the scene for the renaissance of the Wigmore chart method, particularly via the work of Peter Tillers and David Schum (Schum 1977, Tillers and Schum 1988).

Tillers and Schum (1988) cite Wigmore's chart method as an important demonstration that 'research on descriptive and normative matters might proceed hand in hand' (Tillers and Schum 1988, p.935). In their appraisal of the chart method, they cited two factors which they argued lent the method powerful heuristic potential for research on inferential reasoning. First, Wigmore's work was viewed as a clear indication of the extremely wide array of evidential factors that could form the focus of formal analysis. Tillers and Schum used the chart method as demonstration of the existence of 'recurrent generic forms of inferential structures, that, when examined, reveal a wide array of evidential and inferential subtleties' (Tillers and Schum 1988, p.936). Second, the original work by Wigmore was viewed as pointing to an ever wider potential research programme, namely behavioural studies concerning how reasoners evaluate and combine multiple pieces of evidence in forming beliefs. Wigmore's work was viewed as instrumental in helping to identify meaningful variables that might lend themselves to study, and for also helping researchers to structure the kind of tasks subjects could perform in experimental conditions (Tillers and Schum 1988, p.936).

A key difference concerning the re-appraisal of Wigmore's work was the incorporation of probabilistic approaches into the framings of evidential issues. Tillers and Schum (1988) saw the chart method as providing a basis for exploiting new developments in probability theory relating to the study of evidence. In discussing Wigmore's work in an updated context, Tillers and Schum (1988) thus provide a rather telling re-description of the chart method. They redefine his charts as: '*directed* *acyclic graphs whose vertices represent propositions and whose edges represent "fuzzy" probative force qualifiers'* (Tillers and Schum 1988, p.936, original emphasis). In doing so, they place the work in the context of inference networks, which at the time of writing, had began to attract a significant degree of scholarly interest. Tillers and Schum praise Wigmore for his precedence, achieved seemingly without the knowledge of many working in the newer field.

More importantly however, Tillers and Schum argue that the Wigmore chart method is a significant contribution to efforts to better understand the process of what they call ' discovery-related activity' (Tillers and Schum 1988, p.944). Whilst Wigmore provides the means of representing an inference problem, more consideration is needed in order to determine how to put together that representation in the first place. They define Wigmore's work as a *relational* means of representing reasoning about evidence, but argue that it needs to be combined with a serial, or *temporal* aspect if it is to aid a discovery-led process. The key method of combining the two aspects is via the construction of case theories about a specific incident. Wigmore's method is therefore able to act as a set of tools for outlining possible theories which relate to a specific case. Whilst the Wigmore method does not in itself act as a way of generating new hypotheses, it may act to strongly facilitate the process by enabling one to view the relationships between different pieces of evidence. However, Tillers and Schum argue that it is vital for the reasoner to consider new incoming evidence, but only in a way that may lead to new explanations and theories. For example, the introduction of a new piece of evidence, N*, about event N, may lead to a reasoner to formulate a new theory wherein N* is able to explain N. However, according to Tillers and Schum, the 'new' theory generated on the basis of evidence N* is of little use if it *only* explains N. If N* is to lead to a new theory which accounts for other evidence relevant to a case however, it may be considered significant. This new theory may then be used to generate evidentiary tests in order to assess its feasibility.

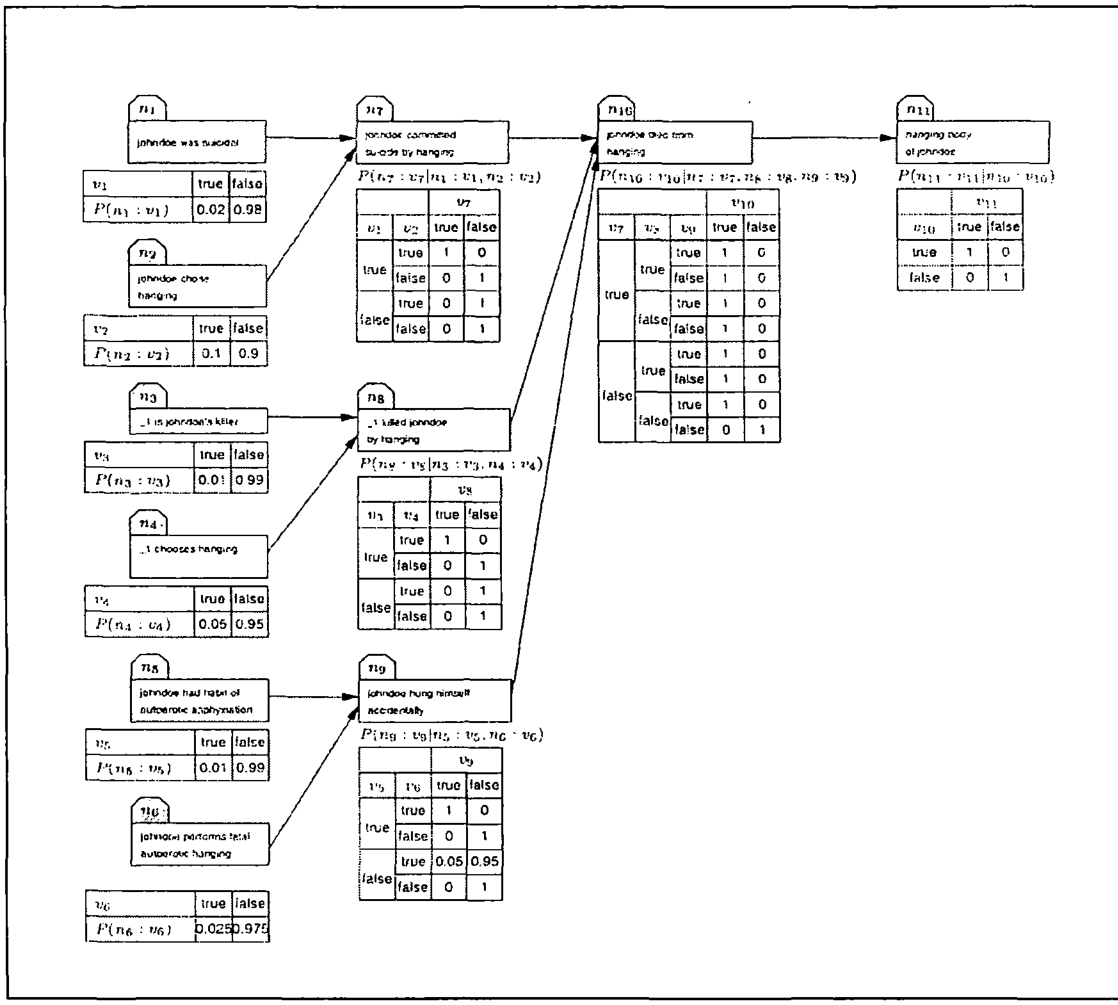
5.5 Wigmore Meets Bayes

The emphasis on the serial nature of combining evidence allotted a role for which Bayesianism was used in subsequent works, and prompted further research which sought to reconcile Wigmore charts and probability. The work of Kadane and Schum

(1996) who comprehensively applied the Wigmore method to the analysis of the controversial Sacco and Vanzetti case, represented a particularly notable fusion of Wigmore charts and Bayesianism[5] In doing so, their work exhibited strong parallels with the development of *Bayesian Networks*, inference charts incorporating Bayesian algorithms, which had been pioneered by the work of figures such as Judea Pearl (Pearl 1988). These can be used to depict chains of reasoning in which the probability of propositions may be conditional on the outcome of other probabilistic hypotheses (see Figure 5.3). Each of these hypotheses may be represented as a *node*, linked in relation to their perceived interdependencies by a series of *arcs*. Using Bayes Theorem it is then possible to model the spread of conditional probabilities.

[5] Ferdinando Sacco and Barthelemeo Vanzetti, were Italian-born anarchists who were convicted and subsequently executed for the murder and robbery of two pay-clerks in South Braintree, Massachusetts, in 1920. Their conviction has long been contested, and it is claimed that political motivations underpinned their guilty verdicts and subsequent executions.

Figure 5.3. An example of a Bayesian Network Developed for Crime Science Analysis (Reproduced from Shen et al (2007)



Although parallels have been drawn between the two, Bayesian Networks can be seen to differ from the Wigmore chart approach in a number of ways (Leucari 2005). Wigmore charts have been viewed as a technique for describing and representing a reasoning process, and hence for constructing arguments. Bayes Networks on the other hand, are viewed as tools for deriving statistical measures of inference given available evidence. Whilst Wigmore charts are formulated retrospectively, Bayes Networks may be built in a manner which adduces a certain amount of predictive capability; they may act as a means for representing the probability of a course of events taking place if reasonably accurate background data is available. Whilst Wigmore charts do not contain quantitative information enabling one to draw inference from the relevant evidence, they do provide a means for showing how particular arguments may be constructed; this is something that Bayes Networks, given their simpler representational structure, are less able to depict. The latter may be construed more as tools for assessing and measuring inference. Wigmore charts are also able to represent hypotheses at different levels of analysis, being able to depict the ultimate probandum of guilt or innocence, the penultimate probanda (hypotheses which need to be proved in order to lead toward a guilty verdict), and intermediate probanda (hypotheses which need to be assessed at earlier stages of the evidential chain). It is unclear to what extent Bayes Networks may be able to reflect this.

Nonetheless, the graphical form of mapping chains of evidential reasoning can be seen to reflect an influence from the Wigmore method. Bayes Networks have featured prominently in discussions concerning the application of Bayes to issues in both legal proceedings and forensic science (Dawid et al 2006; Shen et al 2007, Taroni et al 2006). Bayes Network-based methods have been developed to model the investigation of whole criminal cases (Shen et al 2007), but have also been used to model a number of more specific problems within forensic science, including fire incidents and DNA profiling (Evett et al 2002, Biedermann et al 2005, Taroni et al 2006). Due to the more simple means of representation, Bayes Networks can be built on a modular basis, and hence are considered applicable to a range of forensic problems. Computer programs such as HUGIN, which enable Bayes Networks to be modelled, are also now freely available.

The question remains however, as to what extent this field can be considered as a discipline in its own right. Wigmore's aims appeared to relatively modest. Whilst not rejecting the contribution of scientific evidence to a case, Wigmore appeared to caution against regarding the use and comprehension of evidence as a topic of science in its own right. This contrasts with certain current evidence scholars, who regard the rigorous study of evidence as potentially leading to a greater understanding of the nature of inference as a whole. In a separate development, the incorporation of Bayes into chart-based forms shows at least the potential for predictive power which Wigmore charts are unable to provide.

Around the same time as the advent of New Evidence Scholarship, the Bayesian approach to evidential reasoning began to receive application in the work of the UK Forensic Science Service, most notably via the efforts of Dr Ian Evett. A statistician by training, Evett had been approached whilst working for the FSS by the distinguished Bayesian Denis Lindley (Evett and Joyce 2005). Lindley and Evett would subsequently work together on the problem of applying Bayes to forensic science, and the FSS made progress in applying Bayes to PTE analyses such as glass and fibres, documented in a number of scientific articles (Evett 1984, 1986, 1987).

Evett saw the use of Bayes as providing a solution to the apparently unique problems experienced in forensic scientific work, and sought to argue that it provided a means of evaluating evidence in cases where the transfer of material had occurred from 'criminal to crime scene' (Evett 1984, p.26). Evett states that an investigator will seek to ascertain the odds on contact having taken place, or:

$$\frac{P(C|E,I)}{P(\hat{C}|E,I)} = \frac{P(E,I|C) \times P(C|I)}{P(E,I|\hat{C}) \times P(\hat{C}|I)}$$

Where C denotes a hypothesis that contact has occurred between an individual and a source, and $\hat{C}$ denotes a hypothesis that contact has not occurred. E denotes PTE of interest, for example glass fragments, and I denotes relevant background information (in the case of glass transfer, this might involve details about the source: for example, the size of window if one is suspected to have been broken, its height from the ground, the length of time elapsed between the incident and the apprehension of the suspect etc).

Notably, Evett argues that this derivation highlights the differences between the role of the scientist and the investigator (Evett 1984, p.28). According to Evett, the investigator is concerned with prior and posterior odds, and is interested in questions of the type 'what are the odds on contact having occurred between a suspect and source?' The scientist, on the other hand, is concerned with the derivation of likelihood ratios, and thus poses questions of the type 'what are the probabilities of the *evidence* under each of the alternative hypotheses C and $\hat{C}$?' In this way, the scientist provides data corresponding to the first term on the right-hand side of the above, P(E, I| C)/ P(E, I| $\hat{C}$), which informs the investigator, who is concerned with the left-hand term. Hence, the scientist is able to provide information which aids the investigator in formulating posterior odds from the likelihood ratio and the investigators prior odds.

This formula has since been adapted for further forensic use, whereby C and $\check{C}$ has been replaced by hypotheses which attempt to account for specific events of forensic interest. This method involves the use of two competing hypotheses, one pertaining to a prosecution argument (Hp), and an opposing hypothesis relating to a corresponding defence argument (Hd). In this version the formula is rewritten as:

$$\frac{P(Hp|E, I)}{P(Hd|E,I)} = \frac{P(Hp|I)}{P(Hd|I)} \times \frac{P(E|Hp,I)}{P(E|Hd,I)}$$

In applying Bayesianism to a different context, namely the evaluation of evidence in a forensic scientific laboratory, as opposed to courtroom proceedings, Evett also defined a certain delineation of roles between investigator and scientist. The former is concerned with the assessment of probabilities of guilt or innocence given the evidence, whilst the latter evaluates this evidence, providing likelihood ratios and incorporating information based on other circumstances relevant to the case.

In a later paper however, Evett admitted to experiencing a 'marked reluctance to think in terms of prior and posterior odds' (Evett 1987, p.103), amongst scientists and investigators (Evett 1987). In order to help facilitate acceptance, Evett proposed a verbal convention in order to aid understanding of the significance of the numerical information generated by the use of Bayes. Evett also identified five interrelated challenges that had to be faced 'if any mathematically based system is to become an operational reality' (Evett 1987, p.103). Firstly, Evett points to the *complexity* of real-life cases. Admitting to the simplified nature of the examples presented in his work, Evett mentions the possibility of multiple forms of PTE being involved in a case, and the possibility of two-way transfer further complicating calculations. A second major challenge is the problem of *realistic modelling*, in which the key issue is seen as one of tradeoff: 'the more assumptions we make the less complex the mathematics, the more tractable the solutions and the less the demands on our computing power; but at the same time we move further from operational reality and the smaller we make the domain of applicability' (Evett 1987, p.104). A third key issue concerns the *collection of adequate data* upon which to assess propositions. Evett admitted that the collection of data was largely a pragmatic affair and that FSS scientists were generally limited by the evidence that was presented to them in the course of casework: 'we have to resort to collecting whatever data we can lay our hands on and this generally comes down in to our laboratories' (Evett 1987, p.104). Evett cites a New Zealand colleague as describing the disparate collection of evidence as leading to 'lab door statistics', namely the formulation of statistical inferences being dictated by the small and possibly unrepresentative amount of data available to the scientist concerning a particular question. Fourth, Evett points the problem of *imponderables*. This refers to the multiplicity of background information that might play an influential role in the way a forensic scientist evaluates evidence. With regard to an individual instance of a window being broken: '...the scientist may have no information whatsoever about how close the suspect was supposed to be to the window when he hit it, or how hard he hit it, whether he ducked or swerved as he hit it' (Evett 1987, p. 104). Finally, there are the *circumstances* of a particular case to consider. This brings forth the fact that the flow of information relevant to a case may be ongoing, and hence subsequent information may be contradictory to the previous circumstantial knowledge that a forensic scientist was using in order to evaluate evidence.

Despite the acknowledgement of these challenges, Evett's work represented an application of the ideas outlined in Finkelstein and Fairley's paper, and Evett himself has cited the latter as highly influential (Aronson, unpublished transcript). Now a consultant with vast experience of testifying on evidence in high-profile cases, Evett, along with other colleagues at the FSS, has initiated the Case Assessment and Interpretation (CAI) model, which forms the basis of the case study in Chapter 7. The CAI, which has been the subject of a number of scientific articles, aims to apply Bayesian principles in the development of a set of procedural guidelines to guide the progress of forensic investigation in relation to criminal cases. The work of evidence scholars such as Wigmore, and the more recent work of scholars such as Robertson and Vignaux, and Kadane and Schum, have been cited as strongly informing the development of the model. In taking a holistic approach to the issue of evidence interpretation, the CAI attempts to encompass the issues mentioned above, but more importantly also serves to use Bayes to clarify the role of forensic scientists in an investigation:

> '*the Bayesian framework tells you what kinds of questions you should be addressing*. And if there are numbers, the Bayesian framework shows you how to take account of those numbers' (Evett in unpublished interview with Aronson, emphasis added).

In terms of the introduction of Bayes into forensic science, Lindley's key intervention, occurred at a time when it appeared that forensic science was seeking a stronger sense of disciplinary identity, as shown by the programmatic remarks by Paul Kirk, proclaiming forensic science as the 'science of individualisation'. That Bayesianism was viewed at that time as possibly contributing to the disciplinary identity of forensic science is suggested by the subsequent work performed by Kirk's pupil, Charles Kingston. In the UK, the collaboration of Denis Lindley and Ian Evett proved crucial in introducing Bayes into the powerful FSS. Lindley's intervention originally came about in direct response to the previous use of frequentist methods by FSS scientists, with Lindley effectively arguing that Bayes provided a more suitable way to assess similarity rather than difference. In this way, the frequentist-inverse probability debate, outlined in Chapter 4, is carried forward. Once again, it is possible to identify to identify different interpretations as to what various modes of probability are suited, and what they can and can't achieve. Furthermore, in considering the problem of applying probability theory to forensic science, and outlining the practical issues to be overcome, Evett shows how the problem of interpreting evidence in such a manner is inextricably linked to a wider set of circumstances and issues. Moreover, Evett shows how Bayes may come to be able to guide the behaviour of scientists in the pursuit of an inquiry.

5.6 Conclusion

In this chapter I have drawn upon academic literature to present an overview of approaches towards evidential reasoning. Questions of the nature of evidence and proof received significant attention from John Henry Wigmore in the early part of the twentieth century, and his studies on the nature of inferential reasoning continue to be influential to scholars of evidence today. His influence can be felt in the emergence of technologies of inference. Some of these have taken the form of graphical depictions, whereas others have involved mathematical operations, yet they all have been intended to provide guidance to reasoners in apprehending the potentially complex series of pieces of evidence in a trial, or to help them assess the probative weight of evidence.

Over time, statistics and probability have come to play an increased role in these technologies, with Bayes Theorem featuring prominently in these initiatives. In such a way, the story outlined here concerns the evolution of an identity for Bayes Theorem itself in a particular context, for much of the debate has concerned the role it may play in the nexus of law and science. It lies at the heart of a wider set of discussions which encompass questions about the nature of judicial reasoning itself, namely issues of whether mathematical systems will ever be fully able to cope with the potential complexity of criminal trials, or whether such deliberations are best left to human minds which possess some other, currently poorly understood mechanism, as ruminated upon by Kaplan. Furthermore, the emergence of Bayes Theorem in this context has also informed discussions over whether criminal cases can be considered as exhibiting commonalities which can inform statistical models, which in turn has shaped particular ideas and representations of Bayes, most notably the seeming incorporation of frequentist notions into an ostensibly Bayesian framework by Finkelstein and Fairley.

Regardless of the way in which these kinds of technologies have been realised, together they represent efforts to develop technologies of expertise. They are designed to aid users to reason about evidence in a manner which circumvents supposedly 'irrational' behaviours. As well as removing sources of bias and prejudice, these technologies help one to convert conjectures of evidence into quantitative measures. Hence technologies of expertise have a two-fold function: They convert subjective beliefs into tangible, intesubjectively comprehendable and standardisable measures. More importantly however, by highlighting sources of epistemological weakness in the reasoner, and acting as guides for reasoning, they serve to act as blueprints for behaviour. They provide sophisticated models for reasoning about evidence in a rational and scientific way, and in doing so help define conceptions of 'expert' reasoners.

These technologies, although discussed in US legal literature, have appeared to have received relatively little uptake by legal professionals. Although the shortcomings of the current Anglo-American legal system in relation to evidence continue to be discussed on both sides of the Atlantic, technologies of evidential reasoning have stimulated little in the way of actual reform to the system. As I have briefly mentioned in Chapter 2, and as I describe in more detail in the next chapter, attempts to use Bayesian reasoning in UK courtroom deliberations have met with outright rejection. What is notable however is the way that the thinking represented by these technologies has been appropriated more keenly in the domain of forensic science. The remarks of Paul Kirk echo those of forebears such as Edmund Locard in reflecting the desire of forensic scientists to pinpoint a defining epistemic identity for their discipline to distinguish it from other scientific undertakings. The use of Bayes in forensic science can be seen to exert a dual function in this regard. First, it aids the transition of forensic science from the categorical epistemological position towards identification, to a more conditional paradigm which is viewed as more suitably 'scientific'. Second, by providing a guide to reasoning behaviour, it enables actors to construct an identity. It contributes in the construction of the persona of a judicial 'expert'. This is important not only for placing forensic scientists on a similar footing to other scientists, but also forms a crucial part of their personal makeup which enables them to project a persona as a credible expert in the legal arena. Furthermore, the work of Ian Evett shows how Bayes Theorem can help formally define the role of the forensic scientist, not only in helping clarify the questions to ask but in distinguishing their role from those of their colleagues, the police investigators.

CHAPTER SIX - FORENSIC DNA TECHNOLOGY AND MODERN STATISTICAL APPROACHES TO EVIDENCE INTERPRETATION

6.1 Introduction

In this chapter I provide an overview of the development of forensic DNA profiling and related statistical procedures. By doing so, I aim to provide an introduction to the scientific basis of DNA profiling techniques, and to the probabilistic concepts which have been developed for the interpretation of DNA evidence. Applications of probability theory developed for DNA profiles, particularly those involving Bayesian analysis, have also been applied to other forms of evidence. This chapter therefore provides background information for the two case studies to follow, which involve the development of complex Bayesian technologies for the analysis of more complex DNA profiles (Chapter 7), and a generalised Bayesian approach to forensic investigation of evidence (Chapter 8).

In addition, in the course of this chapter I also consider the interdependence of technical, statistical and judicial discourses, by discussing the ways in which the scientific and statistical basis of DNA profiling has been contested. I examine the challenges made in the American legal arena, and how these stimulated a series of wider debates over a range of scientific issues. I also discuss arguments concerning the way in which DNA evidence has been presented to courts in the United Kingdom, and how this has led to controversies concerning the presentation of scientific and statistical information in general.

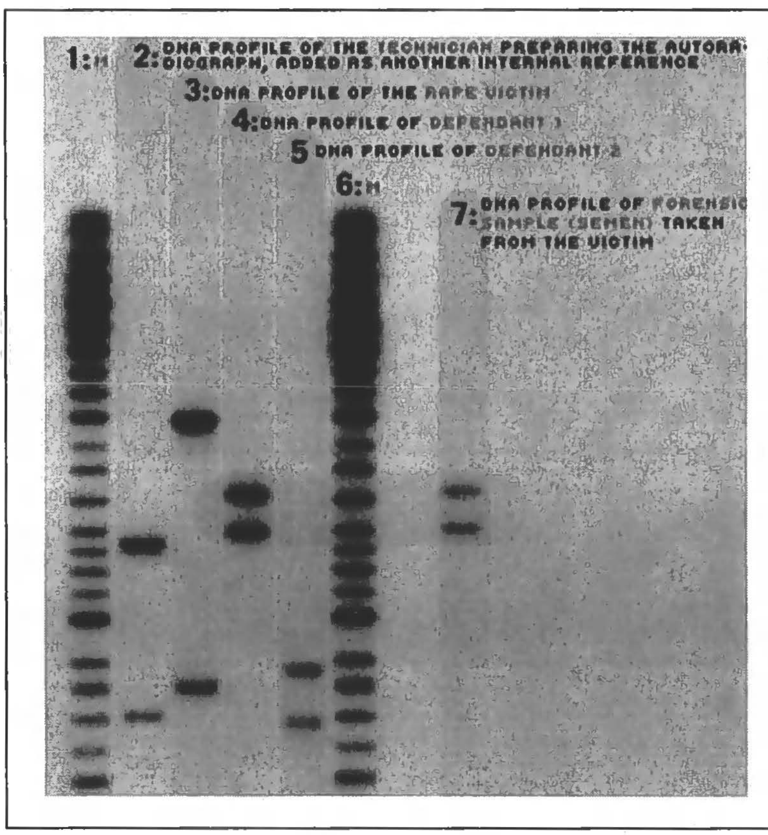
6.2 Initial Developments and Applications

The modern history of the forensic use of DNA is generally accepted to have begun with the work of Professor Sir Alec Jeffreys at the University of Leicester (Buckleton and Gill 2004, p.2). The technique, known initially as 'DNA fingerprinting', but which is referred to more widely as 'DNA profiling', originated largely as a corollary of that group's studies of 'non-coding' regions of the human DNA. Although DNA serves as the molecular blueprint for the construction of amino acids, and hence proteins, large areas of DNA have since been discovered where this function does not arise. The functional significance of these non-coding regions (known as 'introns', in contrast to 'exons', which do code for amino acids) has been, and continues to be, the subject of great interest (Fedorov et al 2003)[1]. Studies on the gene for myoglobin (the oxygen carrier protein in muscle) revealed the existence within it of an intron containing a DNA motif repeated several times over. Further work led to the identification and characterisation of a number of other similar repeat sequences, termed 'minisatellites', existing elsewhere within the genome (Jeffreys et al 1985).

DNA fingerprinting became a possibility once it was realised that the precise number of repeats within certain minisatellites showed considerable variation between individuals. Through the application of a technique known as Southern Blotting, it became possible to obtain unique patterns based upon the number and distribution of minisatellite repeats in individuals. If treated with restriction enzymes, which recognise and cleave specific DNA motifs, minisatellites yield fragments of DNA of varying size. These fragments can be identified via the use of 'primer' DNA, short sequences composed in such a way that they bind to minisatellite fragments. If labelled with a radioactive isotope or fluorescent chemical group, these primers can be used to effectively visualise the presence of fragments within some form of supporting matrix. If gel electrophoresis is used (see Figure 6.1), whereby the fragments are effectively separated on the basis of size, they are visualised as a series of

[1] The completion of the entire human genome sequence revealed that the overall proportion of non-coding DNA was even far greater than originally thought.

Figure 6.1. Gel electrophoresis of DNA fragments. **Bands represent fragments of DNA of a certain molecular weight. The column of bands in lanes 1 and 6 are 'standard ladders', sets of marker fragments used to assess the molecular weight of DNA from other samples. In the other lanes, bands which are horizontally level with each other are deemed to be of the same molecular weight. In a forensic DNA profile, this phenomenon would be deemed to indicate a 'match' between DNA alleles. (reproduced from http://www.americanheritage.com/assets/images/articles/magazine/it/2006/2/2006_2_54.jpg, Accessed 10 November 2008)**



bands arranged in such a fashion that they can be rendered as distinct or identical patterns.

The first legal application of DNA fingerprinting in its original form came about soon after the publication of Jeffrey's findings in 1985, where it proved to be instrumental in settling an immigration dispute involving a Ghanaian family living in the UK as British citizens. The youngest son, having travelled back to Ghana, was detained on his return to the UK under the suspicion of being an impostor. Using the new technique, the DNA of the boy under suspicion was compared with that of his supposed siblings and mother. Although no father was available for comparison, the results showed that the boy possessed a profile which both matched with that of his siblings (indicating common paternal ancestry), and that of the mother. The case against the son was subsequently dropped.

Around the same time as the case, Jeffreys had been collaborating with Dr Peter Gill and Dr Dave Werrett of the UK Forensic Science Service (FSS), and together they published the first scientific article outlining the forensic application of DNA fingerprinting (Gill et al 1985). This paper demonstrated how DNA could be extracted from crime scene stains, involving biological material such as semen. The utility of the technique came to further prominence through its use in the investigation of the rape-murders of two fifteen year old girls, Lynda Mann and Dawn Ashworth, which occurred in 1983 and 1985 respectively. Although the prime suspect, seventeen year old Richard Buckland, originally confessed to the murder of Ashworth, biological analysis showed that he did not share the same blood group as that of the person whose biological materials were recovered from the victim's body. Moreover, despite confessing to the attack on Ashworth, he denied any involvement in Mann's death (Williams et al 2004, p.17).

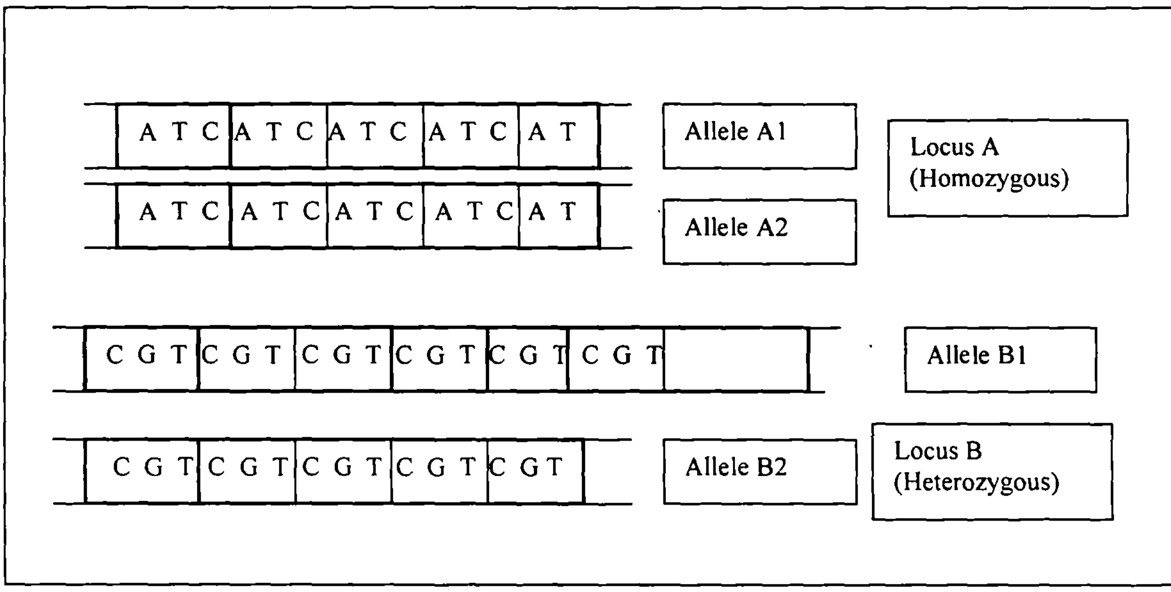
Police from Leicestershire constabulary, who were investigating the crimes, contacted Jeffreys about the possibility of using his technique in order to aid their inquiries. FSS scientists were also approached to further verify Jeffreys work (FSS 2007). A DNA profile from Buckland did not match the profile recovered from the crime scene, although subsequent analysis revealed that the crime scene profiles from each of the Ashworth and Mann cases matched each other. In January 1987, police initiated a mass screening of DNA from men aged 16 to 34 living in villages near to the crime scenes. One of these men, Ian Kelly, supplied his sample under the name of one Colin Pitchfork. However, having let slip that the former had deliberately misled the police, he and Pitchfork were arrested. Once the latter's DNA was sampled and profiled, it produced a conclusive match with the crime scene samples. Pitchfork was subsequently charged with the murders, convicted and sentenced to life imprisonment in 1988 (Williams et al 2004, p.18).

6.3 The Science of Forensic DNA

A forensic DNA profile consists of a number of fragments of an individual's DNA taken from specific sites, or *loci*, of the genome. Each locus is composed of a pair of 'alleles', two DNA strands which consist of a repeated series of pairs of short nucleotide sequences (nucleotides are four compounds, adenine, thymine, guanine and cytosine, normally represented as A, T, C and G. Together these nucleotides constitute the central 'building blocks' of DNA). Variation arises according to the different type of short sequences at each locus, and on the number of times a specific nucleotide sequence, or motif, is repeated at each allele (see Figure 6.2).

If treated with restriction enzymes, which recognise and cleave specific DNA motifs, minisatellites yield fragments of DNA of varying size which correspond to specific alleles. These fragments can be identified via the use of 'primer' DNA, short sequences composed in such a way that they bind to the allele fragments. If labelled with a radioactive isotope or fluorescent chemical group, these primers can be used to effectively visualise the presence of fragments. When a number of loci from an individual are analysed together, the resulting aggregrated data results in the generation of a profile which may be used to match an individual suspect with a DNA sample taken from a crime scene.

Figure 6.2 Variation in DNA sequences at non-coding loci. **Locus A depicts a *homozygous* pair of alleles, so called due to an identical number of the repeat sequence ATC. Each allele at Locus A is 4.2 base pairs in length. Locus B is thus *heterozygous* due to the difference in the number of the repeat sequence CGT at allele B1 (6 repeats) and allele B2 (5 repeats).**



6.3.1 Short Tandem Repeats (STRs) and the Polymerase Chain Reaction (PCR)

A major step forward in the development of the technology manifested itself in the discovery of 'microsatellite' DNA sequences, and the coupling of these with the application of the Polymerase Chain Reaction (PCR). Microsatellite sequences, also termed Short Tandem Repeats (STRs), typically exist as alleles of 100-400 base pairs (bp) in length, and are comprised of a series of repeat motifs of 2-8 bp (Buckleton and Gill 2004, p.7). Due to their smaller size in terms of both allele and repeat motif length, STRs proved themselves to be amenable to analysis via the relatively recent technique of PCR. This latter procedure, now well-established within molecular biology, provided the power to synthesise and replicate DNA sequences at a far greater speed and efficiency than before.

PCR reaction occurs via a number of steps. Firstly, the DNA is heated at high temperature in order to separate the original two strands. Once separated, the reaction temperature is lowered and DNA primers, consisting of short sequences able to selectively bind to the strands, form the site of synthesis of new copies of the DNA of interest, a process catalysed by the use of a heat-stable enzyme. Once new DNA is formed, the temperature is increased again, in order to separate the strands once more. More primer and enzyme can then be added to initiate another round of DNA synthesis. This cyclic process of heating to separate DNA strands, followed by the addition of primer and the production of new DNA, can be repeated numerous times, leading to the production of copies of a specific sequence of DNA at an exponential rate (Newton and Graham 1997).

The use of PCR conferred a number of methodological advantages to the DNA profiling technique. Being an automated process, analysis time was drastically reduced, to less than 24 hours. The shorter STR loci were relatively resistant to degradation, and were thus far more appropriate for forensic use, with crime scene samples being susceptible to a series of contaminating factors from the external environments from which they were recovered. More importantly, the use of STR in conjunction with PCR technology allowed unambiguous assignment of genotypes. DNA fragments that differed by one repeat could be amplified, analysed and visualised using gel electrophoresis. The highly sensitive nature of PCR also enabled the utilisation of extremely low amounts of DNA, equivalent to the quantity from a single cell (Findlay et al 1997).

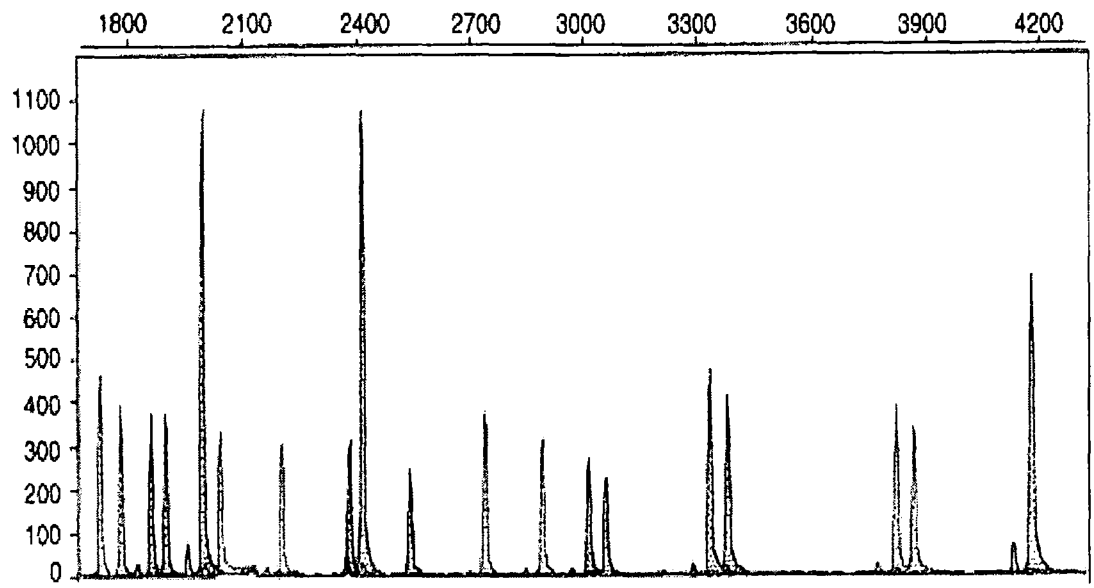
The ability to identify and target individual loci represented an important development, allowing analysts to visualise products with a far greater level of precision. However, this improved procedure, known as the single-locus probe (SLP) method in the UK, and as variable number of tandem repeats (VNTR) in the US, was hindered by other shortcomings related to instrumentation. For example, the procedure was reliant on the use of relatively primitive separation technology, in the form of agarose-based gels. Agarose, a naturally occurring substance, is disadvantaged in that the size of pores existing within gels (which act as passage for molecules) are large and often arranged in irregular fashion (Ream and Field 1999). This irregularity of pore distribution in agarose hindered the reliability of analyses, whilst the large pore size prevented fine resolution of alleles separated by one or a few repeat units, because of the tendency for these molecules to travel within the same pores. Due in part to the problems relating to resolution, fragments were measured by molecular weight rather than fragment length.

The introduction of polyacrylamide gel electrophoresis (PAGE) allowed greatly improved resolution of fragments. Polyacrylamide gels, being synthetic products, could be formed and manipulated with a greater degree of control over the shape and size of the pores within them. A more regular pore shape guaranteed more reliable analyses, and the ability to create smaller pores allowed separation of DNA fragments at the appropriate required level of resolution. Coupled with the use of 'standard ladders', consisting of DNA fragments of known length, samples could be identified on the basis of base pair length rather than molecular weight. Further improvements in instrumentation and visualisation techniques have enabled workers to create multiplexes of several different STR primer pairs, allowing the amplification of several loci at once (Gill and Buckleton 2005, p.8). This development was further enhanced by the introduction of dye-labelled PCR primers. Through the use of different dyes, PCR products could be characterised via the use of fluorescence detection coupled with an appropriate electrophoresis method (Balding 2005, p.44).

However, although gel technology improved, issues remained over the interpretation of DNA evidence using this method, as it meant that comparisons between DNA to determine possible matches were made on a largely visual basis. As I describe in another section of this chapter, this meant that a considerable degree of subjective judgement was involved in the comparison of profiles, and considerable controversy was generated over problems concerning measurement error. In many jurisdictions, the gel-based method has since been superseded by capillary electrophoresis (CE), which confers certain advantages, namely a significant reduction in the degree of subjective interpretation involved in the process (Balding 2005, p.44). In this case PCR products are characterised according to the time taken to emerge from a capillary or gel, rather than being identified on the basis of their location within the latter. The whole process can be recorded entirely digitally via relevant software packages, with alleles represented as a series of peaks. PCR products from several different loci can be separated on the basis of length (proportional to elution time), through the addition of size standards, and by locus through the use of different dyes. An example of a typical piece of CE data is depicted in Figure 6.3.

Figure 6.3. A typical DNA profile obtained via capillary electrophoresis.

Single peaks correspond to homozygous alleles, pairs of peaks to heterozygous peaks at particular loci (Petricevic et al 2005, 23).



Aside from improvements in instrumentation and other such technological developments, efforts have continued to enhance the potential evidential weight proffered by DNA profiling. Whilst some of these have involved new developments relating to statistical analysis of data, the introduction of increasing number of loci has also provided ever greater discriminatory capacity. In 1996 the second-generation multiplex (SGM) system was introduced within the UK, replacing SLP analysis. Originally SGM involved six loci plus the amelogenin sex test, but has since been superseded by SGM^+, which increased the number of non-sex loci to ten. The SGM systems utilise complex and highly polymorphic loci such as HUMD21S11 and HUMFIBRA/FGA, which has contributed to the potential to report match probabilities in the region of 10^{-10} to 10^{-13} (Gill and Buckleton 2005, pp.9-10).

The advent of improved multiple loci systems facilitated the introduction of national DNA databases. The UK has led the way in the implementation of forensic DNA databases, introducing the National DNA Database (NDNAD) in 1995. Using SGM^+ as a platform, the NDNAD now numbers over 3.5 million suspect samples. In the US, the Combined DNA Index System (CODIS) operated by the FBI contains over 2.8 million suspect samples and uses a 13-loci STR system. Numerous other countries have also established DNA databases of their own, and the spread of databasing has led to efforts to harmonise the use of certain loci across national jurisdictions (Kimpton et al 1995; Gill et al 2000).

The following paragraphs have provided a brief introduction to the development of the science of DNA profiling. In what follows I describe how challenges in the US courts led to a far deeper interrogation of the scientific basis of the technique. This section not only demonstrates the high level of complexity involved, but also indicates the relationship between the scientific, technological and statistical aspects of DNA profiling.

6.4 Probabilistic Assessment of DNA Profiles

The evidential weight of a DNA match is ascertained via the calculation of a *random match probability* (RMP), from which can be derived a *likelihood ratio*. The match probability is taken as the chance of an individuals DNA profile randomly matching with a crime scene sample. In order to calculate this figure, data are required concerning the frequency of each type of allele at each locus in a population. A match probability is typically obtained by multiplying the frequency of the occurrence of particular alleles at each locus. Match probabilities can range to as low as $< 1 \times 10^{-6}$.

The RMP is used by forensic scientists to assess the significance of a possible match between the DNA profile taken from a crime scene stain, and that of a possible suspect. Scientists use the RMP to help determine whether DNA profile data is able to make a meaningful contribution to judging the probability of whether the suspect is guilty or innocent. Two hypotheses are considered, with a 'prosecution hypothesis' normally taking the form 'the suspect is the offender', and a 'defence hypothesis' of the form 'a person other than the suspect is the offender'. The prosecution hypothesis, where the suspect is assumed to be the offender, will consider the probability of the individual's profile matching a crime scene stain to be 1. The defence hypothesis on the other hand, considers the probability that the suspect's profile could have matched at random, and it is this hypothesis which the RMP directly informs.

The two hypotheses are compared together, and it is accepted practice to generate *likelihood ratios* which utilise the RMP figure using the formulation of Bayes Theorem as introduced in Chapter 5

$$\frac{Pr(Hp \mid E)}{Pr(Hd \mid E)} = \frac{Pr(Hp)}{Pr(Hd)} \times \frac{Pr(E \mid (Hp)}{Pr(E \mid (Hd)}$$

For example, the prior measure of belief in a prosecution hypothesis in a burglary may be informed by eyewitness testimony. Hence one's prior belief in the guilt of a suspect accused of a burglary may be informed, *inter alia*, by testimony which claims that an individual, resembling the accused, was seen fleeing the area of a burglary shortly after an incident was reported. DNA evidence may then be presented which purports to show a match between the profile of the suspect, and DNA recovered from the scene. Although a likelihood ratio figure will have been calculated, this is normally translated into a verbal measure for ease of comprehension in court. If the DNA evidence is seen to be particularly incriminating for example, a scientist reporting in court may claim it provides 'very strong' support for the hypothesis that the suspect is the offender. The idea behind the use of such an approach is that the report of DNA evidence will contribute to the modulation of juror's beliefs. Although an explicitly numerical approach is not used in the courts, DNA evidence in this case is said to aid the estimation of a posterior probabilistic belief in the guilt of the suspect.

6.5 Legal Challenges to DNA Profiling in the USA

When DNA evidence was first introduced in the 1980s, it was portrayed by some as a largely immutable means of determining guilt or innocence. A positive DNA match between a suspect and a crime sample was often presumed to virtually amount to a guilty verdict (Derksen 2000). However, towards the end of the 1980s, a number of legal challenges were made in the US courts which served to question the validity of DNA profiling procedures. The resultant series of controversies, which engulfed the scientific community in the US, was coined 'The DNA Wars', and led to the publication of two reports by the US National Research Council. Prominent in leading the legal challenges were two criminal defence lawyers, Peter Neufeld and Barry Scheck. It was they who were able to mount the first successful challenge to the admissibility of DNA evidence (Derksen 2000).

In her study of these controversies, Derksen (2000) provides a detailed account of how Scheck and Neufeld made their first challenge during the *New York* vs *Castro* murder case (Derksen 2000). This case involved the murder of a pregnant woman and her two-year old daughter, who had been stabbed to death in their apartment in the Bronx, New York City. The woman's husband, who had discovered the bodies, told police he had seen a man leaving the building with what appeared to be blood on his hands. Joseph Castro, a caretaker's assistant, was later identified as this suspect, and during questioning detectives noticed a small bloodstain on his watch. Subsequent DNA testing of the watch, along with blood samples of the two victims, reported a match between the former and the blood of the adult victim. Lifecodes Corporation, the company who carried out the testing, reported that the probability of such a match occurring at random was 1 in 189,200,000 (Derksen 2000, p.807). Although the prosecution attempted to submit this result for evidence, the defence decided to challenge the figure via a pre-trial admissibility hearing.[2]

Scheck and Neufeld's justification for conducting an investigation into DNA testing procedures was based around the issue of the fact that Lifecodes was a private and unregulated firm, with no accountability. Concerned about the lack of transparency regarding the methods used by Lifecodes, Scheck and Neufeld obtained subpoenas to enable them to gain access to Lifecodes data, which included laboratory notebooks, computer printouts and electropherograms. They discovered that Lifecodes had reported one locus at the DNA bands of the watch stain and the victim sample as being exactly the same size, 10.25 kilobases (kb) long. However, on further examination of the original computer printout data, the watch stain locus had actually been recorded as 10.16 kb, whilst the corresponding victim stain result was 10.35 kb. Lifecodes had reported that it confirmed visual matches by measuring the bands and confirming a match by checking the bands fell within three standard errors of each other. This finding led Neufeld and Scheck to argue that Lifecodes had failed to follow their own matching rules, an argument strengthened by the subsequent discovery over similar discrepancies in the case of other band comparisons. Determination of matches was thus actually more dependent on visual, subjective judgements.

[2] Often referred to as a '*Frye*' hearing.

Furthermore, with the help of Dr Eric Lander, a geneticist and mathematician with appointments at MIT and Harvard, Neufeld and Scheck were also able to attack the scientific basis of Lifecodes statistical calculations, which involved a three-step method to calculate the random match probability (Derksen 2000, p.810). First, the frequency of each allele, across a certain population was calculated. The relevant population as determined by ethnicity, and in the Castro case a Hispanic database was used. Second, each locus was tested to check for Hardy-Weinberg equilibrium, which is used to ascertain that random mating is occurring in the population. Third, the allele frequencies at each locus, calculated in the first step, were multiplied together, with the assumption that there is no linkage of inheritance between loci. According to Lander (1989), none of these steps stood up to scientific scrutiny in the Castro case (Lander 1989, p.503). Although Lifecodes used a three-standard deviation error for matches between two samples ('forensic matches'), it used a far smaller standard deviation, two-thirds, when comparing allele bands within a population database. This had the dangerous effect, Lander argued, of erroneously slanting the random match probability in favour of the prosecution (Lander 1989, p.504). Furthermore, a test for the presence of the Hardy-Weinberg equilibrium in the Hispanic population database showed a significant deviation, and also indicated the presence of genetically distinct subgroups. Together these placed the assumption of no linkage between alleles under doubt, possibly invalidating the use of the product rule to multiply allele frequencies together to derive the random match probability.

Although Joseph Castro was eventually found guilty, Neufeld and Scheck's efforts led to the DNA evidence being ruled inadmissible, the first time this had occurred in the US. The case led to a number of similar challenges to the validity of DNA evidence, and eventually involved a number of leading academic figures.

Another case which featured prominently was *US vs. Yee et al.* The *Yee* case involved three members of a Hell's Angels motorcycle gang accused of murder. In this case, the reliability and reproducibility of DNA work carried out by the FBI came under scrutiny. In addition to questioning the effect of environmental insults and degradation on DNA, as well as issues concerning proficiency testing, the defence sought to challenge the calculations made by the FBI on the basis of the allele frequency databases they had developed and used. The FBI's Caucasian frequency database had used DNA samples taken from 225 of its own agents. However, the FBI had lost the original data, and had to take the samples again (Derksen 2000, p.818). In the original analysis, the FBI had used a technique known as 'binning'. Due to the limitations of the gel-based technology used at the time, the band fragments could not be accurately measured into discrete lengths. Instead, allele bands conforming to a certain measurement range were grouped together, and placed in 'bins' which conformed to a certain kb range, or group of lengths. As such, this amounted to the conversion of discrete values into continuous data, and then back again into an altered, approximated discrete format; specific alleles subsequently corresponded to certain bins.

A subsequent study by Fung (1996) attempted to demonstrate that, although the FBI claimed that the 'match window', (the allowed apparent measurement difference between two bands alleged to demonstrate a match), used by them was 5% (±2.5%), it was actually significantly greater in reality (Fung 1996). Fung's analysis indicated that the actual match window was around 10%, a figure with serious implications as this amounted to a value greater than the size of the bins used by the FBI to allocate allele sizes to molecular weight measurements.

On the second occasion however, the imprecision of the gel technique meant that different fragment lengths were recorded, which in turn meant that bands were classified into different bins, resulting in different frequency distributions from the same set of individuals. When asked in Court about the issue of calculating random match probabilities, Lander opined that no consensus existed with regard to a reliable method for deriving match probabilities (Derksen 2000, p.819). A second defence challenge raised questions over the nature of the Caucasian database used by the FBI. The use of the epithet 'Caucasian' was viewed as arbitrary, and it was argued that this masked the possibility of so-called ' population substructure', a phenomenon by which multiple alleles, the inheritance of each being supposedly independent of one another, are inherited together on the basis of ethnicity. The term 'Caucasian' was viewed as too broad, with the possibility that a number of ethnic sub-groups contained within the database could display substructural inheritance patterns.

The issue of substructure, and how it affected random match probability calculations, strongly divided the experts involved in the case. Defence experts pressed further the argument that probability estimates based on this database could not be considered reliable due to the presence of unrecognised population substructure. In response the prosecution experts argued that even if population substructure did exist, it was of a magnitude so small as to bias the defendant as much it might the prosecution (Derksen 2000, p.819). However, both sides were agreed that the extent to which North American and European Caucasian populations might be substructured by ethnic groups was unknown.

The work of Neufeld and Scheck in exposing the relative fallibility of DNA profiling led to a number of areas being subjected to further legal and scientific scrutiny. In the next section I discuss in further detail the science involved in these areas. I show how these debates touched upon statistical and probabilistic issues, discussed in two controversial reports commissioned by the US government. I continue by describing other debates which have impacted upon discussions of DNA database searches. In a subsequent section I consider the use of statistics and probability in forensic science, and I focus upon the use and presentation of statistics and probability theory in the UK courts.

6.6 Population Substructure

The occurrence of a particular set of alleles at an individual locus is generally taken to be an independent event. The generation of RMPs is calculated according to the product rule of probability, which states that the probability of a number of independent events occurring in conjunction is derived by multiplying the probabilities of each individual event. In the case of DNA profiling, the use of the product rule rests on certain assumptions, namely no co-inheritance of loci, and that random mating has occurred within the population of interest. These assumptions, especially the latter, have been subject to question ever since DNA profiling was first established as a widespread forensic technique. It is clearly evident that mating across the global population is not random, and instead there is a tendency for it to occur within certain ethnic and geographical subgroups. This being the case, the likelihood exists that certain alleles which might be more prevalent in certain subgroups, may become more preserved within that subgroup at the expense of less common alleles. In theory at least, there will be a tendency within that subgroup for certain alleles to tend toward homozygosity if inbreeding within that subgroup continues. Thus independence cannot necessarily be assumed for alleles, and has prompted considerations amongst forensic scientists regarding the phenomenon of population substructure.

Although it was obvious that random mating was not widespread, it was unclear whether substructuring had any truly significant impact on the calculation of match probabilities. Some geneticists argued that the true extent of population substructuring had to be assessed before assertions could be cast about the nature of match probabilities, and that this could only be achieved by detailed empirical studies (Lewontin and Hartl 1991). According to them, only specific sub-population databases could be used to confer accurate data. Opponents claimed that the generation of such data was unnecessarily time-consuming, and that such adjustments made little difference given the extremely low figures associated with match probabilities.

Furthermore, critics such as Eric Budowle of the FBI asserted that there were indeed no significantly large variations between allele sub-populations, and that the existing demographic and genetic conditions did not allow this to be the case. In response, Lewontin, Hartl and other colleagues pointed to an excess of homozygotes in populations as evidence of deviation from independent inheritance. The FBI's reply to them was rather unconvincing, and returned full cycle to the issue of unreliable measurement. Their rejoinder was that, due to the inability of gels to resolve bands fully, it only *appeared* that homozygotic alleles were apparent, when in fact they were just as likely to be heterozygotic (Derksen 2000, pp. 826-827).

The controversies over population substructure and other issues that followed the *Castro* case would eventually lead to two reports being published by the US National Research Council in an effort to settle the disputes. The first of these, often referred to as 'NRC 1', was published in 1992 (NRC 1992). NRC 1 concluded that population substructure had to be taken into serious consideration, and that measures had to be taken to account for any possible substructuring effects. Whilst proposing widespread sampling of ethnic groups to assess the issue, the report called for the application of a 'ceiling principle' to induce conservancy in genotype frequency estimates before the true effects of substructure were known. It was recommended that random samples of 100 individuals should be taken from 15-20 populations, 'each representing a group relatively homogeneous genetically' (NRC 1992, p.83). For each allele at each locus, the relative frequency was to be assessed, with the largest frequency in any of these populations being used in the calculation of genotype frequencies, provided that it was greater than 5%. If this did not happen to be the case, then 5% would be used instead. In the absence of the necessary population work, an 'interim' ceiling approach was also put forward whereby population studies could be carried out 'on at least three major "races"', with the data being used to calculate allele frequencies provided it met Hardy-Weinberg (HW) and linkage disequilibrium (LE) criteria. The NRC claimed that these methods conferred certain advantages. The calculation of high-frequency alleles did not, it was claimed, require large population samples. The ceiling principle also claimed to circumvent the problem of the existence of rare alleles which were possibly more susceptible to genetic drift induced by substructuring effects (NRC 1992, pp.83-84).

In itself, the ceiling principle was perhaps redolent of a self-conscious lack of knowledge about population substructure, as well as a recognition of the limitations of the technology as it existed at that time (the SLP platform in widespread use at the time relied on less loci than modern systems).

Nevertheless, both the ceiling principle and the application of confidence intervals were subject to vehement criticism, most notably in the second NRC report ('NRC 2'), published in 1996 (NRC 1996). NRC 2 acknowledged a number of inherent flaws present in the first report. The ceiling principles were seen as excessively conservative, and more seriously still, the 0.05% threshold value was attacked for being a completely arbitrary figure with no scientific justification (NRC 1996, p.157). Furthermore, NRC 2 criticised the lack of acknowledgement, on the part of the previous authors, of 'standard procedures long used by population geneticists to study subdivided populations' (NRC 1996, p.157)[3]. There was also reference to the poorly specified guidance concerning which population groups to include in the calculation of ceiling frequencies. With regard to database size, NRC 2 also pointed out to the inherent circularity in the use of small databases to check for HW conformity and linkage disequilibrium (NRC 1996, p.159).

In response to this, NRC 2 recommended a more sophisticated theoretical approach. The report saw the use of the inbreeding coefficient, θ. This value gave an indication of the tendency towards 'fixity' of an allele in a given population, i.e. the extent to which homozygosity was established in the population. This allowed for the possibility for the occurrence of mating within distinct subpopulations. Thus NRC 2 proposed a new formula for determining allele frequencies of homozygotic loci which accounted for the tendency toward fixity. This formula however, only applied where precise genotypes were ascertained; the primitive gel technology in general use at the time had difficulty in resolving heterozygotic alleles of similar MW. These often appeared as a single band and could be easily mistaken for a homozygote. In cases where resolution was problematic, NRC 2 recommended certain mathematical adjustments to ensure a conservative estimate.

Furthermore, the use of confidence intervals, in conjunction with the ceiling principles as suggested in NRC I, were considered unsuitable for use with the

[3] Without specifying exactly what was meant by ' standard procedures' (NRC 1996, p.157).

PCR-based systems that were emerging during the time of publication of NRC 2. The use of SLP systems involved loci each exhibiting several allele types, none of which existed with high frequency. In contrast, PCR-based systems involving smaller allele distributions at loci produced larger allele frequencies that added up to >1 given the application of upper confidence limits, which also led to heterozygote frequencies exceeding the HW threshold of 0.5 (NRC 1996, p.158). This was in addition to other criticisms of basic errors made in the application of confidence intervals. For example, the erroneous multiplication of upper 95% confidence limits for each allele frequency was also highlighted, for as according to Weir, 'confidence limits for products are not obtained as products of confidence limits' (quoted in Buckleton and Curran 2004, p.201).

Although the NRC 2 solution was regarded as being grounded on sound mathematical principles[4], it was deemed unsuitable for use with small databases due to the inversely proportional relationship between range of confidence interval on a logarithmic scale and square root of database population *(N)*. It was also questioned whether the method would be amenable to use with additional loci due to the additional variability that this adduces to the estimate, leading to a wider log scale interval (NRC 1996, p.146). However, the logarithmic multiplication method[5] was ultimately seen to suffer from a lack of consideration of other sources of uncertainty, e.g. the effects of population subdivision.

In the light of this NRC 2 proposed an alternative method, which can be summarised as thus:

> 'Within a racial group, geographic origin and ethnic composition have very little effect on the frequencies of forensic DNA profiles, although there are larger differences between major groups (races). It is probably safe to assume that within a race, the uncertainty of a value calculated from adequate databases (at least several hundred persons) by the product rule is within a factor of about 10 above and below the true value. If the calculated profile probability is very small, the uncertainty can

[4] This method is still applied in certain jurisdictions in Australia (Buckleton and Curran 2004, p.202).
[5] Referred to as Asymptotic Normality of the Logarithm in Buckleton and Curran (2004, pp.202-204).

> be larger, but even a large relative error will not change the conclusion.' (NRC 1996, p.156).

Here the authors of the NRC report appear to have taken a substantially different direction. Disregarding matters of statistical theory, this conclusion was reached not by calculation, but through empirical means. Comparisons were made of genotype frequencies between and within 'races'. Individual genotype frequencies were plotted, using a logarithmic scale, on scatter graphs, with abscissae and ordinates representing differing racial, ethnic and geographic groups. These graphs, the report argued, proved that genotype frequencies gathered around a factor of $\pm 10^1$ if the two populations analysed were of the same racial composition, but varied by a greater factor if they differed (NRC 1996, pp. 151-156). This effect was observed, it was claimed, regardless of whether or not issues such as potential substructuring were taken into account.

In addition to the obvious lack of theoretical foundation, this proposition rested on relatively little in the way of scientific evidence. The claims made with regard to racial differences in NRC 2 are made largely on the basis of comparisons within and between white and black populations, and it is not clear how large the study populations were. As such it can clearly be seen to serve as a set of assumptions rather than a systematic methodology.[6]

From this brief discussion of some of the controversies which raged in the US during the 1990s, it is possible to see how these were stimulated due to interactions between the realms of law, science and statistics. Prior to the *Castro* case, DNA profiling was largely considered to be an immutable means of establishing guilt or innocence, so much so that a successful DNA match was deemed to be akin to a guilty verdict. The intervention of lawyers Neufeld and Scheck however led to the opening up of a series of discussions which eventually encompassed complex scientific issues. As described by Derksen (2000), their investigations into the procedures used by Lifecodes Inc

[6] The leading journal *Forensic Science International* has published genetic data on a wide variety of 'population groups', a practice continued by its offshoot journal *FSI:Genetics*.

exposed a number of discrepancies involved in supposedly 'objective' scientific methods (Derksen 2000). Some of these shortcomings were related to technical limitations, with analysts having to use measurement error to overcome the lack of accuracy inherent in the visual gel-based system of assessment. These limitations however impacted strongly upon the calculation of match probabilities. Such figures required databases of DNA profiles and allele frequencies, yet the gel system caused serious problems with the statistical manipulation of such data. For example, the organisation of allele populations into 'bins' was found to be nonrepeatable due to the lack of the accuracy of gel electrophoresis.

Legal questioning also raised issues with the construction of databases used to generate match probabilities. Possibly reflecting certain assumptions about the relationship between race and genetics, the construction of a 'Caucasian' database by the FBI led to a further debate about the possibility of alleles being 'substructured' on ethnic lines, sparking a heated debate in scientific and forensic circles. The report which arose from this controversy, NRC 1, recommended a statistical approach which was heavily criticised for error and arbitrariness, and only served to compound the situation further. The problems with NRC 1 may have been partially caused by the swiftly-changing nature of scientific opinion at the time, but, as Derksen (2000) reports, extra-scientific factors may have also served to shape the final report, which was authored by a committee of leading scientific figures (Derksen 2000, p.823). Personality clashes, time and resource constraints, and the pressure of the need to make the report comprehendable to non-scientists may have all contributed to the outcome of NRC 1.

Although generally regarded as an effort to ameliorate the mistakes made in NRC1, NRC 2 established a distinction between mathematical and empirical attempts to establish confidence intervals. Although it recommended a sophisticated mathematical solution, it also conceded that this was of limited use, depending on the size of the population database. Hence it also proposed a more practical method involving empirical analysis. This led to discussions over the efficacy of each system. Whilst the empirical method has been criticised for being less scientifically sound, the mathematical method has been considered by some to be too conservative for forensic use, in that it may be biased toward the defence position. Hence even through these discussions, technical, statistical and legal concerns continue to intermingle.

The story of the modern development of DNA evidence is one in which probability theory and statistics continues to play a major role, and has arguably helped to drive the development of probability theory in the service of others areas of evidence interpretation. Central to these developments is the ongoing involvement of the Bayesian approach to probability. With reference to controversies raised in the UK courts, I therefore continue to describe the impact of Bayes in discussions concerning statistics and probability in the context of forensic science and criminal justice.

6.7 Probability and the Courts

Challenges to DNA evidence in UK appeal cases have concerned both alleged scientific failures and flawed presentations of evidence (Williams et al 2004). In *R vs Gordon*, concerns over statistical reasoning were accompanied by challenges to the laboratory method (Williams et al 2004, p.22). Gordon's conviction had been based on DNA evidence which had seemingly established a match between two profiles obtained from two separate semen samples from rape cases. The appeal argument attempted to refute this evidence by claiming that the gel technology used to assess the evidence was seriously flawed. Although the distribution of bands from both crime scene samples was found to precisely match, only a series of bands matched those of Gordon. The original, supposedly incriminating match, had been made using a procedure which had incorporated a degree of measurement error between the bands that constituted the profile (Williams et al 2004, p.23). This was allowed by scientists due to the occurrence of stochastic variations in the length measurements represented by the bands each time the procedure was carried out. Gordon's appeal argument was that the degree of measurement error was higher than the 'window' that was commonly allowed. In response, the scientist who carried out the measurement agreed about the matter of degree, but claimed that the previous variation criteria had been too strict, and hence the profiles could still be considered to match (Williams et al 2004, p.23).

UK law has also been faced with a number of controversies relating to the presentation of statistical evidence in court, often involving DNA. One particular problem in this regard concerns the 'prosecutor's fallacy', a term made use of in a paper published by Thompson and Schumann (1987). The prosecutor's fallacy is essentially a straightforward example of a logical fallacy known as 'the transposition of the conditional' (Thompson and Schumann 1987). Thompson and Schumann explain the reasoning behind the prosecutor's fallacy by presenting the following example:

> 'Suppose you are asked to judge the probability a man is lawyer based on the fact he owns a briefcase. Let us assume all lawyers own a briefcase but only one person in ten in the general population owns a briefcase. Following the prosecutor's logic, you would jump to the conclusions that there is a 90% chance the man is a lawyer. But this conclusion is obviously wrong. We know that the number of nonlawyers is many times greater than the number of lawyers. Hence lawyers are outnumbered by briefcase owners who are not lawyers (and a given briefcase owner is more likely to be a nonlawyer than a lawyer). *To draw conclusions about the probability the man is a lawyer based on the fact he owns a briefcase, we must consider not just the incidence of briefcase ownership, but also the a priori likelihood of being a lawyer.*' (Thompson and Schumann 1987, p.170, emphasis added)'.

Hence, the fallacy occurs when the probability of occurrence of a single piece of evidence (the ownership of a briefcase), is used as a measure of the probability of a hypothesis (that the owner is a lawyer). The fallacy reflects the lack, on the part of the reasoner, to take into consideration other pieces of evidence. Or, as it was expressed in more simple terms by one interviewee:

> 'All elephants have four legs, but if we see a four-legged animal it would be wrong to assume it is an elephant on that basis alone!' (Interviewee 6B)

In a forensic context, the prosecutor's fallacy therefore represents instances in which a single piece of evidence is directly associated with the probability of guilt of a defendant. Using forensic nomenclature, this can be considered as a belief that the probability of a piece of evidence given a guilty hypothesis P(E|Hp), is the same as the probability of guilt given a piece of evidence, P(Hp|E).

The prosecutor's fallacy has become strongly associated with DNA evidence, due in part to the magnitude of RMPs which are capable of being generated using modern methods. It formed the basis of a notable controversy surrounding *R v Deen*. In this case, evidence was presented in which the RMP of DNA evidence was given as one in three million. The evidence was presented by the prosecution thus:

> 'Counsel: So the likelihood of this being any other man but Andrew Deen is one in three million?
> Expert: In three million, yes.' (Redmayne 2001, p.58).

The primary objection advanced in this case concerned the ambiguity between the probability that the defendant's DNA matched the crime scene profile, and the probability that the defendant had actually left DNA at the crime scene (Williams et al 2004, p.22). The first calculation involved the random match probability, whilst the second was actually a measure of *likelihood* that Deen had deposited the crime scene stain, generated using a Bayesian schema. The crux of the argument was that, in the exchange quoted above, the *random match probability* figure had been used to answer the question about the *likelihood* of Deen's guilt. Hence, a single piece of evidence (the RMP), had been used to prove the hypothesis that Deen had been the depositor of the crime scene stain, without taking any other circumstances into account. In other words, what should have been regarded as pertaining to 'E' in the above formula (the RMP), was actually confused for 'P(E|Hp)'. This, the defence argued, was an instance of the prosecutor's fallacy.

The controversy over the prosecutor's fallacy also came to light in *R v Doheny and Adams* (1996), a case in which two men were charged with rape. DNA match evidence was presented against one of the accused, Alan Doheny, returned an RMP of 1 in 40 million. The following exchange took place:

> 'Reporting forensic scientist: ...I calculated the chance of finding all of these (matching) bands...to be about 1 in 40 million.
> Prosecution counsel: The likelihood of it being anyone other than Alan Doheny?
> Reporting forensic scientist: About 1 in 40 million.'
> (Evett 2005)

Likewise another DNA match was presented against Adams, the co-accused:

> 'Prosecution counsel: Is it possible that the semen could have come from a different person from the person who provided the blood sample?
> Reporting forensic scientist: It is possible but it so unlikely as to really not be credible...I can estimate the chances of this semen having come from a man other than the provider of the blood sample...I can work out the chances as being less than 1 in 27 million' (Evett 2005)

The convictions of both men were brought to appeal, on the basis of 'the misleading and inaccurate manner in which forensic evidence was presented, in particular the way in which the random occurrence ratio, the frequency with which the matching characteristics were likely to be found in the population at large, was expressed' (*R vs Doheny and Adams* 1996, quoted in Williams et al 2004, p.26).

Doheny's conviction was quashed on the grounds that 'the approach demonstrated by the prosecution's expert was not legitimate and the conviction was unsafe.' The decision led the Appeal Court to make specific recommendations for how match probabilities, or 'frequency calculations' should be presented to the Court:

> 'The scientist should not be asked his opinion on the likelihood that it was the Defendant who left the crime stain, nor when giving evidence should he use terminology which may lead the jury to believe that he is expressing such an opinion' (*R v Doheny and Adams* 1996, quoted in Williams et al 2004, p.26).

Doheny also ruled that, in strictly statistical terms, the match probabilities associated with DNA evidence will incriminate a small number of the population, including the suspect. The role of the jury therefore, was defined as determining whether or not the additional evidence inculpates the accused to an extent that transcends any reasonable doubt. The Appeal Court produced a template for the presentation of DNA evidence to juries:

> 'Members of the jury, if you accept the scientific evidence called by the Crown, this indicates that there are probably only four or five males in the United Kingdom from whom that semen stain could have come. The Defendant is one of them. If that is the position, the decision you have to reach, on all of the evidence, is whether you are sure that it was the Defendant who left the stain or whether it is possible that it was one of that other small group of men who share the same DNA characteristics'. (Williams et al 2004, p.26).

The phrase 'on all of the evidence' is informative here, as this indicates the sense that juries cannot convict purely on positive DNA evidence alone, nor should it be regarded as providing a definitive match between the suspect and the crime scene sample.

Another notable controversy concerning the use of statistics in court concerned the trial, and subsequent appeals, of Dennis John Adams (*R vs Dennis John Adams*). These trials were notable for the employment, by the defence, of a strategy which explicitly used Bayesian reasoning. Lynch and MacNally (2003) present a history of this case, which I now draw upon to describe the details of the method and the controversies it provoked.

Dennis Adams was originally convicted in 1995 of the rape of 'Miss M', which had occurred two years previously. A crime scene stain profile had been recovered and was stored on a database held by the Metropolitan Police Forensic Science Laboratory (now part of the FSS). This was found to match with the DNA of Dennis Adams which was sampled following his arrest for another sexual offence.

The original trial took place in January 1995. The DNA evidence presented in court reported that Dennis Adams was 200 million times more likely to be the perpetrator than any other male. This statistic was, however, virtually the only piece of evidence put forward by the prosecution, and was countered by a relatively strong defence case (Lynch and MacNally 2003). Miss M had claimed that her assailant had been a Caucasian man aged between 20 and 25 years at the time of the attack, whereas Dennis Adams was 37. Miss M had failed to identify the suspect in a police lineup, and subsequently said that Dennis Adams did not resemble her attacker, and looked significantly older than the offender. Furthermore, Dennis Adams claimed to have an alibi, saying he had spent that night with his girlfriend, who in turn corroborated this statement.

In the trial, the defence strategy involved two main areas. First, they questioned the accuracy of the gel autoradiogram on which the profile was generated, and the ability of forensic scientists to generate the likelihood figure based on the data. Second, they adopted a novel line of argument, in which they attempted to use Bayes Theorem to show that the prosecution had failed to take all the details pertaining to the case into account when calculating their likelihood statistic. They wanted the jury to use Bayesian methodology for translating all the evidence relevant to the case into probability estimates. The defence justified this strategy by claiming that all of the evidence be treated the same. They claimed that, in using Bayes Theorem to calculate their likelihood statistic based on DNA evidence, the prosecution had themselves led the way for the Bayesian approach to evidence. The defence however advanced the argument that the prosecution approach was flawed and incomplete, in that they had used Bayes only to calculate a likelihood measure based on the DNA evidence alone. The defence sought to show how Bayes, by accounting for all the information relevant to the case, could be used to demonstrate that the likelihood of Dennis Adams was actually significantly reduced, well below the threshold of reasonable doubt.

The defence recruited Professor Peter Donnelly, a statistician then at Queen Mary and Westfield College, University of London. Donnelly developed a questionnaire that jurors could use in order to guide them in Bayesian calculation. This prompted the juror to make four probability estimates (Lynch and McNally 2003, p.91):

1) The probability that a local man committed the offence
2) The probability that the victim would not have identified the offender
3) The probability that Adams would have given the evidence he did in favour of his innocence
4) The probability that Adams would have been able to call the alibi evidence on his behalf

Donnelly did not instruct the jurors in which figures to use in making their estimates, but presented information for the purposes of illustration. The questionnaire consisted of the following questions (Lynch and McNally 2003, p.91):

1) *What is the chance, assuming nothing else about the case, that the rapist came from the local area?*

Donnelly cited local council data that 153,000 men between ages of 18 and 60 lived within a 15-kilometer radius of the crime scene, a figure rounded down to 150,000 for ease of calculation. He then gave an example to show the effects of assuming a 75 percent chance that the attacker was a local man, dividing 150,000 by 0.75, resulting in a figure of 200,000. According to the facts of the case, Adams was a local man, so that *in the absence of any other evidence* the odds of his being the man who deposited the semen sample recovered during the victim's examination were 200,000 to one.

2) *What is the chance that the victim would fail to identify Adams?*
This part of the questionnaire consisted of two relevant sub-questions. One asked the jurors to state the probability that, if Adams were innocent, he would not match the victim's description. The second question obliged jurors to state the probability that, if Adams were guilty, he would not match the victim's description.

The key issue concerned the ratio between these two figures, for these corresponded to a likelihood ratio of the victim's eyewitness evidence between the defence and prosecution hypotheses. Again, by way of illustration, Donnelly provided an example whereby he assumed that, if innocent, there would be a 90 percent chance that he would not match the victim's description, but a 10 percent that he would not match the description if he were guilty. According to the Bayesian logic of the questionnaire, this ratio was to be multiplied by the ratio given for the first item, 1 in 200,000, which represented the prior odds, to produce a posterior probability estimate of guilt reflecting the combination of these two pieces of evidence. In Donnelly's example, the multiplication of the two figures returned a figure of 1 to 1.18 million.

3) *What is the chance of the defendant's evidence?*

In his example estimate, Donnelly assumed that there was an equal chance that a guilty or innocent defendant would give the same evidence that Adams had done, which returned a 1:1 ratio, hence exerting no effect on the figures based on the first two issues.

4. *What is the chance of the alibi evidence*?

Again this question involved a consideration of defence and prosecution hypotheses. Donnelly assumed a 25 percent probability that a guilty defendant would be able to produce the alibi evidence in this case, and a 50 percent chance that an innocent defendant would produce the same evidence. According to this example, it would therefore be twice as likely that Adams would have the alibi evidence if he were innocent, giving a ratio of 1 to 2. Multiplication of 1 to 1.18 million ratio by 1 to 2 returns 1 to 3.6 million. Hence, Donnelly put forward a probability of
1:3.6 million on the basis of the non-DNA evidence alone.

This figure, of 1 in 3.6 million in favour of guilt, was then used as a prior probability to be combined with the likelihood of guilt from DNA evidence, cited as 200 million. A multiplication of these two figures returned a posterior probability of *innocence* of 1 in 55; hence Dennis Adams was taken to be 55 times more likely to be guilty than innocent. Questions concerning the gel autoradiogram were then taken into consideration. One of the bands from the crime scene profile that had allegedly matched with Dennis Adams' gel data had been questioned by the defence, due to its indistinct appearance. Donnelly argued that if this particular band was ignored, the likelihood figure could be seen to reduce by a factor of ten, to 20 million, which in turn would render Dennis Adams only 5.5 times more likely to be guilty than innocent (Lynch and McNally 2003, p.92). Donnelly went on to argue that another discrepancy in the DNA profile evidence could force the prosecution figure down to 2 million, returning a figure of approximately one in 2 that Dennis Adams was guilty. The defence even went as far to argue that if the fact that Dennis Adams had a brother was taken into consideration, he actually became more likely to be innocent, as this turned the probability of *guilt* into 1 in 18[7]

Despite this approach, Dennis Adams was convicted of the rape of Miss M. An appeal was launched, on two grounds, namely that DNA evidence alone was insufficient to prove guilt, and that the original trial judge had inadequately instructed the jury in the use of the Bayesian defence. Whilst the first argument was accepted as suitable grounds for appeal; the second argument, concerning the Bayesian defence, was however rejected. The appeal judge criticised the way in which the defence had used Bayes:

> '…the mathematical formula, applied to each separate piece of evidence, is simply inappropriate to the jury's task. Jurors evaluate evidence and reach a conclusion not only by means of a formula,
> Mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them.'

This despite the defence's assertion that the Bayesian procedure enabled 'common sense' (or non-DNA) evidence to be weighed on the same scale as ostensibly 'scientific' DNA evidence.

[7] See Balding for a discussion of how relatedness affects likelihood calculations based on DNA evidence.

During the re-trial, the defence once again utilised the same Bayesian methodology, but once more a guilty verdict was returned. The defence immediately pressed for another appeal. This time however, their arguments were rejected. In making the rejection, the Court proclaimed that the Bayesian method was only suitable for DNA evidence, claiming that this involved 'empirical statistical data', as opposed to the seemingly 'non-scientific' data put forward by Donnelly.

In drawing such a distinction, the Court's decision can be viewed as arbitrary. The Court's argument was that 'non-scientific', 'commonsensical' evidence was being given a scientific veneer, and that this 'common sense' evidence should be assessed via suitably intuitive and consensual methods. A 'scientific' technique, such as DNA profiling, was however entirely suited to quantitative Bayesian analysis.

In the UK, it may appear that the rulings on *Deen, Doheny and Adams*, and *Dennis John Adams* have served to act as closure on issues relating to the presentation of statistical evidence in court. The issue of the reporting of statistical evidence has, however, been debated in the literature.

For example, Redmayne (2002) argues that the uptake of a statistical mode of thought has serious consequences for conceptions of evidence and proof. He claims that the use of probabilistic reporting enables the gap separating evidence of identity and proof of identity to be fully revealed, as opposed to eyewitness testimony where it is not instantly clear whether such a gap can be identified. Moreover, Redmayne points out that the *Doheny* approach presupposes that a uniform prior probability is assigned to all those possibly incriminated by DNA evidence, an assumption that can be easily taken to be fallacious. Redmayne counters by suggesting that if juries were told that an accused's profile already existed on a criminal database, it might influence their thinking. However, this points to a paradox. Taken to a logical extreme, the use of statistical reasoning potentially negates any hope of reaching proof. Redmayne cites Rawling (1999) in arguing that if a full Bayesian approach were ever to be adopted, extremely high likelihood ratios would need to be generated in order to pass the high threshold set by the concept of reasonable doubt. Redmayne views jurors as requiring certain presuppositions to be able to make an informed decision (such as trust in the police to arrest an individual who has a high chance of being the offender), but this too is problematic, for this invites charges of irrationality.

Regardless of this, even if a certain mode of reasoning such as probabilism is adopted, Redmayne argues that it may yet be underdetermined in the face of more naturally intuitive forms of reasoning. An example here is given involving DNA evidence. Redmayne equates a suspect being arrested via a DNA database search (as opposed to a one-to-one match between the crime scene profile and a suspect profile), with the arrest of a suspect on some other basis – such as suspicion due to previous form. In both cases, if the defendant decides not to reveal the reason behind his arrest (database search or previous form), with a DNA match the only evidence against him, then the jury may become more suspicious that there is undisclosed evidence against him. Otherwise, it could be considered untoward that the police would arrest an individual by random on the basis of a DNA match. According to Redmayne, the fact that a database search might raise (possibly unjustified) suspicion underpins the argument in favour of deflating the DNA evidence. However, this argument is not what Redmayne primarily focuses on. Redmayne's claim is that the whole debate over database search effects does not rest on a level where logical analysis can be taken to settle the matter, and that instead the argument concerns claims over *how* juries are likely to reason. (Redmayne 2003, pp.881-882).

Redmayne ultimately raises the issue over to what extent we allow a particular mode of reasoning to guide our way of acquiring knowledge, and when we should instead rely on intuitive or experiential knowledge (Redmayne 2002, 2003). He argues that no concrete rules can be ascribed. If this is so, such that any attempt to apply normative standards is inevitably relative, this may indicate that the roles of specific technologies, existing as they do at the intersection between differing epistemologies, are instrumental to the eventual outcome. DNA evidence can be taken to represent a site whereby the epistemic interests of both forensic science and the legal realm have to be accommodated.

6.8 Conclusion

Whilst the probabilism debate within evidence studies has remained fairly esoteric in nature, the use of statistics and probability theory has found ever increasing application within the practice of forensic science. In particular, the field has seen the rise in popularity of the Bayesian approach to probability theory. The enthusiastic embrace of Bayes by forensic scientists and by a section of evidence scholars appears to have created a tension between them and members of the judiciary, who have been somewhat less than forthcoming to grant Bayesian reasoning any further foothold in the courtroom.

Nevertheless, supporters of the adoption of probabilistic reasoning in court cases, and particularly those of a Bayesian persuasion, have continued to state their case, and have often adopted something amounting to a fundamentalist position. The prescriptive stance taken by many Bayesian evidence scholars has naturally provoked close scrutiny not only of the philosophical and logical basis of the position, but also over the practical issues involved in the use of Bayes in the course of the legal process (Jackson 1996). Redmayne (1996) views Bayesianism as a useful analytical tool for forensic science, but argues that any wholesale application of Bayes is problematic. Redmayne argues that for Bayesianism to work in practice, one needs to simplify and downplay the complexities of the real world: '...one of the values of the [Bayesian] model lies in its ability to uncover some of the very complexities with which it is poorly suited to coping' (Redmayne 1996, p.760) Although the Bayesian approach is useful for comparing two diametrically opposed hypotheses, he argues that prosecution and defence case may not necessarily be defined in such idealised terms.

Regardless of such debates, it seems that the advent of DNA evidence has reinforced the pre-eminence of Bayes amongst forensic scientists. Most of the methods currently used for the interpretation of DNA evidence utilise a

Bayesian epistemology, and it is highly likely that future technologies in this area will use Bayes in ever more expansive ways (Taroni et al 2006). With DNA evidence serving as a means of perpetuating the use of Bayes, its immediate future in the context of forensic science is most probably assured, but it is clear that there remains no place for it in the procedures of the courts.

In describing the controversies experienced in the US and UK in the 1990s, I have also begun to demonstrate the interdependence that can be identified between the domains of science, statistics and law. In the US, judicial challenges to DNA profiling led to the wholesale questioning not just of the technology and practices involved in the technique itself, but eventually sparked a series of debates which touched upon complex and esoteric areas within the realms of population genetics and statistics. Legal challenges in the UK have centred on the presentation of DNA match statistics in court. Cases such as *R vs Deen*, and *R vs Doheny and Adams*, have introduced fine-grained logical discussions centring on the concept of the 'prosecutor's fallacy', and have led to the successful challenge of seemingly unassailable DNA evidence. The case of *R vs Dennis John Adams* however, perhaps indicates an important area of uncertainty and contestation; here a statistical approach was seen as unnecessarily impinging upon areas considered the preserve of jury reasoning.

The interdependencies apparent in these examples reinforce the notion of the impossibility of isolating science from a wider set of social worlds. The legal challenges described above led to a certain unravelling of supposedly immutable claims. Moreover, they show how this process of unravelling may occur in numerous ways; either by recourse to a linked series of scientific issues, or in regard to the manner in which this evidence may be presented. In this way, juridico-scientific inquiry can be seen to be a performed and constructed achievement, consisting, essentially, of a process of marshalling reality.

CHAPTER SEVEN – CASE STUDY 1: AUTOMATED DNA EVIDENCE INTERPRETATION SYSTEMS

7.1 Introduction

On 20th December 2007, Sean Gerard Hoey was found not guilty of all fifty-six criminal charges that were brought against him in relation to his alleged role in the bomb attack which took place in Omagh, Northern Ireland, in August 1998. This included twenty-nine counts of murder corresponding to the number of victims of the attack, which caused the largest single loss of life during the so-called 'Troubles'. The prosecution case during the trial, which lasted over a year, was heavily reliant on forensic evidence. Particular prominence was given to DNA evidence generated from a technique known as 'Low Copy Number' (LCN) DNA profiling, which the prosecution argued provided a match between the DNA of Hoey and minute quantities of genetic material supposedly found on the bomb timers used in the attack. LCN DNA analysis had been employed by the Forensic Science Service (FSS) over 21,000 times previously, and offered as evidence in over 40 cases. It was claimed that the technique could facilitate the construction of DNA profiles from material extracted from of a small number of human cells, or potentially even a single cell (Findlay 1997).

However, over the course of the trial, the LCN DNA evidence, as well as other aspects of forensic evidence relevant to the prosecution case, was subject to a rigorous challenge from the defence. In particular, experts testifying for the defence questioned the scientific basis of the LCN technique. Speaking for the defence, Professor Dan Krane from Wright State University, Ohio, argued that the results of LCN DNA testing were susceptible to a far greater level of subjective interpretation than conventional forensic DNA analysis. He also voiced concerns over the relative ease in which LCN DNA samples could become contaminated (BBC News 2007). Cross-examination of DNA scientists called by the prosecution led to one expert admitting that some of the supposedly incriminating results were 'valueless', and that the area of LCN was susceptible to a high degree of ambiguity (BBC News 2006d).

In his criticism of the manner in which LCN had been presented in court, and in particular his rejection of the existence of two peer-reviewed articles as constituting sufficient ground for validation of the technique, the presiding Judge Weir highlighted a number of highly significant issues in contemporary forensic practice. First, his summation highlights the continuing existence of a gulf between the realms of science and UK law, with regard to the issue of scientific admissibility. Not only did he acknowledge the inability of judges to make sole decisions over the admissibility of evidence, but his judgement also exposed differences concerning the means by which the worlds of science and law determine the 'truth' of scientific evidence. Moreover, the trial has also served to expose considerable differences of opinion within the relevant scientific community itself. In this instance, Judge Weir can be seen to have acted in this context as something of a 'gatekeeper' himself, and the trial, as a whole, has served to open the 'black box' of LCN evidence, exposing differing judgements over the efficacy of the technique between experts (Jasanoff 1998).

This case has also served to emphasise how the interpretation of DNA profiles remains an area of contestation. Not only has it raised serious questions about the nature of the technical and interpretive work involved, but it has also exposed the fact that some aspects of these technologies are not fully accepted amongst the forensic science communities themselves, notwithstanding the opinions of the courts.

The Omagh trial represents a significant event in the history of the relationship between the law and forensic science, insofar as it was the first time that LCN DNA had met any form of concerted challenge from defence counsel. Moreover, the defence testimony which challenged the LCN evidence appeared to strongly influence the final conclusions of Judge Weir. It is possible that this case may come to represent the start of a period where DNA evidence may become potentially more susceptible to similar challenges. Following the collapse of the case, the Association of Chief Police Officers (ACPO) ordered a temporary suspension of the use of LCN technique in criminal investigation, with forces in Scotland and Northern Ireland following suit. The Crown Prosecution Service also announced a review of all cases involving LCN evidence. In Australia, the Omagh verdict led to calls for a review of the Falconio murder case (Murdoch 2007).

There is perhaps a certain amount of irony given the confidence with which the LCN evidence was put forward by the prosecution in this and other cases. The LCN technique had been involved in some high-profile successes in criminal investigations in the UK and in Australia, and LCN has played an especially prominent role in resolving many so-called 'cold cases', serious offences which had previously lain unsolved. In fact, LCN has formed part of a repertoire of sophisticated techniques developed by the FSS and made available for casework. These techniques encompass a range of technologies which facilitate the interpretation of DNA profiles and extend the informational potential of DNA. In addition to LCN, the FSS have developed Pendulum List Search (PLS), an automated method for the resolution of mixed DNA profiles from multiple contributors, and the 'familial search' technique which exploits similarities between the DNA profiles of family members in order to compare unknown profiles with possible relatives located on the NDNAD. These techniques have been used both singly and in combination, with successful results being publicised by the FSS and police forces, as well as being keenly received by the media (BBC News 2006a, 2006b, 2006c).

The existence of such technologies, and the benefits claimed for their use in casework, should indicate that the process of interpreting DNA profiles, and of matching them to suspects, is not always an unproblematic procedure. Nor does it result in categorical assignations of guilt or innocence. Instead, the deployment of these technologies demonstrate the quantity and quality of scientific labour required in order to generate meaningful results from biological material. This is a process in which the use of probabilistic approaches such as Bayesian reasoning plays an important role in the development of the algorithms which drive relevant systems. The results yielded by these systems produce conditional values, likelihood ratios, which can provide powerful guidance with regard to the consideration of hypotheses concerning the guilt or innocence of suspects. Certainly the FSS (as developer and provider), and the media, have been keen to hail the success of these interpretation technologies as representing the continuing ability of science to make dramatic interventions into the resolution of criminal cases, many of which having been seemingly impossible to solve. Novel DNA interpretation techniques, such as LCN and PLS, are represented as playing an important role in facilitating the progress of specific police operations, and on occasions, media imagery closely resembles the depictions of forensic science found in fictional dramas such as the American series *CSI*.

In this chapter I first focus upon some of the controversies which surround these technologies, with particular regard to questions concerning their scientific basis. With particular focus on the controversy surrounding LCN profiling, I discuss the ways in which certain areas of contestation have developed, and show how debates concerning these technologies continue to play out. By subsequently extending the discussion to encompass new technological developments, I argue that such attempts at constructing methods for resolving interpretation issues in an ostensibly more 'objective' manner do not succeed in negating these areas of subjectivity. On the contrary, I argue that such technologies serve to create new spaces of subjectivity and ambiguity within the Bayesian architecture that these systems employ.

I then continue by presenting a brief history of the employment of these techniques in criminal investigations. Here, my intention is to analyse the relationship between scientific methods for criminal investigation and other investigative issues. In pursuing this aim I seek to address the following questions: to what extent have these technologies altered the organisation and structure of criminal investigation; or, to what extent have they altered the framework of decision-making with regard to the pursuit of criminal cases? Conversely, to what extent are these technologies embedded in, and reliant upon, a wider set of government policies, police practices, and other exogenous factors? In considering these questions it may be possible to gain an improved understanding of the potentially variable relationship between technology and the organisation of criminal investigation.

In what then follows I introduce the practical complications that may be encountered during the analysis and interpretation of DNA profiles, and describe in more detail efforts to ameliorate these issues, via the application of techniques based on applications of probability theory. I focus upon these techniques in an attempt to demonstrate the high level of calculative labour involved in the generation of systems for the interpretation of DNA profiles, an issue which is often masked by popular representations of DNA profiling.

I seek to further this discussion to highlight how the process of DNA profile interpretation can be forced to deal with a number of sources of ambiguity which are susceptible to resolution via subjective judgement. Essentially I aim to determine to what extent the use of Bayesian theory masks certain assumptions, which, when brought to light, may lead one to question to what extent the whole process of these new technologies is entirely 'objective'. Thus, this case study seeks to build on the work of authors such as Latour, (1999) who have aimed to show how technologies may mask sources of subjectivity to convey an image of objectivity, a process known as 'black-boxing'. In particular, the research outlined herein attempts to reflect and build upon the research performed by Derksen (2000) who, in an earlier study of DNA profiling, sought to demonstrate how social practices determined what was construed as 'objective' and 'normal' in the course of DNA measurements (Derksen 2000) I also aim to examine the manner in which these assumptions of objectivity were upheld for so long, and to explore the objections to them. In examining the manner in which these technologies have been utilised by prosecuting counsel (namely the practices by which evidence from them is represented and conveyed in court), and the practices through which they have been challenged, I hope to improve the understanding of how scientific debates are re-focused in judicial settings.

7.2 DNA Profiling - Controversies

In Judge Weir's report, criticism of DNA analysis focused on two main issues. The first concerned shortcomings in the ways in which the DNA was recovered, packaged, stored and transported in the course of its analysis. Via a series of thorough examinations of police and forensic laboratory records, the Defence exposed several instances of what were seen to amount to inappropriate practices, and which, in the view of the Court, seriously compromised the integrity of the DNA evidence (Weir 2007, para.46). A number of concerns were raised over the handling of related exhibits at a range of stages in the investigative process, including the lack of appropriate labelling of bagged exhibits and the lack of anti-contamination procedures. One instance, which drew particularly heavy criticism from Judge Weir, involved the recovery of exhibits by a Scene of Crime Officer (SOCO) and a Detective Chief Inspector, who had both testified that they had worn suitable protective clothing at the time. These testimonies were subsequently revealed as untruths in the light of photographic evidence to the contrary (Weir 2007, para 50)[1]. Other concerns centred on the lack of adequate record-keeping of the movement and tracking of exhibits, and the wholly unsatisfactory nature of the conditions in which exhibits were stored (Weir 2007, paras.52-53). For example, no systems were found to exist for verifying precisely what items had been recovered or what had been placed in police storage, nor were there any methods for recording the removal of items from the store. The store at the Newry police station was described as 'a mess'.

Similar problems were also exposed at the laboratory analysis stage. An examination of procedures at Forensic Science Northern Ireland (FSNI), where the analysis of evidence was carried out, revealed serious problems concerning the labelling procedures of items. Identification labels attached to

[1] 'The effect of this, as I find deliberate and calculated deception in which others concerned in the investigation and preparation of this case for trial beyond these two witnesses may also have played a part, is to make it impossible for me to accept any of the evidence of either witness since I have no means of knowing whether they may have told lies about other aspects of the case that were not capable of being exposed as such.' (Weir 2007 para.50).

bags containing evidential items had sometimes subsequently become detached, which in a separate case had led to DNA evidence incorrectly being attributed to a suspect. This, along with other related shortcomings, had led to the temporary suspension of the laboratory's accreditation by the United Kingdom Accreditation Service (UKAS) (Weir 2007, para.56). Other evidence was presented which strongly implied that items may have been handled in a manner which had invited the possibility of contamination. The analysis of items had been carried out between 1998 and 1999, prior to the acceptance of formal procedures for LCN DNA, which only occurred in 2000. Hence witnesses who had worked for the FSNI at the time were unable to confirm that scientists wore gloves during analysis (as now stipulated for working with DNA), and that they had certainly not worn protective hats and masks, now also considered a mandatory requirement.

The second issue discussed by Weir in his report focused upon more fundamental questions concerning scientific opinion on the validity of the LCN technique as a whole. He concluded that the LCN technique could not be regarded as having been appropriately validated by the scientific community (Weir 2007 para 64). In his view, two articles published by the developers of LCN were insufficient to constitute validation of the technique. Weir accepted the Defence argument, which included references to the relative lack of uptake of LCN in other jurisdictions[2], and the lack of international agreement on validation procedures for LCN, in contrast to established guidelines and definitions for the validation of normal DNA tests (Weir 2007, para 62).[3] Attempts to assess the repeatability of the results via the consensus method were also criticised. In an experiment performed at the Birmingham laboratory of the FSS, three LCN tests had been performed on the same sample, with the result being that the consensus results obtained via the first

[2] Apart from the UK, LCN-type methods have only been adopted for evidential purposes in New Zealand and the Netherlands so far (Weir 2007, para. 62). Although LCN is also used in the USA, it is generally only used for intelligence purposes and has only been used evidentially on just one occasion.
[3] From the judgement: 'There has been no international agreement on validation and a conference held in the Azores in September 2005 had ended with agreement only that more work in that area was needed...This lack of agreement on LCN was in marked contrast to the normal SGM+ test for DNA for which there were internationally-agreed validation guidelines and definitions approved by the Scientific Working Group on DNA Analysis Methods (SWGDAM).' (Weir 2007, para.62).

two tests had been negated by the third. 'Thus the normal approach used in the United Kingdom had unintentionally been demonstrated by its own proponents to be potentially (and in that particular instance actually) misleading.' (Weir 2007, para.62.10).

In his report of the case, Judge Weir used the issue of the scientific validity of LCN as indicating the possible need to re-consider the manner in which scientific evidence was assessed for admissibility. He cited a report written by the House of Commons Science and Technology Committee, 'Forensic Science on Trial', which recommended the development of a 'gate-keeping test' of scientific evidence in line with the system used in US courts[4]. He also criticised the apparent lack of activity on the part of the Government to implement such a measure (Weir 2007, para.64).

7.3 The Caddy Report

Concerns about LCN had been voiced prior to the Omagh verdict (BBC News 2005, BBC 2007, Caddy et al 2008). In response to these a review was commissioned by the Forensic Science Regulator, with the findings being published in April 2008 (Caddy et al 2008). The report addressed a number of areas, relating to sample recovery and extraction, transfer and persistence of DNA, quantification and interpretation procedures, validation procedures, and the place of LCN DNA in the criminal justice system, including the Omagh trial. The report concluded that the scientific basis of LCN was scientifically 'robust' and that the technique was 'fit for purpose' for forensic use (Caddy et al 2008, p.1). Furthermore, the report was satisfied that all three companies supplying the generic technique had adequately validated their processes for analysing DNA at 28 and 34 cycles for samples of less than 200pg in volume

[4] In his report Weir quoted paragraph 55 of 'Forensic Science on Trial': '"55. The absence of an agreed protocol for the validation of scientific techniques prior to their being admitted in court is entirely unsatisfactory. Judges are not well placed to determine scientific validity without input from scientists. We recommend that one of the first tasks of the Forensic Science Advisory Council be to develop a "gate-keeping" test for expert evidence. This should be done in partnership with judges, scientists and other key players in the criminal justice system, and should build on the US Daubert test." (Weir 2007, para.64).

(Caddy et al 2008, p.1). However, it did recommend the development and eventual implementation of universal standards for the extraction, quantification and interpretation of profiles under LCN conditions.

The reaction within the forensic scientific community to the findings of the report has been varied. In some quarters the report was hailed as providing a decisive endorsement of the technique (Clayton 2008 Teeside lecture) However, it has also drawn equally vehement criticism elsewhere (Jamieson and Bader 2008, Gilder et al 2008). It has been argued that the report did not fully address many of the issues arising from the Omagh trial verdict, such as the lack of international agreement, and that it did not acknowledge dissention within the UK scientific community itself over LCN by failing to consult with known critics of the technique (Gilder et al. 2008; Jamieson and Bader 2008). It was also argued that the report paid insufficient attention to certain scientific issues concerning profile interpretation, such as the complications caused by the presence of DNA mixtures in LCN analysis[5]. Furthermore, even though the report expressed satisfaction with the manner in which LCN has been validated, it does not reproduce or detail any data relating to these validation procedures.

Regardless of the issue concerning the lack of published validation data, other critics have argued that, in the absence of common guidelines for the interpretation of LCN data, any attempts to argue for the existence of validation amounts to something of a tautology (Gilder et al 2008) The fact that no agreement exists over how data may be interpreted means that attempts at proving validation are futile (Gilder et al 2008). Furthermore, if this argument is accepted, 'the review raises important issues about what it means for a forensic science technique to be validated', and raises concerns over the way in which LCN DNA profiles had been interpreted in the past (Gilder et al 2008).

[5] 'DNA in forensic work frequently involves mixtures. The Report specifically recommends "more work" on the interpretation of mixtures (and indeed mentions mixtures only three times in 35 pages), despite the significance of mixtures in the forensic context and thus in criminal prosecutions.' (Jamieson and Bader 2008, p.2).

The authors of the report have also been accused of not consulting sufficiently with known critics of the technique, and that disproportionate and inappropriate attention was given to the companies who develop and market the technique. Moreover, it has also been alleged that one of these companies, the FSS, was aware of the conclusions of the report at least three weeks prior to publication. A report produced for a criminal appeal case by the FSS was quoted as stating: 'Preliminary indications are that [the Caddy report] makes no significant criticisms of the LCN technique' (Jamieson and Bader 2008, pp.2-3). This has been cited as a clear transgression of a Home Office edict to the Forensic Science Regulator, namely that the report be 'independent of any forensic science provider' (Home Office 2006, Jamieson and Bader 2008)[6].

LCN is clearly a contested technique. As I show below, it has, however, played a significant role in UK policing. The next section concerns the relationship of LCN, with the organisational context in which it is embedded. Here I seek to demonstrate that the success of technologies such as LCN are not entirely due to their scientific basis, but are dependent on the wider context in which these technologies are deployed

7.4 DNA Interpretation in Casework

The LCN technique has been used in the investigation of serious crimes, such as rape or murder, where other DNA profiling options have been exhausted, or in cases where the possibility of obtaining other forms of forensic evidence are limited (FSS 2005b). It is admitted that, due to the sensitivity of the LCN, any interpretation of the evidence must take context into account (FSS 2005b). However, the FSS also claim that LCN has played an important role in the resolution of so-called 'cold case reviews', namely cases which have remained unsolved for years, or even decades (FSS 2005b).

The FSS cite a number of these successes in their literature (FSS 2005b). LCN DNA was recovered from a microscope slide bearing evidence from a

[6] Home Office Consultation Document "Standard setting and quality regulation in forensic science" 31 August 2006,para 23.

case involving the rape and murder of Marion Crofts in 1981. Having supposedly been left deliberately untouched in anticipation of future technological advances, an LCN DNA profile from Tony Jasinskyj was obtained in 1999, leading to his subsequent conviction and sentencing to life imprisonment in 2002 (FSS 2005b). In another case, LCN was used to obtain a profile from a microscope slide retained from an unsolved rape case dating from 1995. When loaded onto the NDNAD, the DNA profile matched that of Mark Henson, who was convicted and sentenced to life imprisonment in 2005 (FSS 2005b). LCN has also been used in foreign jurisdictions. For example in 2003, Swedish authorities requested the FSS to carry out LCN DNA profiling on a knife used in the murder of the Swedish foreign minister Anna Lindh. The subsequent profile contained a mixture of the DNA of Lindh and that of suspect Mijailo Mijailovic, and played an important role in bringing the latter to justice.

PLS has also been hailed as making a decisive contribution to criminal investigation. In 2005 Duncan Turner was convicted of a serious sexual assault which occurred in a subway in Birmingham. PLS was cited as playing an instrumental role in resolving the DNA profiles within a mixture of material found on a pair of sunglasses found at the crime scene. A list of possible DNA profiles were generated that could have constituted the mixture observed on the evidence. These theoretical profiles were then compared to the NDNAD, with one of them eventually matching Turner (Sample 2006).

As suggested earlier in this chapter these technologies have also played a particularly prominent role in a number of large-scale 'cold case' operations conducted by police forces in the UK. Operation Phoenix, launched by Northumbria Police in 2002, has involved the re-investigation of all unsolved sexual offences in the area carried out between 1985 and 1999 (Crown Prosecution Service 2003). Operation Phoenix sought to improve the conviction rate for serious sexual offences, and to also make a substantial contribution to the NDNAD (Dixon 2004). In the process of casework, LCN samples were taken from a variety of objects, particularly samples from material retained during previous investigations, as well as archived material such as microscope slides and tapings (FSS 2005a). Overall, the operation yielded forty-one previously unidentified DNA matches, with twenty-three of these being loaded onto the NDNAD. Fifteen serial offenders have been identified. Operation Phoenix led to ten convictions, and three others charged pending trial (Dixon 2004).

A nationwide initiative, Operation Advance, also sought to capitalise on the improvements in DNA profiling technology. In 2004, the Police Standards Unit (PSU) commissioned the FSS to carry out a study which identified 214 cases with crime stain profiles generated via the older Single Locus Probe method (FSS 2006). Around ninety per cent of these involved serious sexual assaults. Operation Advance aimed to re-examine these SLP profiles using the more advanced techniques available to the FSS, attempting to match any profiles with those found on the NDNAD, and to present the relevant police forces with the scientific results for possible further consideration. 154 of these cases were progressed, and 73 profiles were developed with 33 providing matches to named individuals already included on the NDNAD (Home Office 2005a).

Operation Advance was unique in a number of ways. First, it was unique in that scientific analysis preceded police investigation of the crimes. It was also cited as the first time that re-analysis experiments were instigated under the orders of central government. Furthermore, the operation itself acted as an important means through which to test the value of pursuing cases in such a way, with each the progress of each case being monitored.

As of October 2004, 108 profiles had been obtained from samples that had been upgraded. However, the Home Office Report on Operation Advance states that these ranged from full profiles to uninterpretable mixtures. Of 77 profiles searched on the NDNAD, 42 matches of some kind were recorded, (with 34 of these matching immediately to individuals present on the NDNAD), hence returning a success rate of 28% (FSS 2006).

On 12th July 2005, Operation Advance II was launched, focusing on cold cases dating from between 1994 to 1996. Out of 1012 cases of rape and sexual assault occurring in the period 1994-96, 157 of these were recognised as suitable for further forensic work (FSS 2006). 66 cases were subsequently processed by the FSS resulting in 22 matches with the NDNAD. Four arrests are reported to have been planned on the basis of successful database matches (FSS 2006).

The Report on Operation Advance also provided recommendations to inform decisions on whether or not to commence re-investigation on particular 'cold cases'. It highlights three issues. The first concerns the investigator's perspective. This is viewed as being variable and dependent upon organisational 'memory'. Whilst the latter term is not defined, it could be construed to refer to the possible continuing presence of investigators involved with original pursuit of the case, or the presence of archived documentation relevant to the original investigation. (The MacPherson report was critical of the poor recording practices employed by the Metropolitan Police, which were exposed during the inquiry into the handling of the Stephen Lawrence murder investigation (MacPherson 1999)). A second factor concerns the outlook of the victim (or their family), and the magnitude of their desire to persevere with the case. The Report on Operation Advance cites this as another variable factor. It can be seen how re-examination of cold cases involving sexual offences may cause distress for victims. Although there exists the possibility of psychological 'closure' through the resolution of a case, the re-examination of such cases may involve the re-visitation of trauma perhaps long buried. As in other sex crime cases, there also exists the potentially highly stressful experience of having to face the assailant in a courtroom setting, and the experience of having to face highly antagonistic and intrusive cross-examination.

A third issue, the scientific position, is cited as being a great deal less variable. In their view, 'scientific value is added as the body of knowledge increases with time'. Whilst this may be the case, the first two issues perhaps represent a series of factors which may serve to downplay the effectiveness and usefulness of new scientific developments. If taken all into consideration, in a concerted fashion, it can be seen that the decision to re-visit cold cases does not depend on scientific developments alone; nor does scientific progress guarantee convictions (Home Office 2005).

Scientific advances in DNA profiling have, nonetheless, been cited as catalysing the interest in the possibilities of resolving cases previously considered unsolvable, and are often acclaimed as playing an instrumental role in improving detection rates as a whole. These developments have attracted a considerable of public interest. Many of the technologies which have been hailed in this regard involve new methods for resolving complicated DNA profiles, in which interpretation by the human eye is compounded by factors such as the low quantity of material, or by the presence of mixed DNA profiles. In the next section I discuss relevant interpretation issues, and describe the Bayesian technologies which have been developed to address them. These may involve complex mathematical forms; yet in what follows I question whether the use of Bayes in such a way has led to approaches which can be considered any more 'objective'.

7.5 Interpretation Issues

A number of issues exist with regard to the interpretation of DNA profiles. One example of the ambiguity that surrounds the interpretation of DNA profiles, and which is of particular pertinence in the case of LCN profiles, concerns the manner in which profile peaks are identified from electropherograms. In cases where DNA has been recovered in relatively ample quantities, the identification of allele peaks is unproblematic; such peaks will be large and appear very prominently on electropherograms. In cases where the sample volume is small, and the size of peaks becomes less marked, peak identification may become more ambiguous. Here, it may be more difficult to differentiate DNA peaks on the basis of size, relative to background noise, amongst which there may exist artefactual peaks. A key factor in the identification and recognition of *bona fide* DNA profile peaks is the *peak height threshold*. This is the height at which a peak needs to reach before it may be considered as originating from an allele, as opposed to it being merely part of the background signal. Decisions concerning the precise level at which the threshold is set have a direct influence in shaping judgements concerning the recognition of peaks as originating from alleles. For example, identifications of peaks may determine whether a DNA profile emanates from a single person, or from a mixture of DNA samples:

> 'As you drop the threshold...three new alleles have popped up...here's a problem, because the accused doesn't have them, and this looks like a single source profile, if you knew nothing about the case, you gotta say 'that's a single source'...drop it further, and it becomes even more complicated because you think you've got three alleles here...so here I've got three alleles, so that's a mixture, as soon as you mix DNA together you can't pull them apart!...So I've got a real problem in this case. And my question to the Crown was 'Why did you draw the line at 25? Because it's the best position for the Crown...'
>
> 'So that's what I say, I'm not coming up to 'this is the profile you should have reported' I'm saying 'why did you report that one and not tell us...?' (Interviewee 3, 2008)

Figures 7.1a to 7.1c display this phenomenon more clearly. Each figure 7.1a-c displays the same DNA profile, but in each case the peak height threshold is set at a different level. In Figure 7.1a the threshold is set relatively high. In this case, only peaks A and B would be recognised as emanating from alleles. However, in Figure 7.1b, the threshold is slightly lower, but sufficient enough for peaks C and D to be recognised as originating from alleles as well. In

Figure 7.1. Effect of Peak Height Threshold Variance on Peak Recognition

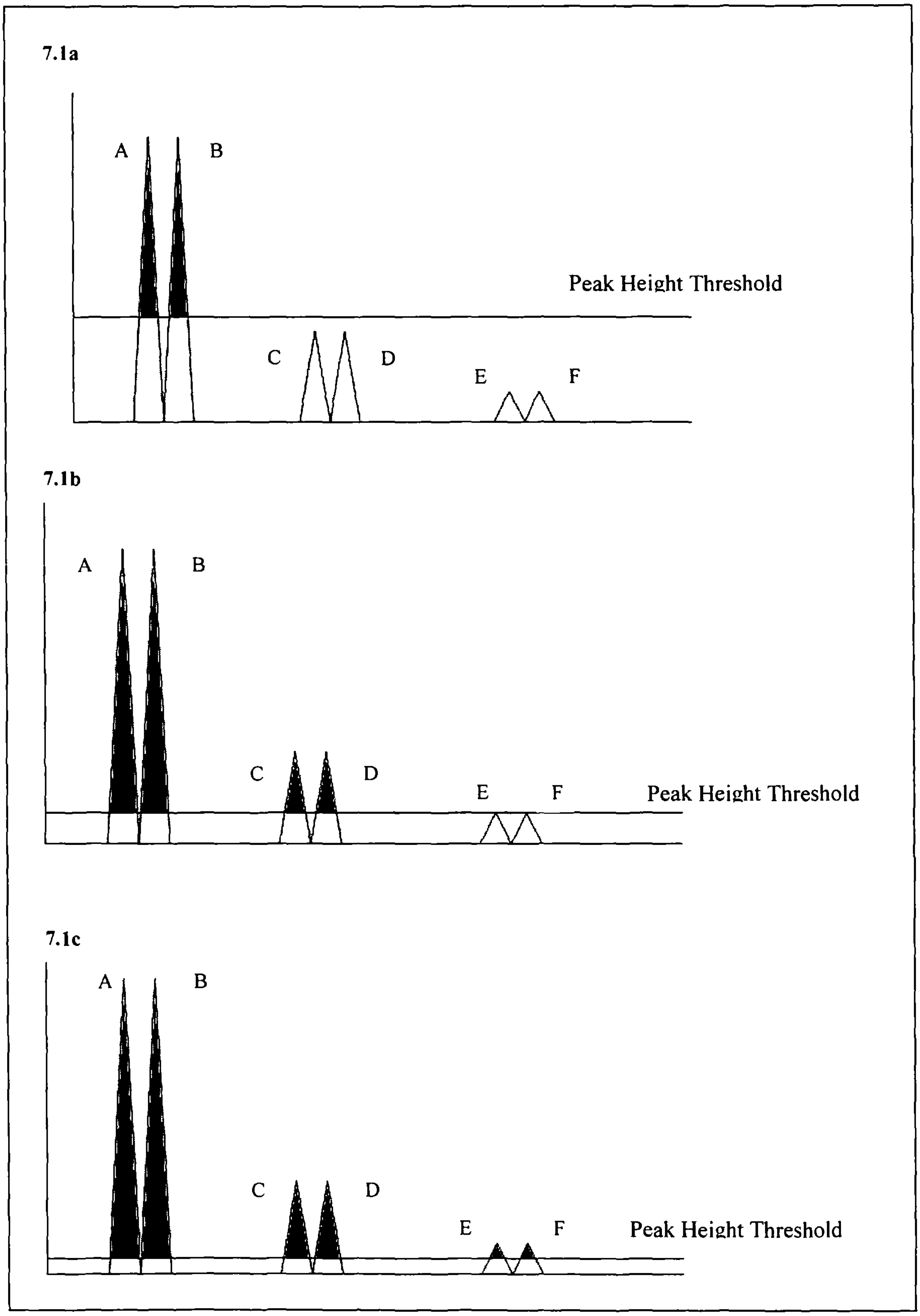


Figure 7.1c, the threshold is set slightly lower still. In this case, peaks E and F may now be considered as *bona fide* DNA peaks.

The peak height threshold is thus an instrumental device for enabling scientists to differentiate between the existence of peaks arising from DNA, and those which correspond to the presence of unimportant background artefacts[7]. However, rather than being a means of settling controversies over what peaks correspond to, the deployment of the threshold actually belies the ambiguous nature of the process of DNA interpretation. No clear guidelines exist concerning the precise point at which the threshold should be set:

> 'I'm not coming up to 'this is the profile you should have reported' I'm saying 'why did you report that one and not tell, not anywhere in scientific, did the scientist say 'if I do this, then that person's excluded and that's a problem...There's the profile...here its just showing how from that profile the Crown say they know that two people have made that up...here's the problem...and when they go that way...there are three possibilities there, there's no way scientifically you can know it's the shaded in one which is the Crown position.' (Interviewee 3, 2008).

Hence the recognition of peaks in CE traces can be a subjective phenomenon. Rather than overcoming subjective judgement, it is in effect controlled by it, and merely serves to reinforce a particular perception of the data provided by the analysis. The decision over where to set the threshold may belie a whole host of experiences, something which the absence of clearly defined scientific procedures only serves to emphasise. The shape of the peaks themselves may defy the possibility of objective recognition even further:

> 'Only question comes [is] if a peak has a certain shape, or a certain signal intensity, is it below a certain pre-determined threshold...that threshold is almost arbitrary to some degree, some of these are arbitrary...'
>
> 'We're striving to get as much information as we can, and the arbitrariness of the minimum peak height threshold encourages us to throw away some information in many instances.'
>
> (e-Symposium Discussion, 2007).

[7] Caused by, for example, minor contaminants.

The use of electrophoresis technology was intended to overcome the sources of ambiguity associated with the use of gel technology, which had previously been used in forensic DNA profiling (see chapter 6)[8]. However, what these examples show is that sources of ambiguity continue to exist, and hence judgemental and experiential reasoning play a continuing part in shaping the representation of DNA data.

7.6 Mixture Analysis

The analysis of DNA profiles arising from the interaction of more than one individual is an integral part of forensic casework. In cases involving serious crime, such as rape or murder, it is almost inevitable that analysis will centre on crime scene samples in which a mixed DNA profile is found. Naturally, mixture analysis raises specific and complex statistical issues. Technological developments have improved the means by which mixed samples can be resolved and distinguished, and these have increased the sophistication of statistical interpretation. New theoretical approaches have sought to maximise the utilisation of available information, yet they have also had to allow for certain limitations that new technologies present, for example the phenomenon of 'stuttering' which is common to PCR analysis. This in turn has served to increase the complexity of such approaches further, yet there still exists scope for reconciling statistical theory and practical analysis.

Many recent developments in mixture interpretation have taken the form of Bayesian approaches. These have sought to take into account both the enablements and limitations of the instrumentation to hand, all in what is claimed to be a 'consistent logical framework' (Clayton and Buckleton 2004, p.223). These appear to be replacing older, more rudimentary approaches. In what follows the development of DNA mixture interpretation methods is traced, taking into account the historical-technological context in which they came about. Following on from the discussion by Clayton and Buckleton (2004), the frequentist method, involving exclusion probabilities, is covered

[8] See Derksen (2000) for a review of the issues associated with the interpretation of DNA profiles from gels.

first, followed by consideration of the development of the major Bayesian approaches (Clayton and Buckleton 2004).

The Frequentist Method – Random Man Not Excluded (RMNE)

Early attempts at the interpretation of mixed DNA profiles involved the use of exclusion probabilities, defined as 'the probability that a random person would be excluded as a contributor to the observed DNA mixture' (Clayton and Buckleton 2004, p.219). Such a definition gave rise to the alternative title of 'Random Man Not Excluded' (RMNE). In mathematical terms, the exclusion probability at a given locus l, (PE_l), is given as:

$$PE_l = 1 - (\sum p(A_i))^2$$

Where $p(A_i)$ represents the probability of the presence of a certain allele at l, given that the alleles observed at l in the mixture ranges from A_1 to A_i. The sum of all values of $p(A_i)$, is squared to acknowledge the presence of two alleles per locus, and subtracted from 1, leads to the probability of exclusion. In this case, the exclusion probability applies to a single locus. Using the product rule, the probability exclusion across a genotype (PE) is therefore:

$$PE = 1 - \prod(1 - PE_l)$$

It can be seen that the RMNE method relies on the accurate calculation of allele frequencies via the use of pre-existing databases. It does not, in itself, represent a direct engagement with the data, in terms of allowing for measurement error etc. RMNE can thus be seen to be a product of the era in which the primitive gel technology available at the time was relied upon to convey and represent profile information. In this context issues concerning measurement error were dealt with through the use of binning systems, and thus it was assumed that the data had been suitably conditioned before PE was calculated. Thus it is perhaps understandable that RMNE can be criticised in the modern context as a wasteful method in terms of the way it utilises information (Clayton and Buckleton 2004, p.223).

Nevertheless, RMNE is still widely used, although it is claimed there is a general trend underway towards replacing the use of exclusion probabilities with likelihood ratios. However, it is claimed to possess some advantages over Bayesian approaches, most notably the lack of assumptions concerning the number of contributors to an unknown mixture, and indeed the ability to assess the evidential value of a profile in the absence of a corresponding suspect profile to hand. It is not reliant on modern instrumentation, which may be advantageous in some cases where resources are limited, and it is also claimed that exclusion probabilities are easier to explain to juries than likelihood ratios (Clayton and Buckleton 2004, p.222-223).

However, Brenner (1997) has argued that a distinct logical shortcoming is apparent within the exclusion probability approach (Brenner 1997). He characterised evidence as possibly consisting of thus:

1) blood types of the suspect
2) blood types of the mixed stain

According to Brenner this enables one to infer that: 3) the suspect is not excluded. In pointing out that, whilst 3) can be deduced from 1) and 2), 1) cannot be deduced from 2) and 3), nor from 1) and 3). This, according to Brenner, amounts to a loss of information. He claims that a likelihood ratio is a summary of the information in 1) and 2), which enables some kind of inference to be made regarding 3). However, the exclusion probability merely summarises 2) and 3), but allows no inference to be made regarding 1) (Brenner 1997).

Qualitative Bayesian Approaches

A relatively early form of Bayesian interpretation was presented by Evett et al (1991), which coincided with the advent of the further development of SLP technology. This method quickly gained approval, being cited by NRC 2, and received favour from the presiding judge in the O.J. Simpson murder trial (National Research Council 1996; Clayton and Buckleton 2004, pp. 223-224).

This method was formulated before the development of advanced capillary electrophoresis systems, which as described below enable additional variables such as peak height and area to be taken into account. However, at the time it was presented as the most efficient means through which to optimise the data available to caseworkers. Furthermore, the likelihood ratio approach was amenable to adaptation depending on the particular scenario[9], and could also be used to account for sources of uncertainty including sampling error and subpopulation correction, lending it a flexibility not open to RMNE.

The basic likelihood ratio (LR) for mixture analysis is thus:

$$LR = \frac{Pr(E \mid G_s, G_v, H_p)}{Pr(E \mid G_s, G_v, H_d)}$$

The numerator represents the probability of the evidence given the genotype of the suspect (G_s), victim (G_v) and the prosecution hypothesis (H_p). The denominator represents the probability of the evidence given the same data, but given the defence hypothesis (H_d), instead.

Given the technology available at the time, this method provided a simple and elegant formula through which evidence could be assessed. It was, like RMNE, dependent on accurate measurement methods, and in itself could not account for any possible shortcomings on the account of technology[10]. The power of statistical methods was constrained by the representative limitations of the technology, which in turn could convey information of only a limited quality to supply the formulae in question.

---

[9] Different hypotheses could be created depending on precisely who the contributors were thought to be. For example, the prosecution could advance a hypothesis whereby the mixture contained the DNA of a suspect and the complainant, countered by a defence hypothesis allowing for the presence of the complainant and an unknown unrelated individual. Another scenario might allow for the same prosecution hypothesis, but with the defence arguing for the presence of the suspect in conjunction with an unknown individual. Furthermore, the prosecution might consider a mixture to consist of two suspects, whereas the defence might counter with the hypothesis that the mixture derived from two unknown and unconnected persons. In each case, a likelihood ratio formula can be derived to assess each case (Clayton and Buckleton 2004, pp. 226-228).

[10] Furthermore, like any Bayesian approach, it's usefulness was reliant on hypotheses from both prosecution and defence that appropriately reflected background information.

The deployment and use of new technologies such as capillary electrophoresis, in tandem with PCR and STR technology, has enabled the development of more sophisticated statistical approaches, able to incorporate a wider range of variables. More importantly still, the way in which data is presented via this technology has allowed analysts to draw some notable conclusions about the nature of DNA mixtures, which in turn has fed back into statistical considerations. It must be pointed out however, that the means by which analysis proceeds are not infallible. Certain potential defects regarding, for example, PCR, are recognised. Analysts are continuing in their attempts to incorporate possibilities raised by such issues into their formulae. However, as the following discussion of the so-called 'binary method' demonstrates, there still exists considerable scope, and need, for subjective judgement in the process of interpreting DNA mixture interpretation.

The interpretation of mixed samples is carried out according to a number of distinct steps (Clayton 1998). A DNA profile must be identified as a mixture in the first place, a process which may not be entirely straightforward even using modern instrumentation. A mixture can only be identified if the alleles from the minor contributor are able to be identified above background noise. According to Clayton and Buckleton, a mixture ratio of 1:10 is taken as the threshold for minor components (Clayton and Buckleton 2004, p.233). However, this figure appears to have arisen 'in practice' and it is unclear as to what grounds it is based upon. Regardless of this, the presence of multiple peaks, particularly those presented in unbalanced size, is normally taken to indicate the presence of a possible mixture[11]. However, certain effects observed in the electropherogram may serve to cause obfuscation. Some peaks may arise as a result of so-called 'stuttering' whereby repeat units may be lost from alleles, normally as a result of miscopying during the PCR amplification cycle. Furthermore, flaws in data interpretation software may serve to misidentify artefacts within the trace as peaks. Conversely, in cases

[11] Notwithstanding the obvious issue of certain background information, most notably the type of reported offence being investigated.

where contributors share common alleles, it may not be clear from the electropherogram if two contributions are present. This may also be the case where contributors are homozygotic at a particular locus. Another source of confusion may arise from the possibility of genetic phenomena such as trisomy (whereby individuals carry more than a pair of chromosomes), or the duplication of specific genetic regions to other sites in the genome. Mixture interpretation using modern means is therefore a potentially complicated process, and may require a significant period of experience before users develop the requisite skills.

Once a mixture is declared and allelic peaks have been designated, the potential number of contributors is ascertained. Although some studies have sought to develop methods that enable contributor number to be derived via statistical means (Egeland et al 2003), it is more likely that at present this is a process which occurs through previous data interpretation experience and also in relation to the circumstances of each particular case. Certainly, it has been reported in the literature that two-person mixtures constitute the 'overwhelming majority of mixtures encountered during casework' (Torres et al 2003). Following this, the mixture proportion (M_x) or mixture ratio (M_r) is determined. Calculation of this figure is a complex process, and a computer package, known as Pendulum List Search (PLS), is used by the FSS to estimate M_x (Clayton and Buckleton 2004). The underlying logic of the estimation of M_x relies on the assumption that it is invariant across loci; however according to empirical studies there appears to be a certain lack of agreement as to whether this is precisely the case (Gill 1998; Clayton and Buckleton 2004).

The estimation of M_x is a key step in the process as this figure is used in the identification of the possible genotypes of the contributors, along with the degree of heterozygote imbalance (H_b). This latter parameter is first used to judge the compatibility of certain alleles if they occurred in pairs, and involves the comparison of peak areas. Clayton and Buckleton report the use of ≥ 0.6 to indicate possible paired alleles (Clayton and Buckleton 2004, p.240). Possible genotype combinations are then assessed using M_x. A similar value for M_x at each locus may indicate a correct profile. Putative genotypes may then be compared with any existing reference samples from suspect and/or victim, if available.

Even at this crucial latter stage, decisions over what may be accepted as genotypes in a mixture is largely dependent on human judgement. H_b and M_x essentially serve as guidelines for the decision by analysts; there appears to be no theoretical basis for calculating the threshold value for H_b, instead it seems to be derived on the basis of past experience. However, the use of M_x has been criticised for revealing an inherent circularity in the process. This value, being calculated from putative genotypes, is then used to confirm the final genotypes. In practice it may also be the case that between them, loci will yield results which combined may serve to adduce ambiguity; and thus the final say will essentially rest on human judgement. This decision may be complicated by the possibility of stutter peaks, or the potential for alleles exhibiting low peak area to be undetected by the software. The binary model used in PLS rests on a number of assumptions, namely: the invariance of mixture proportions across loci, the proportionality of peak area to the amount of DNA; and the area of shared peaks being equal to the sum of the contribution of the contributing individuals (Clayton and Buckleton 2004, p.242). In practice it is debatable as to what extent these assumptions can be relied upon. Furthermore, there are risks associated with the use of reference samples to assign genotypes. Clayton and Buckleton have pointed to the possibility of the presence of these samples to cause bias in the way that analysts perceive mixed profiles (Clayton and Buckleton 2004, p.242). Finally, it is recognised that there remains further work to incorporate other sources of uncertainty into the theoretical underpinnings of the binary approach (Clayton and Buckleton 2004, p.238).

7.7 Problems in LCN Analysis

The analysis of trace elements of DNA was initially envisaged by Jeffreys et al (1988) (Jeffreys et al. 1988). Following on from this work, Findlay et. al. (1997) reported the successful DNA fingerprinting of material derived from a single cell, and moreover, during the 1990s, DNA was successfully profiled using a wide variety of sources, including fingerprints, fingernail debris, discarded cigarettes, single hairs and items of clothing (Findlay et al 1997; Buckleton and Gill 2004)

The decision to opt for LCN analysis may depend upon a number factors, such as the nature of the surface material from which DNA may be recovered and the circumstances of the case (FSS 2005b). The decision to proceed with LCN analysis will occur if it is felt by the investigation team that a DNA profile would benefit the pursuit of the case, but it has been impossible to obtain such a profile via conventional profiling procedures.[12]

Certain sources of controversy can be located in practices surrounding the assignation of alleles to individuals on the basis of LCN profiles. The process of assigning specific alleles in an LCN profile to particular individuals is subject to a potentially considerable degree of complication, largely due to issues such as contamination, 'drop-out' and other stochastic effects as described above.

Allele drop-in: It may not be entirely accurate to discuss allele drop-in in terms of 'contamination'. 'Background' has been discussed as a possibly more appropriate term (Buckleton and Gill 2004, p. 280-281). Nevertheless, it has been argued that spurious allele peaks may arise from two relatively distinct sources. Instances involving single allele drop-in have been distinguished from occurrences where multiple spurious peaks are sighted (Buckleton and

[12] Approximately forty per cent of all samples submitted for DNA analysis are deemed as 'sub-optimal', namely they are incapable of yielding profiles that may be generated and interpreted in normal conditions (Greenhalgh 2008, Forensic Science Society Conference, Teeside University 17 April 2008). It has been opined that the generation and interpretation of DNA profiles via conventional means requires over 250pg of recovered DNA (Greenhalgh 2008, ibid), although it may be possible to obtain some form of result if DNA is recovered in the range of 100-250pg. LCN DNA profiling has been estimated to involve quantities of DNA below 100pg.
The issue of whether to quantify the amount of recovered from casework samples is a matter of some contention. Of the three forensic science providers in the UK, only Orchid Cellmark and LGC routinely perform quantification (Caddy et al 2008). The FSS do not perform quantification on the grounds that potentially valuable amounts of DNA may be unnecessarily discarded (Caddy et al 2008).

Gill 2004, p.282). It is thought the former may occur via a number of potential stochastic instances constituting 'background' sources, whilst in the latter case, contamination may be seen to arise from a more easily identifiable source.

Heterozygote Imbalance: The predominant method for LCN analysis simply involves an increase in the number of PCR amplification cycles, from 28 (as used in conventional profiling), to 34. This however, has been shown to induce greater imbalance between heterozygotic allele peaks. Furthermore, comparisons across individual loci revealed a high degree of variance from the multilocus H_b value (Buckleton and Gill 2004, p.279). The degree of variance for H_b in smaller allele peaks was found to become so great as to render the latter totally uninformative (Buckleton and Gill 2004, p.279).

Allele drop-out: According to Gill et al (2007) the problem of heterozygote imbalance is exacerbated in LCN analyses (Gill et al 2007, p.128). An extreme form of heterozygote imbalance may result in the disappearance of one of the alleles at a locus, a phenomenon which has been referred to in the literature as 'drop-out' (Gill et al 2007). When a PCR amplification is prepared, only a portion of the extract is analysed; in the case of the particularly small volumes associated with LCN, it is claimed that there is a greatly increased likelihood that some alleles will not be included in the reaction. The phenomenon of dropout may also occur due to sample degradation.

The interpretation of LCN profiles is recognised by the relevant experts as a significant issue, and attempts have been made to overcome the related problems. Many of the methodological problems associated with conventional DNA profiling are exacerbated in the case of LCN. If sufficient quantities of an individual's DNA are present at a particular scene, the resulting profile should be prominent enough to be easily visible over any contaminating material. However, the non-sterility of the environment may also lead to the deterioration of DNA samples, and thus analysts may only be able to recover a partial profile. Another source of complication involves the recovery from the crime scene of mixed DNA profiles from two or more contributors. Whilst these issues add a considerable degree of complication to the interpretation of conventional DNA profiles, the small starting quantities involved in LCN analysis mean that the problems in interpretation caused by these phenomena are rendered even more intractable, exacerbated by the possibility of 'drop-in' and drop-out'.

A paper by Curran and his colleagues attempts to show how such issues may be taken into account. First, there is the order in which 'drop-in' alleles may come to contaminate an LCN profile. No assumptions are made regarding the order in which alleles are added to a profile (Curran, Gill et al. 2005). Hence, the assumption of random order of contamination needs to be taken into account in mathematical terms. If the set of potential contaminating alleles contains k members, then there are $k!$ possible arrangements of the order in which these alleles came to contaminate the profile. Second, population substructure may have to be taken into account. It is here that a potentially serious controversy arises. Although the authors state an equation to take this effect into account, it uses the coancestry coefficient θ. This raises something of a contradiction if it is argued that allele probabilities for contaminating alleles are more likely to be related to sources within the laboratory (i.e. analysts) rather than the population of the offender. The use of θ may actually have serious implications for the calculation of likelihood ratios based on LCN data, as the probability of drop-in for certain alleles may vary considerably from a calculated figure based on the genotypes of analysts. It is unclear from the literature whether this practice is carried out in UK forensic laboratories on a routine basis: certainly the literature does little to emphasise this precaution.

Originally, analysts had used a so-called 'consensus' system, whereby an allele was reported only if duplicated in a replicate analysis of a sample extract (Gill et al. 2000). Furthermore, if negative controls displayed duplicated alleles corresponding to those found in the samples, they were not reported and samples were reanalysed if possible (Gill et al 2000, pp.38-39). However, the consensus approach was viewed as inefficient in that it did not make full use of the information yielded by the profiles (Gill et al 2006) In order to make full use of the data, including the option of making calculations for mixed profiles, and to apparently take into account population substructure, an expert system, known as *LoComatioN*, has been developed to relieve the computational burden.

LoComatioN is a hypothesis-driven expert system which has been designed to generate likelihood ratios (LRs) for any number of different propositions to be evaluated (Gill, Kirkham et al. 2007). The derivation of each LR requires an evaluation of the probability of observing evidence under a prosecution and defence hypothesis, H_p and H_d respectively, and involves the formula:

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)}$$

The use of likelihood ratios to evaluate evidence based on the consideration of prosecution and defence hypotheses shows that LoComatioN utilises a Bayesian mode of probability. The system allows the analyst to formulate and test propositions of their choosing, although in most cases it is likely there is little variation in the type of propositions being formulated. For example, in a rape case, the prosecution proposition H_p may be that a crime scene stain may consist of material from victim *V* and a suspect *S*. A corresponding defence hypothesis, H_d, may assume the same stain contains contributions from *V* but from someone unrelated to the suspect. In this case the stain material is denoted as *V* plus unknown (*U*) material. However, it is claimed that LoComatioN is able to evaluate more complex propositions, such as cases involving contributions from up to five individuals, and over five replicated analyses (Gill et al 2007).

A number of assumptions are built into LoComatioN in relation to the issue of drop-out and contamination. With regard to the former, two issues are encompassed in the probabilistic calculations; firstly, the need to estimate the probability of drop-out (Pr(*D*), and second, the need to include the probability of a specific allele being subject to drop-out (Gill et al 2007). In previous LCN interpretation approaches, an approach was used in which a possible drop-out event was signified by the designation F. Using this approach, the probability of a given allele a being subject to drop-out can be given as $\Pr(a\mathrm{F}) = 2p_a$ (multiplication by 2 allows for the possibility of the dropped-out allele occurring at either of the chromosomal locations of the locus). However, this approach has been viewed as potentially favouring the prosecution hypothesis, particularly as $\Pr(D)$ decreases; the assumption of a low value of $\Pr(D)$ possibly creating an erroneously high estimate of a match. A different concept has been introduced in LoComatioN. In this case, if drop-out is required to support a proposition it is assumed that the identity of the unknown allele Q may be any allele except those observed in the DNA profile:

$$\Pr(Q) = 1 - \sum_{i=1}^{n} p_i$$

where n alleles are observed in the profile and p_i is the frequency of the allele (Gill et al 2007, p.129-130).

LoComatioN can use this formulation to calculate whether drop-out may have occurred. For example, an LCN sample taken from a crime scene may yield evidence giving a profile in which three alleles, *abc,* are present at a particular locus. A suspect's profile for the same locus is *ab*. Hence a prosecution hypothesis will have to account for the presence of allele c. The assumption made here is that another contributor's DNA is present, containing c, plus another allele that needs to be accounted for. The Q designation allows for this missing allele. In practice, the use of Q manifests itself simply as $p(Q) = 1-p_a-p_b-p_c$. The first step to formulating a proposition involves calculating the probability of the evidence given H_p and the instance of drop-out (D), namely $\Pr(\mathrm{E}|\, H_p, D)$. This is given by Gill et al (2007), in line with Bayes Theorem, as:

$$\Pr(\mathrm{E}|\, H_p, D) = \Pr(\mathrm{E}|\, H_p)\Pr(D), \text{ where } \Pr(\mathrm{E}|\, H_p) = 2p_c p_Q$$

Any prosecution hypothesis will however also have to account for the possibility that no drop-out has occurred. Here, the assumption is that the crime scene stain is a mixture of the suspect and another contributor. Hence, if the suspect is ab, and is assumed to be the offender under H_p, then the prosecution has to account for what the genotype of the unknown contributor might be. In order to explain such a mixture, the unknown contributor would be assumed to be either one of three genotypes: ac, bc, or cc. If no drop-out has occurred, then the probability of there being an unknown contributor is $p_{ac} + p_{bc} + p^2_c$. Hence if no drop-out has occurred:

$\Pr(E| H_p, \check{D}) = \Pr(E| H_p) \Pr(\check{D})$ where $\Pr(\check{D})$ = probability of no drop-out, and $\Pr(E| H_p) = p_{ac} + p_{bc} + p^2_c$

The prosecution hypothesis allowing for *both* the possibility of drop-out and no drop-out is therefore:

$$\Pr(E|H_p, D\& \check{D}) = 2p_c p_Q \Pr(D) + p_{ac} + p_{bc} + p^2_c \Pr(\check{D})$$

A corresponding defence hypothesis H_d, would assume that the crime scene stain is the result of a mixture between two unknown contributors, and hence the suspect is not involved. Allowing for the possibility of drop-out of unknown allele Q, H_d has to take into account all the possible permutations of alleles in a possible mixture $abcQ$, of which there are 24. If no drop-out is assumed, then the only permutations involve abc, of which there are 12.

Hence for $\Pr(E| H_d, D) = \Pr(E| H_d) \Pr(D)$ where $\Pr(E| H_d)=24p_a p_b p_c p_Q$

For $\Pr(E| H_d, \check{D}) = \Pr(E| H_d) \Pr(\check{D})$ where $\Pr(E| H_d)=12p_a p_b p_c(p_a+p_b+p_c)$

Hence H_d, allowing for both the possibility of both drop-out and no drop-out, is:

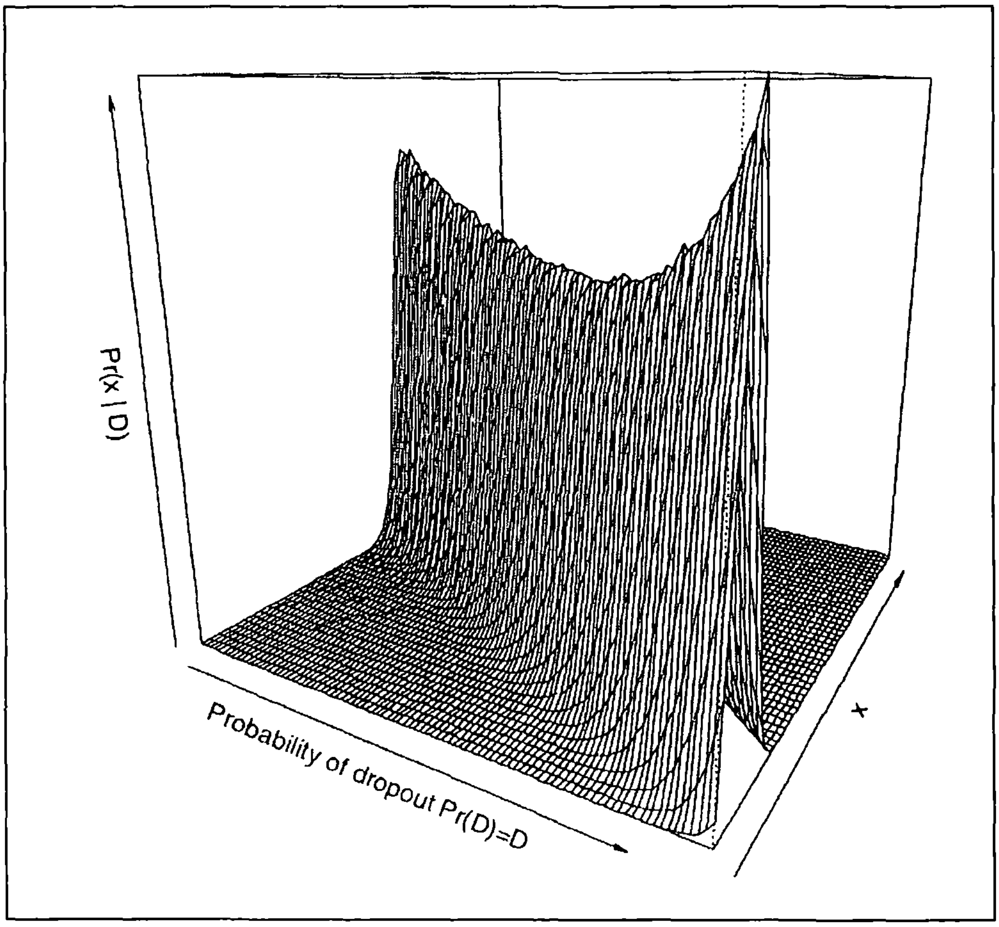
$$\Pr(E|H_d, D\& \check{D}) = 24p_a p_b p_c p_Q + 12p_a p_b p_c(p_a+p_b+p_c)$$

The full LR equation, for the assessment of H_p and H_d can be simplified to:

$$\text{LR} = \frac{\Pr(\check{D})\,(2p_a+2p_b+p_c) + \Pr(D)(2p_Q)}{2p_a p_b[\Pr(\check{D})(p_a+p_b+p_c) + \Pr(D)(2p_Q)]}$$

Pr(D) is formulated via simulation. In this case, the simulations estimate the probability of observing x alleles (at a certain number of loci), given that the probability of drop-out is equal to D (Pr(D)=D). Hence the aim is to estimate Pr (x | D). This is problematic however, given that the true probability of drop-out, in any situation, cannot be known. Instead, the likelihood of D is estimated from the observed data x. This observed data is generated, via the repeated production of random profiles with n_c contributors. Each iteration of the simulation will produce a random profile that could have resulted from the contribution of n_c unrelated individuals profiles. From each profile the number of observed alleles, x, is counted. Pr(x | D) is therefore estimated by determining the frequency of the occurrence of different values of x at a given value of D. Repeated iterations of random profiles are generated at a range of values of D from 0.0 to 1.0, and the frequency of values of x is plotted in a manner shown in Figure 7.2 below (Gill, Kirkham et al. 2007).

Figure 7.2. Reproduced from Gill et al 2007, p.135)



Using such a graph it is possible to investigate particular cases, and enables the analysts to answer questions of the type 'what is the most likely value for Pr(D) if x=32?' This can be answered by identifying when the maximum likelihood occurs for x=32, which is, (as for any value of x) Pr(D)=0. However, as Gill et al (2007) demonstrate in the figure reproduced in Figure 7.3, which represents a 'slice' of Figure 7.2 taken from x=32, it is quite possible to observe 32 alleles even when Pr(D) is as high as ~0.2 (Gill, Kirkham et al. 2007). Given such a potential range, the analyst is obliged to apply a confidence bound on Pr(D). Hence a value of D (designated D*) is chosen so that 95% of intervals of the form 0 to D* would contain the true value. This is achieved using a cumulative distribution function The likelihood function is converted to a probability function by 'normalising' it, i.e. rendering it so the area under a curve equals 1. The resulting graph represents the probability function of D (f(D | x)). From this a cumulative distribution function can be generated (Fig 4, reproduced from Gill et al 2007). In Figure 7.4, the y-axis indicates the probability that the value of D is smaller than the value on the x-axis. If a vertical line is drawn from x=0.15, it intersects the dotted line at ~0.95. Hence the analyst concludes that 'assuming only two contributed to this mix, it is possible to be 95% sure that the true value of Pr(D) is less than 0.15' (Gill, Kirkham et al. 2007)

Figure 7.3. Likelihood function for the probability of drop-out when x = 32 and n_c=2, (Reproduced from Gill et al 2007, p.136).

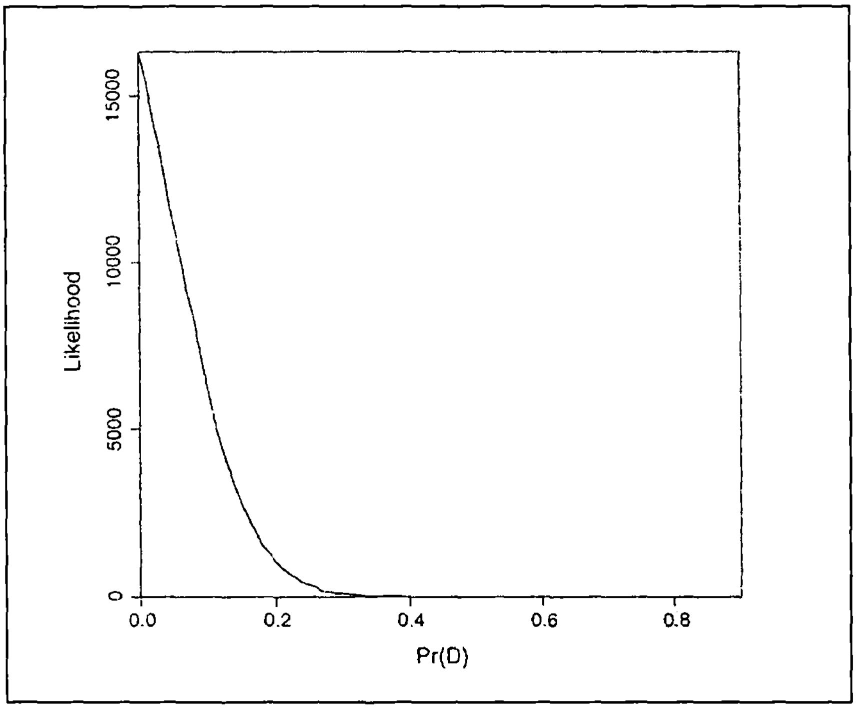
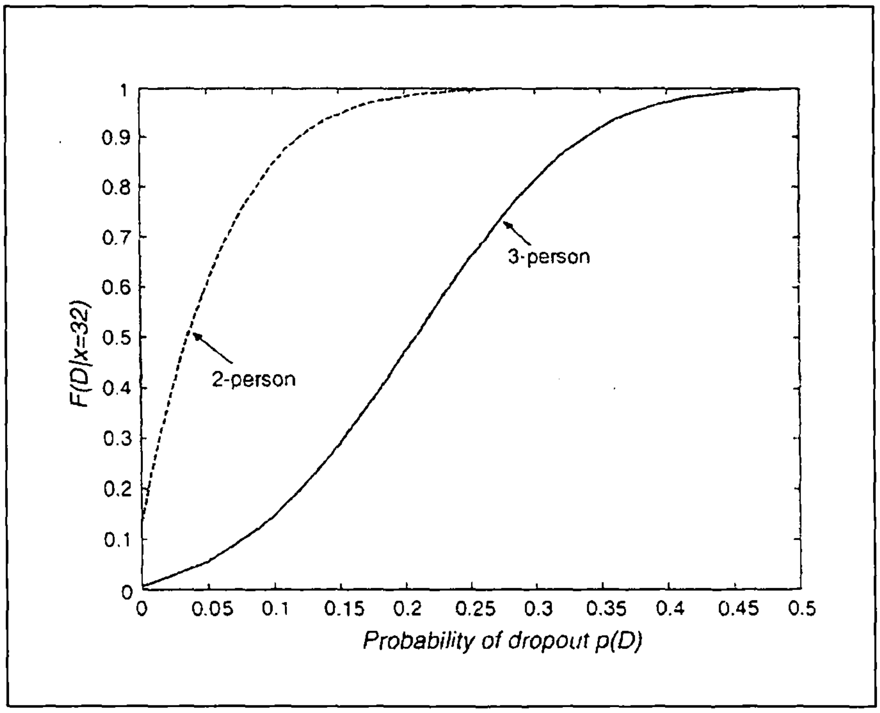


Figure 7.4. The cumulative distribution function for *F(D| x*=32) for a profile with 32 alleles, and for a 2- or 3-person mixture (Reproduced from Gill et al 2007, p.136).

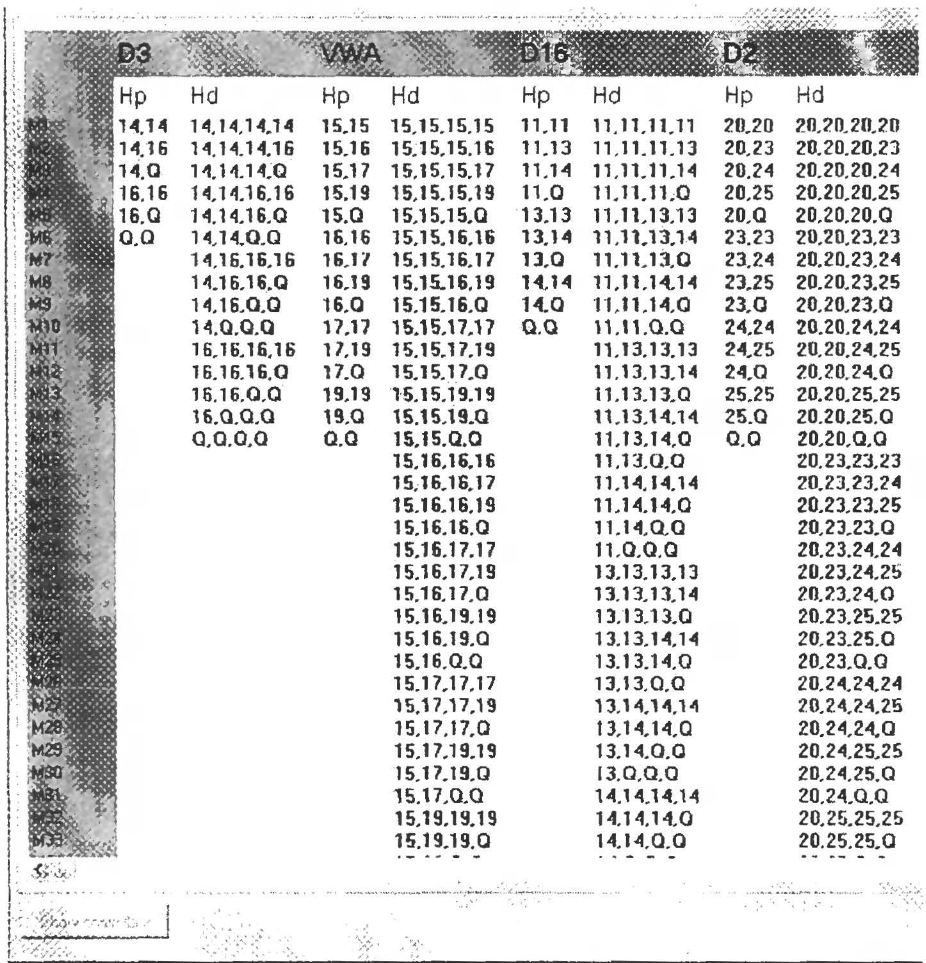


In practice, the LoComatioN system functions by presenting an results screen displaying the possible alleles for each locus (see Figure 7.5.). The analyst selects the alleles observed at each loci for each specific LCN profile, from a list presented on the input screen. Once these data are entered (a process which is apparently fast and convenient), the program requests the user to specify the hypotheses they wish to consider by way of a separate interface. Possible individual contributors may be considered under a particular hypothesis by the simple operation of selecting their name and adding them to a particular list. The number of unknown contributors can be specified by selecting the 'unknown' option the required number of times (Curran et al 2005). The final step involves the user/analyst requesting the program to consider the likelihood of possible genotypes occurring under different scenarios. Through another separate interface screen LoComatioN will return LR's for each genotype under each scenario. All proposed combinations are generally tested unless there is a reason to believe that certain combinations may be unreasonable in light of the data (Curran et al 2005, p.53).

The LoComatioN system may appear to represent an improvement over the 'consensus' method for the interpretation of LCN DNA profiles, at least in that it is able to return seemingly objective, numerical assessments of the likelihood of hypotheses, against the previous method, which was open to a significant degree of subjective judgement. However, a number of assumptions are built into LoComatioN, which I now discuss.

Some of the assumptions made in the construction of LoComatioN are questionable when considering the environment from which LCN crime scene samples are recovered. Due to the very nature of forensic work, LCN DNA samples are likely to be recovered from highly contaminated environments, and where there is a high likelihood of samples being degraded. As already discussed, whilst this a problem for conventional DNA profiling, contamination and degradation effects are even more prominent in LCN DNA analysis. Yet a closer examination of LoComatioN indicates that these effects have not been appropriately taken into consideration at an appropriate level of rigour.

Figure 7.5. User Interface Screen for LoComatioN (Reproduced from Gill et al 2007, p.137)



	D3		VWA		D16		D2	
	Hp	Hd	Hp	Hd	Hp	Hd	Hp	Hd
M1	14,14	14,14,14,14	15,15	15,15,15,15	11,11	11,11,11,11	20,20	20,20,20,20
M2	14,16	14,14,14,16	15,16	15,15,15,16	11,13	11,11,11,13	20,23	20,20,20,23
M3	14,Q	14,14,14,Q	15,17	15,15,15,17	11,14	11,11,11,14	20,24	20,20,20,24
M4	16,16	14,14,16,16	15,19	15,15,15,19	11,Q	11,11,11,Q	20,25	20,20,20,25
M5	16,Q	14,14,16,Q	15,Q	15,15,15,Q	13,13	11,11,13,13	20,Q	20,20,20,Q
M6	Q,Q	14,14,Q,Q	16,16	15,15,16,16	13,14	11,11,13,14	23,23	20,20,23,23
M7		14,16,16,16	16,17	15,15,16,17	13,Q	11,11,13,Q	23,24	20,20,23,24
M8		14,16,16,Q	16,19	15,15,16,19	14,14	11,11,14,14	23,25	20,20,23,25
M9		14,16,Q,Q	16,Q	15,15,16,Q	14,Q	11,11,14,Q	23,Q	20,20,23,Q
M10		14,Q,Q,Q	17,17	15,15,17,17	Q,Q	11,11,Q,Q	24,24	20,20,24,24
M11		16,16,16,16	17,19	15,15,17,19		11,13,13,13	24,25	20,20,24,25
M12		16,16,16,Q	17,Q	15,15,17,Q		11,13,13,14	24,Q	20,20,24,Q
M13		16,16,Q,Q	19,19	15,15,19,19		11,13,13,Q	25,25	20,20,25,25
M14		16,Q,Q,Q	19,Q	15,15,19,Q		11,13,14,14	25,Q	20,20,25,Q
M15		Q,Q,Q,Q	Q,Q	15,15,Q,Q		11,13,14,Q	Q,Q	20,20,Q,Q
M16				15,16,16,16		11,13,Q,Q		20,23,23,23
M17				15,16,16,17		11,14,14,14		20,23,23,24
M18				15,16,16,19		11,14,14,Q		20,23,23,25
M19				15,16,16,Q		11,14,Q,Q		20,23,23,Q
M20				15,16,17,17		11,Q,Q,Q		20,23,24,24
M21				15,16,17,19		13,13,13,13		20,23,24,25
M22				15,16,17,Q		13,13,13,14		20,23,24,Q
M23				15,16,19,19		13,13,13,Q		20,23,25,25
M24				15,16,19,Q		13,13,14,14		20,23,25,Q
M25				15,16,Q,Q		13,13,14,Q		20,23,Q,Q
M26				15,17,17,17		13,13,Q,Q		20,24,24,24
M27				15,17,17,19		13,14,14,14		20,24,24,25
M28				15,17,17,Q		13,14,14,Q		20,24,24,Q
M29				15,17,19,19		13,14,Q,Q		20,24,25,25
M30				15,17,19,Q		13,Q,Q,Q		20,24,25,Q
M31				15,17,Q,Q		14,14,14,14		20,24,Q,Q
M32				15,19,19,19		14,14,14,Q		20,25,25,25
M33				15,19,19,Q		14,14,Q,Q		20,25,25,Q

For LCN DNA, the assumption is made that the probability of allele drop-out is independent of locus (Gill 2005). However, this assumption may only hold where no sample degradation has occurred. If significant degradation has occurred, then loci where high molecular weight (MW) alleles are in existence may be preferentially affected (Gill et al 2007, p.130) However, in the LoComatioN system there is no sign that Pr(D) has been modelled to take into consideration MW-dependent degradation. In fact, the authors state that they estimate Pr(D) under the assumption that allelic drop-out occurs randomly of locus, yet this clearly does not take into account the possibility of preferential degradation which they acknowledge earlier in the same article (Gill et al 2007, p.133). Furthermore, there is no sign in the literature of how the problem of analysing a partially degraded allele may be approached. An allele may be degraded in such a way that it may appear as another genotype, yet this possibility is not mentioned in the literature. This reflects another more serious issue. It appears that, in the case of LCN analysis, degradation is not assumed, simply because degradation is something that is extremely difficult to be ascertained in LCN work. Hence the integrity of the profile cannot still be guaranteed. The lack of degradation can only be assumed in the intact non-crime scene samples used to determine Pr(D).

Other issues concern the estimation of the probability of contamination. Unlike the probability of drop-out, which involves a relatively complex process, the probability of contamination, Pr (C), was calculated from empirical data. Gill et al (2007) report that the calculation of Pr(C) was based on data taken from negative controls. This is most likely to have involved of 'blank' samples, containing no sample DNA, which are analysed alongside casework samples in order to verify for the quality of the work and to monitor the level of contamination based on either user error or environmental effects in the laboratory. Such blank samples will, if analysed collectively, contain extremely small quantities of alleles. The laboratory records used by Gill et al (2007) indicated an estimated contamination probability (Pr($\acute{C}$)= 0.05 per sample where:

$$\Pr(\hat{C}) = \frac{n}{LN}$$

Where n is the number of alleles observed in a series of negative controls and N the total number of negative controls analysed, and L is the number of loci tested per sample. The probability of any specific allele appearing as a contaminant is assumed to be the same as the frequency of its occurrence in the white Caucasian population, taken from a frequency database (Gill et al 2007)

Aside from the assumptions about allele contamination in relation to other ethnic databases, questions must be raised about the use of $\Pr(\hat{C})=0.05$ in the LoComatioN calculations. Basing this estimate on blank samples, taken from a laboratory, may not necessarily be reflective of the contamination challenges to be found within crime scenes. In the latter case, $\Pr(\hat{C})$ may be considerably higher. However, to date no such study has been carried out to evaluate the levels of contamination of DNA found in crime scenes.

The issues raised by these assumptions demonstrate the problems faced by scientists, in trying to model effects such as contamination and degradation. In this case a combination of both empirical data and theoretical modelling is used to help estimate these key parameters. Yet it is far from clear whether these techniques accurately reflect the kind of environmental challenges to DNA found in crime scenes. Of course, any attempts to generate empirical data from crime scenes are fraught with complications, and would be decidedly labour-intensive. This is in itself however, can be seen as a reflection of the demands of the Bayesian approach. In a way, the issues outlined here reflect the high informational demands that arise from putting Bayes into practical use.

Furthermore, certain interviewees were highly critical of the use of terms such as 'drop in' being used in any way to explain LCN profiles:

'You can't exclude people because when you invoke dropout, the reason why there's the alleles are not there is because they've dropped out. Or, it may be that the alleles are never there in the first place. Evaluate by Bayesian approach, you can't! (Interviewee 3, 2008)

This Interviewee, who was largely sceptical of the validity of the LCN technique, talked of the process of assigning an LCN profile as fraught with ambiguity and recourse to arbitrary decision-making:

> The LCN technique is so unreliable that what they have to do is say 'well we repeat it if only the alleles come up twice do we count them, Notice again theres a 17 in the same sample that's ignored, right, whereas, if you add it up that way, it discounts, now, here's the problem, you get a 14, and that becomes a mixture, and there's a 29, which is not in the accused. So you're looking good along here for the accused, 15, 17 16 19 11, OK? All matching the accused, 11 13 matches the accused but also matches this person. 11 14 matches so that there is entirely explicable by these two people here, right, 15 could have come from this person, 17, remember there could have been an unknown. 16 19 could only be down here. The eleven? Well theres no 11 in there. So that's all very awkward, and that's why the DNA analysts work back that way, they say 'well if this is the story, the evidence is consistent with, and that's a crap way of looking at evidence. Because the question really is 'how many other stories is it consistent with?' So that's an example of how they're using this Low Copy Number stuff, and its just guesswork (Interviewee 3, 2008).

Another interviewee described the use of 'drop in' in equally frank terms:

> 'Contamination is contamination. There's no degree of contamination...*contamination ruins evidence.*' (Interviewee 10, 2008, emphasis added).

The controversies surrounding the interpretation of LCN profiles shows that these 'inscriptions' cannot be regarded as 'immutable mobiles' which facilitate the unproblematic translation and passage of scientific information (Latour 1990). Instead, they represent spaces where ontological claims are contested, subject to both the critical scrutiny inherent to the scientific tradition and that of the adversarial legal system. Certain terms, such as 'drop in', and 'drop out', have been invoked to account for the imperfections supposedly apparent in the data, which have actively been constructed, in complex fashion, during the course of the generation of LoComatioN. From this discussion however it can be seen that these constructions have not fully satisfied critics from within the scientific community, and it remains to be seen how LCN will cope with future legal challenges.

The need to model these constructions in the manner described strongly reflects the informational demands of a Bayesian system, yet it is not clear whether the assumptions upon which they are based, accurately represent the reality they are trying to model. The construction of 'drop in' and 'drop out' in LCN analysis represents another set of practices which appear to be necessary for the propagation of Bayesian reasoning. The contestability of these notions show once more how Bayesian interpretation of DNA profiles involves a distinctly assumptive element, in contrast to the image often portrayed by Bayesians of a perfectly logical schema. As in the case of mixture interpretation, where peak height thresholds may be manipulated and experiential assumptions are made about key parameters, LCN analysis relies on a similar combinations of assumptions and constructive practices. This example strongly demonstrates how Bayes does not necessarily remove sources of subjective judgement and practices, but instead creates new spaces where they may be found.

7.8 DNABoost

The previous passages have described the problems associated with the interpretation of complex DNA profiles. They have demonstrated how the task of making assignations based on DNA evidence is by no means an uncomplicated process, with particular difficulties involved with regard to mixed DNA profiles, as well as LCN profiles taken from very small quantities of DNA. However, this section has also described certain technological approaches to these problems, which have used Bayesian probability theory to produce automated systems to allow scientists to overcome the complexities associated with DNA profile interpretation. Further, I have also attempted to show that technologies such as LoComatioN and PLS do not in themselves provide a means of interpretation which is entirely free from human influence. In fact a number of assumptions and ambiguities are built into these systems which call into question the extent to which such systems can be labelled as 'objective' means of interpreting profiles, and whether they present a truly significant advance on previous methods of assessment.

By way of a supplement to the previous discussion, I now wish to briefly introduce another technology which appears to share a number of characteristics with the systems described above. There does not however, exist any scientific literature in the public domain that enables one to take the same critical investigation taken above, despite the fact that this system was widely publicised in the media. In what follows I describe how this system was rendered as a convincing scientific technique, in a manner somewhat different to conventional modes of scientific persuasion.

In October 2006, the FSS announced the introduction of a new technique, known as 'DNABoost', which was heralded as representing another step forward in interpretation technology, a 'world first in bringing clarity to a type of sample that was previously difficult to interpret' (New Criminologist 2006). It was hoped that this new technology could improve identification rates by up to 40 per cent (Journal 2006) and crime detection rates by 15 per cent (Muir 2006, Telegraph 2006). Pilot schemes were launched among forces in West Yorkshire, South Yorkshire, Northumbria and Humberside. These schemes, intended to each run for three months, sought to test DNABoost on volume crimes such as burglary, theft and assault, with some hopes being expressed that the technique could eventually be applied to 'cold cases', namely previously unsolved cases involving serious crimes such as murders and sexual offences (The Journal 2007).

DNABoost appeared to have been intended for use to separate individual DNA profiles in cases where small quantities of biological material had been transferred and deposited. For example, the technique could possibly have been used in cases where more than one individual had touched a surface such as the arm of a chair. It was claimed that DNABoost would be able to separate out multiple DNA profiles in such a scenario. Details of the precise scientific basis of DNABoost have, so far, never been disclosed, although it appears that DNABoost was intended to be used in tandem with LCN techniques to improve interpretation rates (Morris 2006). It seems that, in functional terms at least, DNABoost was largely identical to LoComatioN.[13] The emphasis on the ability of DNABoost to resolve complex DNA profiles involving multiple individuals, each contributing small quantities of starting material, suggests that the technique may share similarities with LoComatioN and PLS, and indeed may have amounted to a combination of the two technologies. Thus, given the previous philosophical and computational approaches utilised by the FSS in the development of DNA profile interpretation technologies, it is perhaps reasonable to speculate that DNABoost involved some form of Bayesian computational approach. What is relevant to this discussion however, is that, to date, no articles have been published in the relevant scientific journals, outlining the theoretical basis of this particular technological form. This is in contrast to both LoComatioN and PLS, which have both been the subject of published papers.

What is certain however, is the extent to which the FSS invested a considerable amount of effort to publicise this new technique. They hired Medialink, a company who specialise in converting messages from companies into news stories to be distributed along a number of channels, including the internet and broadcast media. Medialink's website displays the logos of what appears to be a number of high-profile clients. Along with the FSS, the logos of companies such as Adidas, Ford, Nokia and HSBC are all prominently displayed (Medialink 2008). Further information displayed on Medialink's website provides further details of the work carried out by the company on behalf of the FSS, and yields some possible idea of the strategy of the FSS. As well as raising general awareness 'of the process [DNABoost] and of the FSS as an entity' (Medialink 2008), it states that FSS 'particularly wanted to encourage those police forces who were not taking part in the pilot scheme to

[13] In announcing the technique, FSS DNA manager Paul Hackett said: 'this particular technique is based on the foundations of existing DNA profiling technology so the laboratory-based techniques are exactly the same as we have used over the last 10 years, so that's very robust, very well established' (BBC News 4 October 2006).

buy the DNABoost service from them.' (Medialink 2008). Medialink conducted 'an intensive media relations campaign, dubbing the breakthrough '*CSI Britain*' (Medialink 2008, emphasis added), and 'successfully placed FSS spokesperson Paul Hackett on all major channels for live interviews and offered filming opportunities from the FSS Lambeth laboratories' (Medialink 2008). Furthermore, the efforts of Medialink, at least by the company themselves, were deemed to be a considerable success:

> 'The story really captured journalists' imagination (sic) and dominated throughout the news day, right from breakfast to 10 o'clock bulletins. Medialink obtained some of the most coveted spots in broadcast including: the sofa on GMTV; Radio 4's 'Today', Sky Sunrise, BBC Breakfast and by the end of the day the FSS had been contacted by all of their target police forces in the UK. Even Downing Street commented on the story!' (Medialink 2008)

The information found on Medialink's website raises some very interesting issues. First, it shows the extent to which the FSS wanted to sell DNABoost to other police forces not involved in the pilot schemes. By placing news stories in all the major UK broadcast networks, the FSS were able to obtain extensive advertising for their new product. It is perhaps pertinent here to recall that the FSS is in the process of being converted into an entirely commercial body, and already faces some competition from some private firms, most notably the company LGC. Here there appears to be a certain self-awareness on the part of the FSS to sell itself. However, the work of Medialink possibly goes beyond normal advertising. Rather than using commercials, news channels were used to act as a platform through which DNABoost was able to be exposed to police forces. These news stories effectively functioned as advertisements for the product. What is pertinent here is not only the way it was packaged, but that it was packaged within a particular type of news story – crime – hence an issue of concern to publics as well as police. Hence, DNABoost was not only being projected to police, but by also being disseminated to the public it possibly led them to both trust the science and perhaps also pressurise the police into taking up the technique.

This strategy seems to have worked – according to Medialink all police forces in the country apparently contacted the FSS to enquire about DNABoost.

It is also perhaps worth noting the prominence of the term 'CSI Britain'. Aside from the possible patriotic connotations (the FSS is keen to position itself as something of a world leader in the realm of forensic science), it is interesting to note the association with the American TV series. One can only speculate as to the real motivations behind the use of such a term, but there are perhaps certain connotations. The use of such a term lends a certain glamour to the technique, and could be viewed as reinforcing a sense that this is 'state of the art' technology. However, this possible overshadows the issue of the validation of DNABoost. No details concerning the validation of the technique have been published to date. In this case there have not even been any peer-reviewed articles released, as in the case of LCN.

7.9 Conclusion

The advent of technologies for mixture analysis and LCN can be seen to have contributed toward certain developments in policing practices in the UK. The potential to re-examine cold cases represents an extension of policing power which is also facilitated by the existence of the NDNAD. Thus it appears that these developments could be framed as contributing to a greater ability to extend the temporal reach of investigation. As my discussion has shown however, government policy has played a significant role in regard to this trend, and it is important to note that operations such as Advance were instigated through a centralised route. The overall success rate of these operations has been modest relative to the media interest in singular examples of convictions gained via the employment of the 'cold case' rationale. In this way, it is possible to view the operations as a whole as experiments in applying new technological means to investigation.

The prominence with which the successes of techniques such as LCN and PLS have been reported in the news media shows the interest that these supposedly cutting-edge technologies can attract. This of course, does not detract from the high level of controversy these technologies continue to draw. It appears that a number of figures in the forensic science community view the Caddy report as providing a necessary validation of the LCN technique. The future success of techniques such as LCN in securing convictions remains to be seen, but it is likely that the kind of technologies employed by the FSS will continue to be the source of debate. One issue that remains to be solved is international agreement on the efficacy of LCN. There has been relatively little in the way of progress in establishing a common set of guidelines for the employment of LCN. More seriously still, the response to the development of LCN has aroused scepticism rather than enthusiasm in other jurisdictions (Budowle et al 1991). The UK is relatively unique in employing LCN on such a relatively widespread basis, and the lack of international progress may reflect a less accepting stance to LCN in other countries (Gilder et al 2008).

Much of the debate has surrounded the need for adequate protocols to ensure that LCN analysis can proceed to enable the results to be free of contamination or compromise. Whilst these kind of problems invite relatively feasible solutions, in this chapter I have also sought to introduce areas of discord which concern a far less tractable set of issues, namely those concerning the problem of interpretation. The interpretation of both LCN and mixed profiles represents one of the most instrumental steps of the process in converting DNA material into evidence, and it is here that I have attempted to demonstrate how Bayes plays a vital role, in forming the calculative and philosophical basis of these technologies. In describing these technologies however, I have sought to show their development has rested on the employment on the use of constructions and assumptions whose correspondence to reality may be open to some question. For example, whilst PLS was meant to replace a highly intersubjective process of the assessment of mixed profiles, certain assumptions within the system are open to considerable scrutiny. For example, a key element in this technology is the values assumed to represent heterozygote imbalance (H_b). It is unclear to what extent this assumption has been verified and validated, yet it plays an important role in the production of suggested genotypes in mixture resolution. Furthermore, the calculation of mixture proportion (M_x) rests on the assumption that the mixture proportion is invariant across loci, yet empirical work suggests this assumption may require further testing. Amongst other issues, there is the possibility that the peak height threshold value may be set in such a way as to favour a certain hypothesis; again to date no scientific work has been reported which considers this issue. Despite all this however, PLS has been introduced into casework by the FSS, and has already claimed convictions.

A similar situation appears to exist with regard to LoComatioN, the system developed by the FSS to resolve LCN profiles. FSS scientists have constructed a certain vernacular which appears to be used to justify their propositions, namely the use of 'drop in' and 'drop out'. To date, little scientific study exists which serves to rigorously consider whether such terms are appropriate; this issue is not addressed in the current literature on LCN. Furthermore, the use of laboratory data to assume a certain probability of contamination (Pr(C)) may not necessarily reflect the kind of environmental challenges that DNA from crime scene samples may actually face. The probability of drop-in, Pr(D), on the other hand, is calculated via computer simulation. Whilst the literature reports a seemingly rigorous process for this calculation, it does not appear to consider the possibility that drop-in could be allele-specific, nor does it take into consideration environmental challenges. Both assumptions concerning contamination and drop-in therefore, could be said to rest on matters of convenience however, as it must be acknowledge that further consideration of these issues could take up considerable time and resources.

In this way, both technologies could be construed as involving certain practices of black-boxing (as discussed in Chapter 2). They contain assumptions which, if tested further, could incur considerable costs for the developer. In discussing the controversies surrounding these technologies, I have sought to show that the black boxes are not entirely invulnerable to interrogation. This case study therefore shows, further to studies such as Derksen (2000), that supposed 'technological progress' does not involve the continued excision of areas of subjectivity; indeed, this study appears to indicate that with new technologies come new sources of contestation and ambiguity, as highlighted by the vehement criticisms emanating from other members of the forensic scientific community.

With regard to the consequences for Bayesian reasoning, it must be noted that Bayes has been criticised for placing extremely large informational demands on the reasoner. It is unclear to what extent these demands may stretch to, but it is clear that practical circumstances may limit the precise quality of information being fed into Bayesian algorithms. Yet what this study highlights is not necessarily the problem of obtaining quantitative data, but the assumptions and practices which go into creating this data. In this way Bayesianism can be seen as being centred around a certain set of practices. Without these practices, the construction of Bayesian technologies cannot take place.

Bayesianism is seen as a means of assessing 'subjective' measures of probability; but what is unclear is precisely what kind of subjectivity is being assumed. This case study appears to indicate that this subjectivity, or to be more precise, inter-subjectivity, is actually far more complex than first thought. What this case study demonstrates is that the kinds of 'subjective' probabilities under assessment owe their existence to a heterogeneous mix of 'objective' and 'subjective' underpinnings, which in turn reflect a particular context. The 'Bayesian technologies' that arise are a result of this context, yet the same epithet also acts as a means of organising these underpinnings. It not only provides a basis for making sense of information gleaned from DNA, but it also acts as a framework for facilitating 'scientific' practices.

Finally, one may wish to consider the question of how the continuing development of new technologies, produced through companies operating in an increasingly competitive and commercialised market, may be brought to the attention of governments (and by extension, their electorate), as much as they be made to appeal to police forces who are ostensibly their primary customers. The example of DNABoost appears to suggest that this could be the case. The slightly troubling aspect of this example concerns the manner in which direct appeal was made to policymakers, publics and police. The PR campaign for DNABoost effectively subverted supposedly unbiased news media, leading to the acceptance of a story which effectively functioned as an advertisement for the FSS's product. Moreover, no scientific data has ever been released into the public domain concerning DNABoost. This development leads one to consider whether, in the increasingly competitive market for forensic science services, developers of technology may begin to direct efforts towards convincing a more wider set of stakeholders, namely government and publics, in the apparent worth of their products. As the DNABoost example indicates, such efforts may involve strategies other than the transparent release of information long considered a mainstay of conventional scientific conduct.

CHAPTER EIGHT – CASE STUDY 2: THE CASE ASSESSMENT AND INTERPRETATION (CAI) MODEL

8.1 Introduction

In this chapter, I move from a consideration of the use of Bayes in support of specific instances of DNA interpretation, to examine the wider use of Bayes in forensic science in general. To do this I introduce a model of forensic scientific investigation which has been constructed by a number of senior forensic scientists and statisticians working for, or on behalf of, the UK Forensic Science Service (FSS). This initiative, known as the Case Assessment and Interpretation (CAI) model, has been under development since 1998, and has been the subject of a number of articles, published predominantly in prominent forensic science journals[1]. The stated aim of the CAI model is to optimise the cost-effectiveness of the casework process, by providing a more efficient and focused procedure for comprehending evidence in the context of criminal investigations. A key feature of the CAI is the use of a Bayesian framework for the evaluation of hypotheses formulated in relation to the evidence that becomes available to those investigating a criminal case. The CAI is intended to be applicable to any form of evidence and also represents a strategic attempt to promote and shape the uses of Bayes within a forensic scientific context.

In the first part of this chapter I present an overview of the development of the CAI, and introduce the model's central elements. After outlining the theoretical architecture of the model, I then explore a series of issues that arise when the CAI is directly applied to investigative casework. In the ensuing discussion of these issues, I demonstrate how they comprise a series of practical contingencies that are not fully amenable to a Bayesian representation of reasoning. I question the extent to which the application

[1] See Cook et al (1998a, 1998b, 1999), Evett et al (2000a, 2000b), Jackson et al (2006)

of Bayes to forensic science is able to reconstitute investigative reasoning practices along more explicitly rational lines. Instead, I argue that the application of Bayes in this case is necessarily shaped by a wider array of individual, organisational and institutional practices and processes. The form(s) of 'Bayesian reasoning' outlined in this chapter are not best understood as the top-down imposition of a mode of reasoning which draws justification from supposedly transcendental logical principles. On the contrary, I argue that these reasoning forms are co-dependent on the practical contexts in which they are inextricably embedded. Hence the forms of Bayesian reasoning that are simultaneously applied to, and arise from, these contexts, are fluid and flexible accomplishments.

8.2 Overview of the CAI

The Case Assessment and Interpretation (CAI) model has been in development for over ten years (Cook et al. 1998a) Responsibility for the formation and evolution of the model can be traced to work done for the FSS by a team of senior forensic scientists. These originators of the CAI cite new commercial imperatives as representing one of the main drivers for the development and application of the model. Since 1996, the relationship between the FSS and the police forces they serve has been one of customer and provider; the FSS charges police forces for the analysis of each piece of evidence requiring assessment, and the CAI has been developed to enable the FSS to optimise the efficiency of the service it can provide to police forces (Cook et al 1998a). What this has entailed is a model which has been used as a way of estimating the contribution that each piece of evidence is able to make to the investigation – and potential prosecution - of each specific criminal case.

In what follows I provide a brief history of the development of the CAI. I first discuss the organisational changes that saw a change in the way the FSS conducted business with the police. I show how this drove the development of the model, with particular reference to the creation of new systems of customer feedback. I outline the proposed system of consultation between forensic scientists and police investigators, and the new system for managing the relationship of criminal investigation and the analysis of evidence.

8.3 History of the CAI

In the first of a series of papers outlining the CAI, Cook et al (1998a) discuss an apparent change in 'culture' affecting not only the FSS, but police forces as a whole. Following the re-positioning of the FSS as a government agency in 1991, direct charging for casework services was introduced in order that the costs of forensic science would become more directly apparent to operational police officers (Cook et al 1998, p.152). This took place as police forces themselves were devolving increasing financial responsibility to officers who were more closely involved with day-to-day casework (Cook et al 1998a, p.152). With police forces being redefined as the primary 'customers' of the FSS[2], the latter adopted a specific notion of precisely what constituted it's 'products', defining them in terms of: 'an activity; the time taken for it; the cost; the standards to which the activity adheres; the expected outcome; and the chargeable unit' (Cook et al 1998a, p.152). At the heart of the 'new culture' was a sense that customers (in this context, criminal investigators) would benefit from a greater degree of consultation over the way in which the scientific aspect of specific casework proceeded. Thus, the aim of direct charging was to: 'enable the customers to make better decisions about how to allocate their own resources, in turn creating a greater sense of value for money' (Cook et al 1998a, p.152).

It must be noted however, that behind this point is the assumption that customers (police officers) have adequate knowledge of the circumstances

[2] Cook et al (1998a) reported that, at the time of writing, 95% of business was coming from the British law enforcement agencies, with the remaining business involving private law firms and overseas organisations (Cook et al 1998a, p.152). Whether this remains the case in 2008 is a matter of conjecture); however what may be of interest in the future are the possible

behind a case to be able to make informed decisions about the nature of the forensic scientific services they require. One stated aim of the CAI was not only to provide better value for money, 'but also of achieving improvement through a genuine partnership in which the customer has a greater participation than hitherto in decisions about what work is done in the laboratory' (Cook et al 1998a, p.152). This quote is of particular interest, in its reference to the sense of 'partnership' in terms of how laboratory-based inquiry could proceed. This apparent need to accommodate the interests of actors external to the direct task of scientific inquiry points to an obvious sign of the way in which economic imperatives have driven a process which may be construed as the 'co-production' of forensic scientific authority (Jasanoff 2004). This sense of 'partnership' differs from the view of Fraser (2007), who argues that the introduction of direct charging has led to a 'shift from the expert power of the laboratories to that of the police' [Fraser, 2007 p.384]. The CAI however, indicates a more proactive role for forensic scientists. If the assumption of the police having perfect knowledge of their forensic scientific needs is relaxed, then it may actually indicate, if anything, an enhanced influence of forensic science providers on the course of criminal investigations.

Another aim of the CAI has been to convey the notion that the process of interpretation of evidence, defined by the authors as 'the drawing of rational and balanced inferences from observations, test results and measurements' (Cook et al 1998a, p.152), is a process that begins at the start of a criminal investigation. Hence the interpretation of evidence is not simply taken to occur at the final stage of examination, once all the evidence has been accumulated. Instead it is viewed as a constantly recursive and collaborative course of action that involves a significant degree of consultation, feedback and input from other actors from the outset. This approach was taken to contrast with an older model of investigation, in which suspects were identified through non-forensic means, with the collection and deployment of forensic evidence largely being informed by a

consequences of the FSS becoming a wholly private, independent company, as has been mooted several times.

concern to more authoritatively incriminate these suspects in the actions and events of interest (Interviewee 8).

In a subsequent paper, the definition of 'interpretation' in this particular context is fleshed out further:

> 'Interpretation is, of course, a part of everyday life and it is possible to visualise a kind of spectrum. At one extreme there is pure intuition, which defies rational analysis. At the other extreme is pure logic. Scientific judgement cannot be based on pure intuition or 'hunch'...the scientist should, as far as possible, be able to rationalise the opinion that is presented. The opinion might be supported by data: there is, indeed, a kind of data spectrum ranging from the case where the opinion rests largely on an established data collection to the case at the other extreme where there are no data and the opinion is entirely based on experience. In any case, the expert opinion is necessarily subjective, but it should always confirm to logical principles. Those principles are furnished by considering probability theory as a means of reasoning under uncertainty: this leads to the Bayesian view of evidence interpretation...' (Evett et al. 2000b, p.234)

Of interest here is the slightly paradoxical reference to subjective expert judgement adhering to 'logical principles'. However, as one of the authors argues elsewhere, it is Bayesianism which is viewed as the naturally logical means through which to reason in such a way.

> 'It's a good idea not to use the word 'Bayesian' but just to call the approach 'logical'. Nobody is going to attack you for being logical!...That framework—call it Bayesian, call it logical—is just so perfect for forensic science. All the statisticians I know who have come into this field, and have looked at the problem of interpreting evidence within the context of the criminal trial, have come to see it as centring around Bayes's Theorem.' (Evett and Joyce 2005, p.37).

Hence in this instance the application of Bayes to forensic scientific problems is viewed as a natural, and possibly inevitable phenomenon. Indeed as the quote above demonstrates, Bayes is often cited by it's acolytes as the *only* logical way to reason under conditions of uncertainty. In this context Bayesianism is thought of as embodying certain key principles informing the correct practice of evidence interpretation (Cook et al. 1999):

- Interpretation of scientific evidence is carried out within a framework of circumstances, dependent on the *structure* and *content* of the framework.
- Interpretation is *only meaningful* when two or more competing propositions are addressed.
- The role of the forensic scientist is to consider the probability of the evidence *given* the propositions that are addressed.

The authors argue that the primary role of the forensic scientist in a criminal investigation is to consider the repertoire of claims and allegations of investigators, advocates and witnesses which together comprise the framework of circumstances in which the scientist operates (Cook et al. 1999). Furthermore, they assert that it is the responsibility of the scientist to formalise this repertoire in order to assess various case elements in accordance with the interpretive principles outlined above.

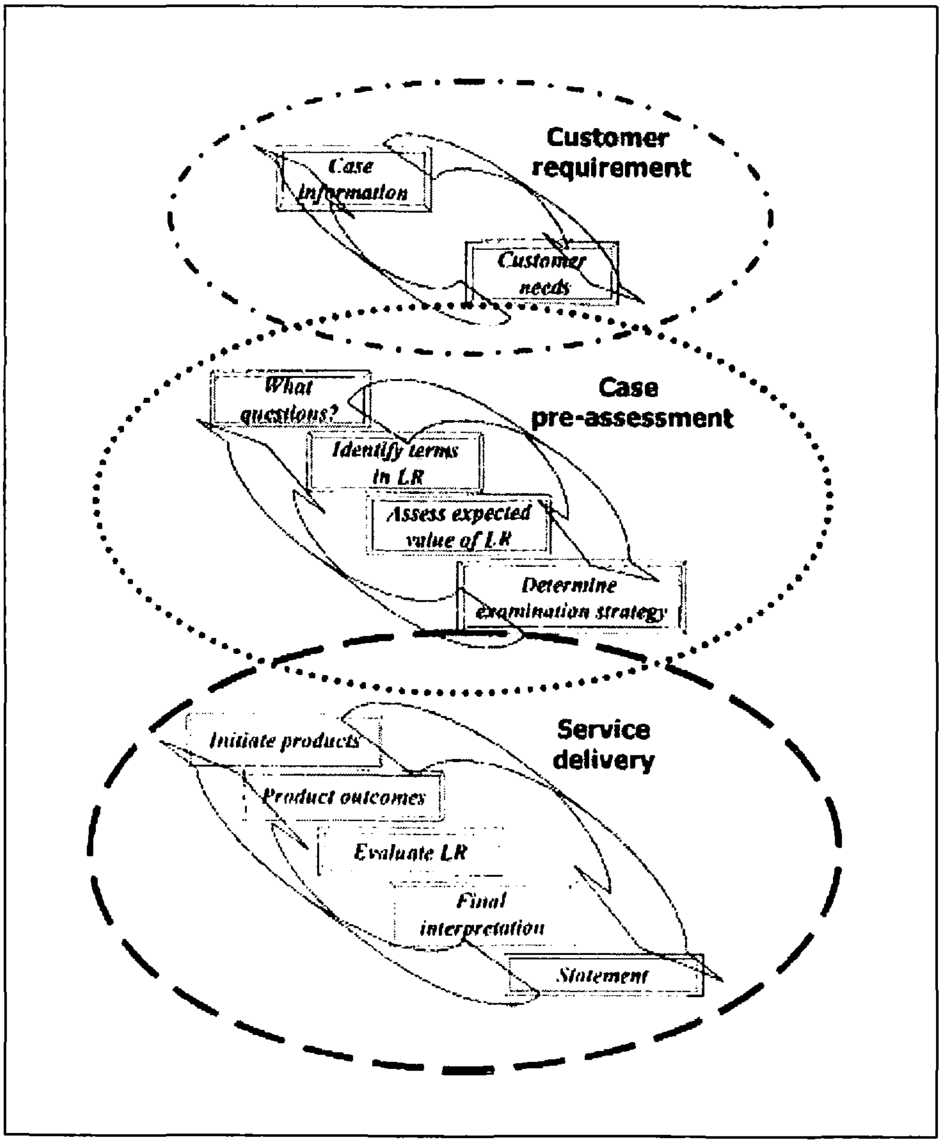
8.4 Principles of Forensic Interpretation

The CAI has been described as encompassing three distinct, but interlinked phases: an assessment of *customer requirement*, a *case pre-assessment* phase, and *service delivery* (Cook et al 1998a, p.152). As shown by Figure 8.1, each phase is broken down into a series of processes, which may interact in the manner of feedback loops, representing the emphasis on the continual review and re-appraisal of the outputs.

The *customer requirement* phase, as the name entails, requires an assessment of the customers needs. The developers of the CAI stress the need for good communication in this regard, in order that the scientist has 'an adequate appreciation of the case circumstances so that he/she can set up a concise framework for thinking about what kind of examinations may be carried out and what may be expected from them'. (Cook et al 1998a, p.153). Construction of such a framework of circumstances is taken to be a necessary pre-requisite for the development of propositions relevant to the evidence and the case. The authors stress the need for scientists to take a 'balanced view' of each case, in line with what they regard as the principles of 'the Bayesian view of evidence, that it is not sensible for a scientist to attempt to concentrate on the validity of a particular proposition without considering at least one alternative' (Cook et al 1998a, p.153). At each stage of the process, scientists are obliged to consider at least two competing propositions, most commonly relating to a proposition relevant to the prosecution hypothesis, and another relevant for the defence hypothesis. A key part of this initial phase involves the identification of precisely what kind of propositions may be assessed. In some cases, the proposition of interest may appear to be somewhat removed from the crime under investigation, possibly relating to the investigation of the origins of transfer evidence, e.g. 'these fibres came from this garment'. Such a proposition may relate to the crime due to the wider framework of circumstances. For example, the fibres in question may have been recovered from the scene of an apparent burglary. Subsequently a possible suspect may be arrested, wearing a garment which might match with the fibres recovered from the scene. Forensic scientists may wish to determine how significant any match may be, particularly if the garment in question is unusual or rare.

In prescribing such behaviours, the CAI marks a notable change from the 'traditional' role of forensic evidence. In this model of investigation, forensic evidence was only identified and analysed if it was regarded as contributing toward the construction of a case against an individual normally already suspected by the police. The decision to carry out particular forensic tests was often the preserve of police officers themselves, who may have perceived certain items of evidence as yielding potentially incriminating forensic information, regardless of their relative lack of scientific understanding. The CAI however, obliges scientists to take a more neutral and balanced view of evidence, requiring them to consider defence scenarios in addition to prosecution hypotheses. The more considered approach to proposition selection also reflects the more commercial orientation of forensic science. The CAI aims to minimise the

Figure 8.1 Outline of the Case Assessment and Interpretation (CAI) Model (Reproduced from Cook et al (1998a) (Cook et al 1998a, p.153)



wasting of resources on redundant tests which may yield little or no useful data for an investigation.

Once the type of questions that require consideration have been identified, the scientist is able to move into the *case pre-assessment* phase. Here, the scientist would be expected to further clarify the propositions they wish to investigate, by possibly re-describing them in more quantitative terms. A key component of this phase is the generation of likelihood ratios (LRs) which take the form:

$$\frac{\textit{Probability of the evidence if prosecution proposition is true}}{\textit{Probability of the evidence if defence proposition is true}}$$

This form enables the scientist to represent his/her expectations in a more precise and quantifiable manner. For example, in an assault case it may be necessary to examine an item of a suspect's clothing in order to determine whether any fibres from a victim's clothing might have been transferred in the course of an assault being carried out. This might lead to the scientist to express the following expectation: 'if this proposition is true, then I would expect to find x numbers of transferred fibres'.

The assessments made in this phase inform the manner in which work is carried out in the final *service delivery* phase. This involves the execution of relevant tests in accordance with the case pre-assessment phase. Statements will eventually be prepared by the scientist outlining the findings of these tests. What is made clear by the authors is that this process is viewed as sensitive to feedback and is thus not regarded as following a strictly linear path, as indicated by the curved arrows in Figure 8.1 (Cook et al 1998a, p.153). Hence the process is continually subject to review and possible further consultation with the customer. Whilst the initial direction of forensic inquiry may be informed by the circumstances of the case, new lines of inquiry generated by subsequent developments are able to be readily incorporated into the framework.

8.5 Investigative vs Evaluative Mode

In a subsequent paper, the authors of the CAI have put forward a three-stage model of hypothesis formation and testing in forensic scientific casework (Jackson et al. 2006). In this work, it is proposed that forensic scientists operate initially in an 'investigative' mode, followed by an 'evaluative' mode, which may break down further into a 'preliminary evaluative' stage and then a 'fully evaluative' stage depending on the availability of evidence (Jackson et al 2006). Table 1 (Jackson et al 2006) compares the procedures and behaviours of forensic scientist in the investigative and evaluative modes. A number of general differences can be seen. In the investigative mode, a suspect may not yet have been identified, and thus the emphasis focuses on gaining information from examination of the incident scene, and attempting to explain that evidence. In the evaluative mode, there is more emphasis on the suspect. Furthermore, in the investigative mode, the conjectures may be based largely on the opinion of the scientist. Whilst these may be largely a matter of experience, they may ultimately take the form of prior probabilities of a particular cause to account for the incident. The investigative mode breaks down into seven activities: first, the making of observations; second, the generation of a set of hypotheses, exhaustive within the context of the case, to explain the observations; third, the assignation of prior probability measures for these hypotheses; fourth, the generation of likelihood ratios for the range of possible future observations given that each of the hypotheses were to be true; fifth, the further search for observations; sixth, the generation of posterior probabilities for each hypothesis and their re-ranking in order of probability; and seventh, the communication of opinions to the customer (Jackson et al 2006).

If, via the investigative phase, it is possible to generate two mutually exclusive, competing propositions, Bayes theorem may be employed. Jackson et al (2006) state that the role of the scientist in this case is to generate likelihood ratios (LRs) for the scientific findings, and it is here that the scientist can be seen to be making the transition into an 'evaluative' mode. Jackson et al (2006) define an evaluative opinion as 'an expression of the magnitude of the LR' (Jackson et al 2006, p.38). The evaluative stage breaks down further into 'preliminary evaluative' and 'fully evaluative' stages. The preliminary evaluative stage involves the construction of propositions against which likelihoods concerning the evidence may be derived. Typically (although not necessarily exclusively), pairs of propositions are compared, pertaining to a prosecution and defence hypothesis respectively. Once a comparison of likelihoods are possible, a likelihood ratio can be generated which indicates the relative likelihood of one proposition against the other. If such a ratio can be derived, it is said that the investigation has entered the fully evaluative stage.

Table 8.1. Summary of the roles of investigative and evaluative scientists. (Taken from Jackson et al 2006, p.43).

Investigative Mode	**Evaluative Mode**
The issues addressed by the scientist tend to be centred on, or generated by, the incident	The issues addressed by the scientist tend to be centred on the suspect/defendant
The scientist begins the process by making observations to help answer the key questions	The scientist begins the process by considering case circumstances, the allegations and the issues
Hypotheses are generated to explain the observations; explanations may be speculative and/or imaginative and/or novel	Propositions are generated generally from prosecution and defence positions
The scientist will seek new observations to strengthen or weaken belief in the explanations	The scientist makes observations and checks against her predictions of the outcomes, given that each of the propositions were true
The scientist may operate from one perspective	The scientist is required to be impartial in approach
If the probability of the observations, given the truth of the explanation, is very low, the scientist will seek clarification and re-assessment of the suitability of the explanation.	The scientist is not concerned, generally, about very low probabilities for the observations. Concern might be raised of the probability for the observations, given the truth of EACH of the propositions, is very low.
Opinions provide direction to the investigation; options for further investigation may be generated or eliminated	Information is presented to assist the court in addressing intermediate and ultimate issues
The scientist may provide an opinion on the truth or otherwise of an uncertain event	The scientist provides information to help the court to form an opinion on the probability of an uncertain event
Prior probabilities, either generated by the scientist or given by others, are taken into account explicitly by the scientist	Prior probabilities are the remit of the court
Investigative opinion may be provided by the scientist as a posterior probability	Evaluative opinion is provided in the form of expression either of the LR or of single likelihoods; posterior probabilities are for the court
The scientist is findings-led	The scientist is propositions-led

The evaluative mode eventually culminates with the scientist being obliged to report likelihood estimates to court which are intended to help the jury in their deliberations. The CAI therefore allows for the fact that the latter have the ultimate say in the matter of guilt or innocence.

8.6 Formation of Propositions[3]

Evett et al (2000a) describe in more detail how the CAI may be used to construct hypotheses in relation to casework (Evett et al 2000a). Through their work with operational forensic scientists they posit that the formation of propositions involves an intermediate stage, in which various explanations for the evidence may be tentatively advanced (Evett et al 2000a, p.4). The formation of explanations will necessarily be dependent upon the framework of circumstances relevant to a particular case, and which may well be a contingent and rather fluid process, depending on certain factors (such as certain pieces of evidence being put into doubt, or being rendered inadmissible etc) (Evett et al 2000a, p.5).

In their discussion of the processes leading to the use of explanations through which to generate propositions, Evett et al (2000a) give a number of examples of the kind of explanations which may be considered during the course of a rape case in which no trace of semen has been recovered:

'1. Intercourse did not take place.
2. Intercourse took place and there was no ejaculation.
3. Intercourse took place but a condom was used.
4. Intercourse took place but the complainant used a vaginal douche.
5. Intercourse took place but all trace of the semen has been lost.' (Evett 2000a, p.5)

[3] It is possible here to view the division between explanations and propositions as relating to the two-stage model of scientific inquiry as put forward by the 19th century American philosopher Charles Sanders Peirce. He viewed the scientific process as consisting first of a stage of hypothesis formation, known as *abduction*, followed by a process of hypothesis testing, which he referred to as *retroduction*. In the case of CAI reasoning, it can be seen that the process of explanation-generation relates to the abduction stage, whereas the proposition stage relates to the retroductive stage.

Evett et al (2000a) use this set of examples to demonstrate the varying degree to which each explanation accounts for the evidence. Using probabilistic terms, they emphasise the relationship between the semantic content of explanatory statements and background information. In this particular set of examples, No.5 can be seen to provide a complete explanation of the observation. Evett et al refer to such an explanation as *prescriptive*. In more formal terms, the probability of the observation, *given the truth of No.5*, is equal to 1. However, for a given observation *S*= 'no semen was recovered', it would be that the probability for each of the other explanations would also be close to 1 (Evett et al 2000a, p.5). Other background information is therefore necessary in order to compare the explanations in a more meaningful fashion. For example, in order to compare No.1 with the others one might wish to know if the suspect and the complainant had ever met, and if so under what circumstances. Other information, such as past behaviour and experience of the individuals might help investigators to more effectively differentiate between the plausibility of the explanations. Assessment of the explanations may enable scientists to re-frame them as propositions, which may aided by the existence of background information able to be translated into quantitative terms[4].

[4] Evett et al (2000a) highlight six different aspects which differentiate propositions from explanations. It may be of interest here to quote these in full:

Propositions come in pairs. An essential component of the approach to interpretation followed in these papers is the need to address two propositions at any one time: one prosecution and one defence. Explanations can be generated without mutually exclusive alternatives.
Propositions need background information. An explanation needs no circumstantial evidence?; it may be completely speculative and introduce new features hitherto not thought of. A proposition can only be generated from a stated framework of circumstances and assumptions.
Propositions are formal. Explanations may be generated quite informally, almost in a 'brainstorming' sense. However, propositions must have a logical relationship with the framework of circumstances.
Propositions are mutually exclusive and exhaustive. Within the stated framework, the two propositions will satisfy this requirement. Explanations, being informal, may be interrelated, intersecting and open ended.
Prior odds relate to propositions. The framework of circumstances will enable a jury to assign conditional probabilities to the chosen propositions. Explanations may be without any rational probability.
Propositions relate to inference. It follows that propositions are capable of rational inference. Explanations will, in general, not be testable in a logical sense' (Evett et al 2000a, pp.5-6, original emphasis).

Evett et al (2000a) use the above example to show how the framework of circumstances may affect the kind of propositions that are generated depending on the type of information that may be available. If the possibility is considered that a complainant may have been under the influence of drink or drugs, then the construction of propositions may become more difficult. However, if the complainant is able to recollect clear details about the incident, such as whether the offender wore a condom, whether ejaculation took place, or if the complainant had not had prior intercourse for a month etc, then it may be possible to frame pairs of propositions, along the lines of:

Prosecution proposition: The complainant had intercourse as alleged
Defence proposition: The complainant had not had intercourse for at least a month

The kind of information that may be garnered from the framework of circumstances will vary from case to case. Typical sources of information may include details of times and activities, which could also guide scientists in the formulation of meaningful propositions specific to each case. For example, scientists may be able to use prior data about the persistence of semen in the vagina after intercourse to consider the probability of the evidence given each of the above propositions.

8.7 The Hierarchy of Propositions

The *hierarchy of propositions* plays a central role in organising the way in which propositions are constructed and assessed. This hierarchy classifies propositions along three lines, in order of their relevance to the ultimate issue under consideration by the court (Cook et al. 1998b). Level I, or the *source* level, relates to propositions concerning the origins of the evidentiary material. For example, source-level questions may concern the origin of glass fragments in a case of breaking and entering. Scientists may wish to ascertain the likelihood of the glass fragments originating from a particular broken window. In this case, scientists could consider two hypotheses: a prosecution hypothesis which would seek to ascertain the probability that the glass fragments originated from the window implicated in the case, and a defence hypothesis considering the probability that the fragments originated from another window not involved in the act of breaking and entering. To cite another example, a set of Level I propositions could be formulated to consider the origin of blood on the clothes of a suspect under suspicion in an alleged assault case. A prosecution hypothesis would seek to ascertain the probability that the blood came from the victim of the assault, whereas a defence hypothesis would seek to consider the probability that the blood originated from another individual not involved in the assault. In each instance, assessment of the sets of hypotheses is based on observations, measurements and analyses. In the case of Level I questions, the process of assessment is generally considered as being exclusively within the ambit of scientists, as the means of assessment at this level will exclusively involve the measurement and comparison of quantitative data (Cook et al 1998b, p.232-233).

Level II in the hierarchy, the *activity* level, involves a greater element of reconstruction of the events in each case. Re-visiting the example of the breaking and entering case, one may wish to pose a prosecution hypothesis, concerning the probability that a given suspect smashed the window, against a defence hypothesis seeking to ascertain the probability that the suspect was not present when the window was broken. In the example of the alleged assault, the prosecution hypothesis may concern the probability that a suspect attacked the victim by kicking the latter in the head. The defence hypothesis, meanwhile, may be concerned with the probability that the suspect was not present when the victim was kicked in the head.

A key difference between Level I and Level II propositions is that, whilst the former may be addressed via strictly scientific means, Level II propositions cannot be addressed without taking a specific framework of circumstances into consideration (Cook et al 1998b, p.233). To address the Level I proposition of the probability of glass fragments originating from a particular window, it could be feasible to analyse the refractive index of the fragments against a database containing data on the refractive indices of glass windows from a number of different manufacturers (Cook et al 1998b). However, if the investigators wish to address the probability of finding a matching quantity of fragments about a suspect given that it was they had smashed the window, they would require further circumstantial information. In this case it would be of use to be able to ascertain how the window was smashed, and the time interval between the incident and the removal of the clothing from the suspect for analysis. In order to assess a corresponding defence hypothesis, in this case the probability of finding matching glass on the suspect if they were unconnected with the smashing of the window, it may be useful to have information which would predispose the suspect to have glass present on his clothing, such as lifestyle or profession (Cook et al 1998b, p.233). Hence the more information a scientist may possess in relation to the circumstances of the case, the more fully Level II propositions will be able to considered, and the more bearing that insights generated at Level I will have on consideration of Level II propositions.

However, it is unclear just how much information *is* required before the transition from Level I to II can be made. Here, it seems that the deeply contextualised nature of specific criminal investigations necessarily limits the extent to which the CAI can be fully systematised. Furthermore, it would appear to be the case that the amount of information available to investigators influences the type of propositions that can be formed. Certainly, it does appear to be the case that the role of the forensic scientist becomes more directly involved in the progress of the case. Rather than simply analysing material to gain neutral, quantitative results, the forensic scientist is obliged to learn more about the specific circumstances of each case. This however, can be seen to have certain consequences. In addition to possibly widening the workload of operational forensic scientists, in terms of a greater amount of information having to be considered, this also has a range of ethical and pragmatic implications. The primary concern in this regard is whether knowledge of the circumstances of a case may lead to bias in the assessment of evidence (Koppl 2005). Regardless of this, it can be seen that the transition from Level I to Level II necessitates greater interaction between police investigators and forensic scientists in order for the contribution of the latter to lead to more meaningful inferences.

The emphasis on the role of forensic scientist in generating Level II propositions raises some questions about the division of ratiocinative labour in the course of investigating a criminal case. The kind of level II propositions outlined in the literature appear to be the kind of lines of enquiry that one would expect the police to pursue. Hence the remit of what could be viewed as 'forensic science' is possibly stretched; what the CAI can be seen to represent in this regard is a possible extension of what could be seen as constituting forensic scientific inquiry. Rather than a concern with the recovery, rendition and analysis of items of possible evidential interest and importance, 'forensic science' under the realm of the CAI is being extended to constitute a possible method for the advancement of criminal investigation as a whole.

The final level in the hierarchy, Level III or the *offence* level, concerns the probability that a suspect has committed a criminal offence. Level III propositions are the domain of the jury, assisted by the judge (Cook et al 1998b, p.233). Offence-level propositions may also be construed as activity-type propositions; however the key difference in the case of Level III propositions is that they concern the question of whether an actual crime has occurred. Although Level III propositions are assessed by the jury, and ultimately outside the domain of the forensic scientist, the latter may well be able to assist the court in their deliberations, by reporting information concerning likelihoods in a suitably comprehendable form (Cook et al 1998b, p.233). As the potential work involved in incorporating Level I data into Level II data demonstrates, a considerable amount of work may be involved in producing propositions and inferences of use to forensic scientists before they can then assist courts with ruling on Level III propositions. In many cases, Level II and Level III propositions may resemble each other; what separates them is whether a particular activity may be construed as a criminal offence. For example, whilst it may possible to prove a Level II proposition such as 'the suspect had intercourse with the victim', more persuasive work is required before a corresponding Level III proposition, namely 'the suspect is guilty of rape' (ie no consent), is proven[5].

So far my discussion has focused upon the CAI as it is depicted in the literature published in forensic science journals, and so has considered it as a series of empirically informed theoretical statements. I now turn to a more detailed discussion of the issues that arise when considering the practical implementation of the model. In this section I draw upon additional material from semi-structured interviews which add greater understanding of the contingencies of the application of Bayesian reasoning to support criminal investigations. Hence I attempt to show how the use of CAI in casework reveals a number of issues which challenge the representation of the CAI as an intervention derived from, and shaped by, a set of abstract principles of probabilistic reasoning.

8.8 Application of the CAI in Criminal Investigation

The use of the CAI can be seen to start right from the outset of an investigation:

> 'Say we just find a body in a field, what's gone on? How did the body get there? What happened to the body? So the scientist, whether it be a police scenes of crime type scientist or a...more independent forensic scientist, they would be giving some kind of opinion, an investigative opinion, probably a posterior probability opinion as to what went on. So they say well, you know, looking at the...tracks in the...disturbance at the scene, I think the body's been dragged from that position to that position, I then

[5] In addition to the three Levels outlined above, another Level, known as *sub-source*, has been proposed to describe the possibility of a further set of considerations that may be introduced by the existence of DNA evidence. The consideration of a *sub-source* level separates questions of match probability of DNA from the question of where it was recovered. The question of whether the DNA was recovered from blood, semen or another source may vary in importance depending on the nature of the crime, or the other circumstances surrounding the incident. The use of a sub-source level has been seen as a possible means of preventing the over-privileging of DNA profile data, against other evidence in relation to the case.

> think it's been turned over, you know, and all stuff like that. So these are posterior probability opinions.' (Interviewee 5, 2006).

It is here that forensic scientists are viewed as operating in the investigative mode as described above. At this initial stage, scientists will generally be preoccupied with attempting to establish some form of reconstructive knowledge based on observations made of the scene. As the interviewees comments demonstrate, the ability to make reconstructive inferences in such cases may well depend on one's capacity to interpret a variety of visual cues. These cues which investigators may use to help make such inferences can take all manner of forms. Another interviewee used the term, 'soft data' to describe this:

> 'You can tell straight away that she's probably been murdered...she's clearly been sick...and somebody's cleaned her up...she's probably been re-dressed...now investigators, SIOs, think a lot about that, and their hackles rise when they see people with thongs on back to front, and bras on inside out, and stuff like that. They immediately start thinking 're-dressing', then somebody else had killed and dressed her afterwards. The lab would probably...wouldn't see that and know that, so that's got to come from the investigative side or somebody with a scientific mind who's part of the investigation...that's what that is, *soft data*, 'why is the zip slightly down?', somebody's tried to re-dress her and hasn't got the top button in as he's moved her, its come down, a girl probably wouldn't go out, there's no belt there ...that's intelligence, that's...the life-blood of an investigation...' (Interviewee 8, 2008)

The ability to identify and recognise soft data may depend strongly on the experiential standpoint of the individual scientist. Experience in this sense may refer both to their memory of previous investigations of similar cases, but possibly also to a generalised life-experience more akin to what one may think of as 'commonsense'. The epistemic grounds of the identification of such soft data may therefore, be rather opaque. Identification of soft data may rely to a great extent on individually accumulated experience, or shared understandings. There also exists the possibility that soft data may be recognised through pure intuition, a hunch that 'something isn't quite right here'. Even when one is able to reason in a retrospective manner, it can be seen that soft data which may guide such reconstructive efforts may depend on the experience and standpoint of the individual scientist. For example in the following discussion Interviewee 8 recounted an incident in which a girl's body had been found in a ravine:

> 'I said we can tell the SIO something important...because the scratches on the girl's knees were round the legs...the scratches went round like that , and I said to the scientist, she's obviously been upright going down the bank, she's run down there, so that means its an attack site.'
>
> 'If she'd been dead, and transported in [a car and the body then dumped in the ravine], and she'd have been dragged down the bank with scratches they would have gone down the length of the leg.'
>
> (Interviewee 8, 2008).

Precisely what constitutes 'soft data' may be open to some interpretation, and it may well manifest itself in different ways. It may be seen however to refer to information which may not necessarily be readily quantifiable, and its existence and significance may be contingent on a series of extraneous factors which could include time, location, activity etc. Nonetheless, the comprehension of soft data appears to form an important preliminary step toward the formation of propositions. The identification of these visual cues and the inter-relationships between them at a crime scene may directly inform the generation of prior probabilities, and hence propositions. Consideration of such derived propositions allows the scientist to begin to operate in the evaluative mode. However, in the transition to this mode certain problematic issues in the application of the CAI become increasingly apparent.

A key issue at source level concerns the availability of suitable databases with which to assess the significance of particular pieces of recovered evidence. In some cases, such as DNA, these are relatively well established, and indeed are considered necessary resources upon which match probabilities may be calculated. However, for many other forms of evidence, interviewees argued that the necessary data sets able to aid scientists in assessing propositions were often lacking:

> ...there's things like assigning priors, if we are to assign priors in investigative mode which is what I'm saying we would do, so it's having sufficient data to assign priors realistically and obviously. Perhaps more, more importantly because either in investigative mode or in evaluative mode we're dealing with likelihoods which is the probability of the evidence given propositions but there's huge areas where there just ain't the data...(Interviewee 5, 2006)'

One exception is the field of the fire investigation. In this case there appear to have been efforts to establish databases which may aid investigators in making inferences regarding the cause of particular fires:

> '...they set their priors from a database of known previous fire causes so I can't quote you the figures but just for example what they'll say is that from a known database the previous fires...twenty per cent are caused by accelerants, twenty per cent are caused by electrical, twenty per cent are caused by smoking or whatever...ten per cent are caused by explosions...so they have a database...which gives them priors. So it's testable, it's exposed, it's explicit, it can be challenged, but it's explicit, people can see the priors that we are using. Now we are far, far, far from that in any other field.' (Interviewee 5, 2006).

However, by and large, full implementation of the CAI may be hampered by the absence of data which may facilitate the generation of prior probability estimates[6]. It was stated by one interviewee, that in the absence of background data, recourse could instead be given to less explicitly rationalised means of establishing likelihood estimates:

> '...so people rely on their own experiences, they maybe kick it round with their colleagues and say what do you think about the likelihood of getting this, you know, maybe not phrase it like that but that's the sense they're trying to get.' (Interviewee 5, 2006)

[6] Forensic science providers however, do keep certain databases of information from certain pieces of evidence. For example, databases of glass fragments have been developed to compare glass on the basis of refractive indices.

Hence there is great potential for likelihood estimates to be based on a plurality of epistemic groundings. In the absence of standardised data, such estimates may be based on experiential or intersubjective understandings of the significance of evidence based on previous instances. Likelihoods may then be generated in the absence of transparent calculative procedures. If likelihood estimates are grounded on more experiential bases, then a potentially considerable amount of variance may be encountered.

The issue of data availability becomes greatly complicated on progression through Level II, the activity level. For progression to occur other factors relevant to the circumstances of the case need to be taken into account. In the following example, DNA was isolated from a discarded ammunition cartridge recovered from the scene of a fatal shooting. The DNA subsequently matched that of a suspect. However, as the following discussion shows, the match alone is not sufficient grounds to assume the suspect was the killer:

> 'In, in preparation for arresting this guy there's a discussion with the scientist about the potential strength of evidence and the officer says something like you find me his DNA on that cartridge case and I'll charge him with murder. Now that may or may not have been, you know, a sound strategy but obviously what he was doing he was translating his, matching the, potentially matching DNA on the cartridge case which would be the bottom of our hierarchy, straight through to guilt and I think most lay people would say his DNA's on the cartridge case, he must have fired the gun. And that just shows again what I was saying before about you may get a magnificent likelihood ratio for the source of DNA on the cartridge case but what that is in terms of whether or not he pulled the trigger, there's a lot of intermediary stages to go through to translate the weight of evidence for DNA to weight of evidence for guilt in terms of pulling the trigger.' (Interviewee 5, 2006).

Hence a considerable amount of inferential work needs to be accomplished in order to establish a link between the presence of the DNA match and the contribution of the suspect to the murder.

At this level Bayesian analysis becomes increasingly more dependent on information pertaining to the specific framework of circumstances of the case under investigation. If such information is not readily forthcoming, then the contribution of the scientist may remain at source level. In the following casework example, investigators were pursuing the case of the murder of two people. The perpetrator had entered the house of the two deceased, and had subsequently attempted to use a towel to mop up blood at the scene:

> 'Well there may be things to be done you see because that then starts to test how expert are we at scientific knowledge and understanding of transfer and persistence of DNA. What really would we expect to find given that he handled the towel seven days or ten days ago? If we would expect exactly the same sort of profile that we'd expect given he's handled it on the day of the murder then sure, we couldn't, the likelihood ratio becomes one for that evidence. The evidence is equally likely given he did the murder or given he was there seven days ago. But if say, if say for whatever reason technically you wouldn't really expect that quantity or that form or that amount or that distribution of DNA given that it was seven days ago compared to, you know, handling it as a result of the murder... Then we could say well the findings are somewhat more likely given that he handled the towel on the day of the murder and then given he handled it seven days before, there is still some probative value, some weight of evidence from likely ratio in favour of the prosecution.. (Interviewee 5, 2006)'

The requirement to account for additional variables such as measures of transfer and persistence of material brings about a further analytical burden. The information required at this level may be highly multifaceted in nature, which in turn may impact upon the kind of datasets required. In some cases, such as the above example, it may well be feasible to model these variables[7], and in doing so add greater weight to the contribution of certain pieces of evidence. However, in other cases, the grounds on which activity level assessments are made may not be so readily quantifiable.

The appropriate use and construction of propositions was viewed as central to the aims of the CAI. However, interviewees commonly expressed the opinion that the further one ascended the hierarchy of propositions, the process of constructing propositions involved an increasing degree of complexity. One example given by Interviewee 5 involved the discovery of DNA on a bullet cartridge. Interviewee 5 discussed the issue of having to reason whether this DNA could be used to inculpate an individual suspected of firing the gun, from whom a match had been returned:

> 'In, in preparation for arresting this guy there's a discussion with the scientist about the potential strength of evidence and the officer says something like you find me his DNA on that cartridge case and I'll charge him with murder. Now that may or may not have been, you know, a sound strategy but obviously what he was doing he was translating...potentially matching DNA on the cartridge case which would be the bottom of our hierarchy, straight through to guilt and I think most lay people would say his DNA's on the cartridge case, he must have fired the gun. And that just shows again what I was saying before about you may get a magnificent likelihood ratio for the source of DNA on the cartridge case but what that is in terms of whether or not he pulled the trigger, there's a lot of intermediary stages to go through to translate the weight of evidence for DNA to weight of evidence for guilt in terms of pulling the trigger.'
>
> 'And I don't know how far the case went after that, whether they actually progressed to court on a murder charge but at some stage the murder charge was either dropped or he was found not guilty. I think the murder charge might have been dropped but he was found guilty or he did admit to a charge of handling ammunition.'
>
> 'So now it would be interesting to find out whether the case was progressed through prosecution and trial or whether it was dropped by the prosecution at some stage, I really don't know but again that's a nice example I think of the weight of evidence not being translated through the hierarchy.'

[7] For example, see the work by Lowe et al (2002) on DNA transfer and persistence which could be used to inform such work.

There's kind of a lot of intermediary stages to, to kind of get to weight of evidence of this DNA in terms of whether or not he was the man who fired the gun at this guy...

So we need to know whether or not the DNA, what's the chance the DNA on the... cartridge case came from the person who handled it. Now it may or may not have been, a good probability they say the next stage is, is he the guy who loaded that bullet into the gun. So they say what's the probability that the DNA we found on the cartridge case came from the person who handled it and loaded it into the gun and then so and then we say what's the probability that the person who handled the cartridge case, loaded it into the gun, was the same person who fired the gun...To progress through the hierarchy you'd have to say things like OK we've got DNA matching...the guy so the issue is what we would call a sub-source issue which is the source of the DNA so the issue is whether or not the DNA came from him, we've got a good likelihood ratio for that, the next issue is about whether or not, this is an activity level issue, is whether or not he handled the bullet.' (Interviewee 5 2006).

This example shows the need to introduce the concept of the hierarchy of propositions, in that it highlights the range of considerations that may need to be taken into account before a link can be established between the DNA evidence and the alleged offence. However, it also exposes the difficulties in constructing this link through the formation of relevant propositions. In what follows I discuss this issue in more detail, and demonstrate how a number of sources of ambiguity may be encountered, which are potentially intractable in nature, and may only be circumvented by distinctly social practices.

8.9 Propositional Ambiguity

The complexities involved in the application of Bayesianism to investigation manifest themselves in a number of sources of ambiguity encountered in the process of the construction of propositions. First, *intra-propositional* ambiguity concerns the potential uncertainty relating to the wording of propositions. Second, *inter-propositional* ambiguity may involve uncertainty over which is the most appropriate proposition to address in relation to an activity. It can also involve uncertainties over whether two pairs of propositions are actually mutually exhaustive. Finally, *statistical* ambiguity involves uncertainty over the nature of data used to estimate prior probabilities and likelihoods.

Intra-propositional ambiguity may commonly manifest itself via the use of vague terms in propositions. The authors of the CAI have sought to address this issue (Evett et al 2000b). They cite the example of the word 'contact' as typical of the kind of phrases which are open to interpretation. The phrase, 'Mr Smith had been in contact with broken glass' does not convey precisely what Mr Smith was doing that led him to be in contact. In this case he may have deliberately broken a window, or alternatively he may have been an innocent bystander, located close to a window when it was broken by someone else.

> 'I think in every day language, you know, there's more leeway and we probably understand what we mean...but if we're trying to evaluate evidence we've got to be very careful to specify the proposition and the alternative, quite crisply, because otherwise it's very difficult, if not impossible to assign any probability to the evidence given this woolly proposition...so, you'll see lots of examples of that throughout...' (Interviewee 5, 2006).

There is a possibility that *intra-propositional* ambiguity may arise as a consequence of lack of information at the activity level. If it is difficult to establish the precise series of events, scientists may only have recourse to more vague terms.

Inter-propositional ambiguity may exist where there is more than one proposition, or set of propositions to address when addressing a particular activity based on evidence. One example involves the trial of Barry George for the murder of TV presenter Jill Dando:

> 'The propositions...it was that 'the result we've got...is 'how likely is it that to arise if it got into his pocket at the time he fired the weapon that killed Jill Dando', and 'how likely is it to have got into his pocket...at some other time?' Now that does illustrate my difficulty with the process. One, why is that second proposition the one they used? Why isn't it 'anybody else in the street', why isn't it 'anybody else who has an interest in guns'?' (Interviewee 8, 2008)

This example demonstrates the myriad of factors that could be taken into consideration when constructing propositions. Here, most of the issues lie with the construction of alternative (generally defence-led) propositions. Prosecution propositions may simply need to account for the probability of a certain suspect committing a particular offence. However, in trying to construct an alternative account of events, any number of factors could be taken into consideration.

This kind of uncertainty can lead to questions being raised over whether an offence occurred at all. Another example that may be cited here involves a case involving the death of an elderly woman in the Netherlands. The woman was found dead in her greenhouse with two single stab wounds to the neck, seemingly administered by a pair of scissors. The woman was also found to have had a high concentration of alcohol and diazepam in her system, and had a history of substance abuse. Her husband, who also had a history of alcoholism as well as minor domestic violence, was placed under suspicion of her murder.

> '...the main reason why the police charged the husband with murder was that they couldn't see how anybody could accidentally fall and stab themselves in the neck fatally with scissors...' (Interviewee 8 2008)

Three senior forensic scientists, a psychiatrist, a pathologist and a lawyer all attempted to deliberate over this case using a CAI-led approach. Considerable difficulty was experienced in agreeing on which alternative propositions to formulate with regard to the case. However, it was felt that a possible alternative to consider was that the woman had died accidentally:

> 'the first thing you do when you look at it, is try and get some figures about how likely is it that you would be stabbed in the neck fatally, or stabbed with scissors anyway...and how likely are you to accidentally fall and stumble, and kill yourself like that...(Interviewee 8 2008)'

Initially, figures were obtained which seemed to suggest that a murder had occurred. These statistics indicated that the probability of a murder occurring using scissors was one in ten thousand, whilst the probability of an accident involving scissors was three in four million. When combined these figures returned a likelihood that indicated it was 132 times more likely to have been murder.

> 'But that's not the right population...there are only seven hundred murders per year in the UK, so that's the population available for murder by scissors, so its one in ten thousand...times 700...of all murders...Whereas, it's the entire population that's available for falls onto scissors, so if that's 3 in 4 million, its 60 million times 3 in 4 million. When you work it out, by Bayes, which I never did at the time, ok, it turns out to be instead of intuitively, that its much more likely to be murder, its actually much, much more likely to be 640 times more likely to be an accident.(Interviewee 8, 2008)'

It is possible to question these results. As can be seen, the figures for murder are obtained for the UK, not the Netherlands. Furthermore, the figures relating to accidental death were obtained from the coroner's office in Birmingham. However, the ambiguity became further compounded:

> '****** said 'no we need to know all fifty year old plus ladies in the UK, how likely they are to fall'. And at that point we thought...'but we can't get that'. Can we? Because that includes all the falls that don't injure anybody, because, by definition the ones that do injure somebody enough for them to go to hospital is a subset of that anyway. So its actually, quite difficult...' (Interviewee 8, 2008)
>
> 'And ****** was adamant that that wasn't what we wanted to know, we wanted people who don't fall when they are drunk, and we missed that point. Right, because the alternative is that how likely is she to do that when she's...fallen when drunk and he wanted to know how likely

are you to fall and not injure yourself when you're drunk. And as well as, how likely are you to just fall, and so he wants the proposition to be as close to the alternative as relevant...as possible.' (Interviewee 8, 2008)

Here, can be seen an instance of statistical ambiguity and inter-propositional ambiguity. Controversies over data are related to the kind of propositions one seeks to test. Here, there are a number of other factors that could be considered in the analysis, and these affect the kind of propositions that may be considered in turn:

'But then we realised that the population I realised this bit by myself and ****** said that's right, but then we actually tried to apply it to her individually, why was she the person who fell on the scissors? Right? As opposed to anybody else? We all found that very difficult...by then we were definitely going backwards and forwards'.

'Yes, there probably was influenced by the fact that we could see some of the propositions weren't going to get data on. And that wasn't sensible was it? Really, we should have got the propositions right and worried about the data afterwards...' (Interviewee 8, 2008)

Here, there seems to be a direct inter-dependence on the type of propositions constructed and the availability of the data. Propositions were formulated after it was considered what data was available, as opposed to agreeing on what propositions to test from the outset. Hence a certain degree of pragmatism was involved. Ironically, the case was largely solved through other means:

'But in fact, you can, if you think about it, you can think of an explanation really easily, which is, I can demonstrate, I Bayesianised it, right, but we actually managed to prove it after that. So if you think of yourself as...tottering about, on your feet, and you trip as you're going through the doorway, you might stick out your hand to hold yourself up, right...I'm going through this doorway, I trip, I've got scissors in my hand, I spread my hand to try and catch something, this is an upright of the door frame, and I fall and I turn my head because I can see the knife

> (sic) coming, and I've stabbed myself in the left side of my neck, like that, with the scissor blades open, and them being against the upright, right, so my whole weight is going to push them into my neck...so, the pathologists were thinking, 'how could you fall?' in, scissors in your right hand, and get stabbed there, with the scissors open, you can't. But, if you think of it that you trip and you open your hand...you open your hand to support yourself, that does everything that's required. It supports them against some hard object...and you turn your head, naturally, so if you look at the wound...see how wide apart they are, because she's spread her fingers, so...now that's not Bayes...now I've gone back to just expressing it in layman's logic terms' (Interviewee 8, 2008).

The important point to consider here is that whilst Bayes may have been cited as a guide to reasoning, other more practical modes of reasoning were employed to help establish precisely what happened. What is notable is the observation of the particular type of marks made by the scissors, and how this informed a practical reconstruction of events. The 'layman's logic' is redolent of the 'soft data' alluded to above; a mixture of practical action and observation is used to reconstruct the events, rather than any final recourse to quantitative data. Although the latter might provide some vague guidance to the case, in the sense of *suggesting* a possible alternative explanation, a great deal more work was involved to progress the situation.

More seriously, the concern was expressed that there was a risk that a Bayesian analysis might still be conditioned due to the circumstances of the case, and that it was the latter, rather than the Bayesian approach, which still tended to guide the course of inquiry:

> 'I'll give you an example. Let's say, let's say a crime was committed, a burglary say but there was a witness to say it was a, it was a very young woman in her teens, a young teenager, and the police attend, they find a footwear mark which they believe could be proven say it can be proven that it was the offender's footwear mark. They arrest the suspect, not because of the footwear but from other information, they arrest the suspect, take her shoes, they're submitted and the scientist compares this and, and says oh that's a good match, that's a small shoe, it's a woman, it's a female's shoe and it's really, really unusual that. So, you know, in

> the general population it would be unusual... But then you say well let's have a think about this. What is the...database? You say well what is the alternative? Well let's think of the proposition and the alternative. The proposition is this shoe made the mark and the alternative is this shoe didn't make the mark, it was made by the shoe of another young female. So you've conditioned almost, well you've conditioned your alternative and you're conditioning then the database because the database is young females who are likely to be in the, the potential perpetrators...So the value of the evidence is not what the scientist first thought, this is a very unusual shoe. Sure it's a very unusual shoe in the population of burglars but in the population conditioned on the alternative which is young females in this area at the time, it may be quite common. (Interviewee 5, 2006)'

Here, it can be seen that ambiguity exists with regard to the precise type of database that the evidence should be compared against. A number of factors (gender, location in this case) act to confuse the issue.

8.10 Responses to the CAI

It can be seen that there is a marked plurality of understandings and representations with regard to the concept of reasoning in a 'Bayesian' way. From the above, the latter can be viewed as representing: a process which involves the construction and comparison of divergent propositions; a means for updating one's beliefs in the light of new information; and a necessary and inevitable consequence of the explicit adoption of principles commonly associated with a 'scientific stance' to forensic evidence and its interpretation in the course of criminal investigations.

However, one notable facet of all of the above is a shared sense that Bayes acts as a guide for reasoning through a case. Yet even this is a position not necessarily shared by all:

> 'Statistics is a tool. Its not science' (Interviewee 3, 2008)

Here, there is a strong sense of what Bayes can, and can't be seen to achieve. In this instance the use of Bayes, and statistics in general, is seen as more limited. Rather than functioning as a guiding mechanism for forensic scientists, Bayes is viewed more as a simple means of verifying data. The main issue that lies at the heart of these criticisms is not that of the correctitude of Bayes Theorem *per se*, but more in the way it is currently being prescribed for use in forensic scientific practice:

> 'All of this is about the weight of the evidence and how its used rather than there's any real flaw in the process.....Bayes theorem was really designed for long-running experiments, Bayes used it for rolling balls along a table...now a crime scene is a single experiment, irreproducible in that sense, and the question then is, how useful are prior probabilities after the event has occurred? Now I'll give an example...what was the probability that you and I would meet in this room? If you estimated that a year ago...now tomorrow, after it has happened...can we say, well it was unlikely that they met because the prior probability was, right, so does it really matter, because what we actually use is the corroboration of, you and I saying it happened, it's [the secretary's] diary, [a colleague] seeing you, you've got a train ticket, all of this other information that says it happened...so the fact that it was unlikely in advance...' (Interviewee 3, 2008)

The quote above also highlights another perceived shortcoming of Bayesianism, namely concerns over the kind of measures that are most important in the context of evidential interpretation. Rather than the considering the degree of likelihood of a particular event occurring on the basis of evidence, it was argued that it is the degree of *corroboration* of evidence that is of most importance. It was also argued that the likelihood ratio currently promoted for use in forensic science takes no account of the degree of confidence that may be attributed to the numerator or denominator[8]:

> 'I think the danger is abuse of the method, and I that its...the likelihood ratio on the other hand of, 'which is more likely to have happened', we

[8] Certain criticisms of Bayes may lie at an even more fundamental level. For example, the Bayesian model requires one to assume that the degree of belief in evidence is more relevant than the probability of the actual event. This leads one into potential confusion over what might actually be meant by 'Bayesian probability'. If it is considered a degree of belief, rather

frequently...we simply don't know how to assess the numerator and denominator. And the other problem is, what if...I'm entirely confident in the numerator but have really no idea about the denominator? If I know that something's a million to one, and I'm pretty unsure about the other denominator being ten to one...so the likelihood ratio is hundred thousand, but that hundred thousand is the same as, a dodgy million to one and a certain ten to one...or an uncertain and uncertain, or a certain and certain. So the likelihood ratio doesn't recognise the confidence in either of those. And in reality, for evidence like paint, glass we simply don't know.' (Interviewee 3, 2008).

Hence in this instance, the epithet 'Bayes' was regarded as acting as a mere re-packaging of what was still considered, at a fundamental level, the role of human judgement in forensic casework:

'It's giving a scientific coating to what basically is a human judgement about the belief in something...now of course that's what Bayes was trying to get around, the belief thing, but put it on an objective basis, on measured results...this is not what's happening...we're having it now transformed into this degree of belief, and it doesn't matter how much you believe something beforehand...if the hundred to one outsider comes in before the two to one favourite, you'd say, nah it couldn't have happened. How useful is it knowing that fact if somebody says 'the hundred to one outsider won', how useful is that information in assessing after it's happened? Well intuitively I'd say no...if however, you did a long-running series of measured experiments...we don't have a measure here' (Interviewee 3, 2008)[9]

Here, once again, the question concerning the basis of judgements made in the context of using Bayesian reasoning is called into question. In turn this raises a related issue, namely the extent to which Bayesian reasoning may be considered as a skill, and hence one that may be certain scientists may find more or less easier to adopt then others. Although from consultations and discussions it seems that certain individuals have accepted the approach

than a frequentist definition of probability as one that is readily observable and calculable, then there may be risks in it usage.

[9] Hence a hypothesis with low *a priori* probability is suddenly converted into one with a high *post hoc* degree of belief (Interviewee 3, unpublished). Given that instances of single events are being considered, this was viewed as potentially dangerous.

and found it to be useful, the skill of using Bayes, and reasoning in a 'Bayesian' manner, seems not to have been acquired in a particularly intuitive fashion:

> 'I mean it might take a while, you know, because you need to revisit this stuff time and time again but it does, you know, we kind of liken it to a slippery fish, you know, we grab the fish and you think you've got it and then [snaps fingers] it's gone out of your hand...' (Interviewee 5, 2006).

Thus, while committed Bayesians emphasise the supposed logicality of the approach, it is far from clear that this is a form of reasoning which 'comes naturally' to forensic scientists. Accordingly, it is uncertain whether when used, Bayesian reasoning has been applied in a consistent manner. It is apparent that consultations with operational forensic scientists had exposed some of the difficulties in making probability judgements, and in the absence of tangible data, it is notable that forensic scientists appear to continue to largely rely on personal experience and tacit knowledge. Those relatively new to the profession found themselves taking this experience on board themselves so that the ability to draw conclusions was developed by having:

> 'learnt at the feet of someone else, y'know you kind of look at your more experienced colleagues and see what they did, and they pass on their wisdom, sort of thing, and...I suppose that wasn't very satisfactory really, it wasn't very 'scientific'. (Interviewee 5, 2007)
>
> And I think in the end, in the end, what people do, how people make decisions, can be very intuitive, and most people's intuition is quite good...I think many people see what appears to work well, but on some occasions your intuition can really let you down, and that's what we were trying to expose, get people to expose more, how they are arriving at their decisions in a case.' (Interviewee 5, 2007)

The reference to the 'apparently tried and tested way of doing things' seems to indicate a tendency for forensic scientists to revert to the use of accumulated personal or collective experience in casework rather than the Bayesian framework. For example, a visit to the laboratory of a large metropolitan police force indicated a wholesale lack of regard for Bayes, with identifications and matches being called on the basis of individual judgement.

There may be a number of reasons for this relative lack of uptake. Issues regarding the availability, or lack of, adequate data for establishing prior probability, was cited as a common reason. Other reasons seemed to centre around personal preference. In this latter case it is unclear whether preferences were influenced by the disciplinary background of the scientist, or whether it was a matter of individuals being comfortable with the method as a whole:

> '...in the early days it was quite difficult to turn some of those people round and think in a different way but having said that...the handwriting people in the FSS have been particular, most of them warm to it very much, there were one or two key people who said 'we can't do this, this doesn't work'...and really set the stall out against it...so, again...the experience was a bit mixed, some people warmed to it, and took it up, and some took it forward, other people said 'no way' dug their heels in...and I suppose in the end it doesn't depend on which discipline you're coming from, it depends on your own personal style and...equally some people [in] DNA areas, even though they've been trained to think in likelihood ratios for DNA evidence, if they try to apply it anywhere else they couldn't see it...' (Interviewee 5, 2007)

In the examples of both handwriting and DNA, the respective experiences both seem rather mixed. In the case of handwriting, a sub-discipline which does not lend itself easily to quantitative methods, it is interesting to note that the majority of analysts were open to the methodology of the CAI, although it is perhaps equally important to recognise the existence of some opposition within this area. Furthermore, the experience of DNA analysts is particularly striking. The difficulties experienced in applying Bayes to anything other than the relatively narrow realm of the evaluation of DNA evidence point to a potentially more complex series of factors which may both influence the desire to use Bayes, and the proficiency with which it is applied.

What does seem relatively clear is a certain reluctance on the part of some forensic scientists to open up their methods to the kind of epistemological scrutiny which Bayes requires. Here, the sense of the authority of conventional practice being challenged did not necessarily appear to be welcomed, and it is evident that some analysts may have been very uncomfortable with this:

> 'Another area for objection might be because we were challenging conventional wisdom, and we were almost kinda challenging the basis of their expertise. Or appeared to be, challenging the basis of their expertise, so I think that, you know, once a person has that in their mind then...its very difficult to get past that...I'm thinking...in other areas as well, not just handwriting, I know certainly in the chemical areas...there were certainly a few people who had the same reaction as that handwriting person, and it was almost like a challenging their, maybe even authority in some ways...I know it was very difficult sometimes...when we started to talk about propositions and alternatives and said well 'what's the probability we get this amount of gunshot residue on his hands 4 hours after firing', and what's the probability we get this gunshot residue on his hands given he hasn't, he's a member of the public, you find people very reluctant to give you, not a definitive answer, but a good answer, a good answer that you and I could understand...they waffle on or, well they would kinda say 'well I just know this', y'know, so I suppose theres a, a person has to be willing to open up, and kind of almost admit that their science, their expertise is on shaky foundations.' (Interviewee 5, 2007)

The concerns raised above relate to the role of Bayes specifically in a forensic scientific context. However, this context is quite clearly part of an even wider domain, encompassing the contexts of policing, the judiciary, and increasingly commercial interests. As I seek to argue below, the application of Bayesian reasoning in forensic science is subject to a wider series of negotiations relating to this wider domain, another issue which has largely been overlooked by those who have previously commentated on the subject of Bayes and forensic science.

As discussed in the initial section of this chapter, the employment of the CAI could reinforce the strategy of positioning forensic scientists more closely to the centre of the investigative decision-making process. Indeed, one notable aspect of concern for the developers of the CAI was the awareness that a significant proportion of the burden of decision-making remained in the hands of actors other than forensic scientists, at the expense of scientific propriety:

> '…if we haven't got the data about transfer and persistence, has the blood come from the window or has the blood come from that man. Have these firearms discharge residues come from that gun?...the big problem is that, if we, if we haven't got the information, the data, then and I would say scientific information, scientific data, for transfer and persistence and detection, primarily for the numerator [prosecution], and then background levels, for the denominator [defence], we'd have to stick at source, now if we stick at source, it's simply about mainly within-sample variability and between-sample variability, and if we stay there, then someone has got to progress it to activity level, to get to the offence, and…the people who are gonna do that are gonna be people like lawyers, judges, and jurors, lay people. Who aren't as well placed as a scientist should be to get to activity…so there's that disjunction really, when…there's a lack of scientific data, to move from source to activity. But the big temptation is, for other people to do it then, on behalf of the scientist.' (Interviewee 5, 2007).

There have been some suggestions that the wider deployment of the CAI could elicit a change in the nature of the relationship between forensic scientists and the police, in terms of the manner in which the decision-making process is organised. With regard to the roles and responsibilities of forensic scientists, the CAI could possibly be seen to have attempted to remove some of the pressure on forensic scientists to deliver authoritative and categoric solutions to casework-related problems:

> 'I think as well, it gets into another area I'm fascinated by, and its this bit about the, image of forensic science, the CSI effect, and all that sort of stuff…its almost, I bet theres scientists who almost feel they've got to be authoritative, because they feel that's their role, so I've got to be

categoric, I've got to give firm opinions...and that kind of forces you down a route you may not want to go but you feel you have to because, I'm the forensic scientist, everyone's expecting the answer...everybody loves categoric answers, so you think I'll give 'em what they want, so I think there's an element of that... the kind of history has been authoritative forensic science, and we feel we have to live up to that.' (Interviewee 5, 2007)

One possible consequence of more use of the CAI would be to re-position forensic scientists as acting in partnership with police investigators, and it could be seen as re-positioning the two sets of actors in epistemic terms. In order to make source-level data relevant and useful to the outcome of the investigative process, scientists would be reliant on information garnered from police inquiry in order to gain an understanding of the framework of circumstances of the case. The CAI has also aimed to give both partners a say in which pieces of evidence would be taken forward for analysis:

'What did happen, or what still does, is that police would say, there's a break-in, there's been a glass window broken, look for glass on the clothing, or, you know, examine this for DNA, you know, and so, they kind of directed the scientists to apply certain techniques, whereas what we really try and encourage the police officers to say is, 'what's your problem, what are you trying to establish?' in terms of what are the issues...is the issue, whether or not he broke the window, or is the issue whether or not he went inside the property or is the issue whether or not he handled stolen goods. You know, what is the issue for you, then, I'll go away and have a think about it and then come back with a strategy for you to see if, you know, what do you think of that? A costed time-strategy, and that's...that's the whole essence of CAI, really in terms of the case assessment bit, the 'A' bit...establish good strong propositions at activity level.' (Interviewee 5, 2007)

Thus in the framework of the CAI, the scientists act in a more facilitative role, helping police to clarify the important questions to ask in the course of a case, in return for the supply of circumstantial information from the police. However, it is not clear if this new set of roles and responsibilities is in keeping with how the police view the position of forensic scientists in the hierarchy of criminal investigation. It may well be the case that police investigators may view the role of forensic scientists as somewhat more limited: to provide scientific data about evidence, which may occur in relative isolation to the deliberations made by police concerning the progress of a criminal case. It is difficult to ascertain how easy it may be to change the attitude expressed at the start of the above quote, concerning the tendency of the police to provide direction to scientists with regard to the pieces of evidence regarded as important to the case.

There may also be issues with the imperative for the forensic scientist to consider defence-type propositions. Here, the concern over the potential to force defendants to account for evidence is turned on its head. The need to take defence propositions into consideration runs counter to the fact that more often than not, forensic scientists are generally employed by police forces keen to prosecute suspects. Perhaps for this reason, there appears to have been suspicion on the part of the police with regard to the need to consider defence propositions within the framework of the CAI:

> 'We had another group of officers who kind of mis-interpreted really, one or two things we said...we had to do a bit of work re-assuring them that you know, we weren't out to...how shall I put it? Help the defendant...' (Interviewee 5, 2007)

Here the main objection by these police officers appears to have centred around the possibility that, in conducting a rigorous enquiry into the origins of particular pieces of evidence along Bayesian lines, there lay a risk of 'giving explanations to the defence' (Interviewee 5, 2007). Although the respondent above claims that police 'mis-interpreted' what the originators of the CAI were undertaking, the concerns of the former are perhaps not entirely unfounded. As mentioned above, for evidence to have any use in a case beyond Level I propositions, a wider framework of circumstances needs to be taken into consideration; however for the 'activity' level to be fully considered, it may be necessary for the suspect to be consulted so that a defence proposition be entertained. It is not entirely clear what procedures would be used to elicit a suitable defence-level proposition in such an instance, although interviewers could possibly be forgiven for exercising caution.

The use of the CAI to construct defence propositions has possible consequences slightly removed from concerns over the relationship with police. Another criticism levied at the use of Bayes concerns the onus that is placed on the defence to account for evidence:

> 'The prosecution argument must be proved, but the defence position may simply be that any or all other hypotheses that are not the prosecution hypothesis is true.' (Unpublished article written by Interviewee 3).

With regard to the current system, the defence is under no obligation to provide an alternative account to that of the prosecution. Instead of giving the defendant opportunities to explain away the evidence, a Bayesian schema may force them into fabricating accounts which may prejudice juror perceptions. There lies a risk that the defendant may be forced into concocting a series of potentially contradictory accounts depending on the circumstances of the investigation or the progress of the case through the courts[10].

This latter concern brings the issue of impartiality into relief. Regardless of whether police forces employ forensic scientists on an 'in-house' basis, or contract services to external bodies such as the FSS, it is apparent that police and prosecutorial bodies enjoy a significant resource advantage in terms of the forensic scientific support available to them. Law enforcement agencies constitute the overwhelming majority of the business of the FSS. Hence forensic science, in the UK at least, can be seen to be somewhat more closely aligned with the prosecution domain as opposed to the defence. According to one interviewee who was involved with the defence case in the

[10] The authors of the CAI do address the issue of when defendants give 'no comment' (Jackson et al 2006).

Omagh bomb trial, this apparent close association was seen as prejudicial to the outcome of cases, a situation only able to be ameliorated when the defence had access to sufficient resources, the latter being viewed as an instrumental factor:

> 'In England and Wales, I've not yet got an answer to the bias that's introduced when your entire business is dependent upon the police for funding. If they don't get the results, y'know, work it out! However, while we've got an adversarial system, where the other side are properly funded, as in Omagh for example, then we can see an equality of arms, and that's what I've been most exercised about is equality of arms.' (Interviewee 3, 2008).

Furthermore, this interviewee viewed the market for defence science as 'poor', and that defence testimony on forensic scientific issues tended to be provided by individuals who had experienced some form of dissociation from major organisations such as FSS, such as retirement. This may risk a lack of credibility in the eyes of the jury.

Thus impartiality was seen as a major issue, and it may have been the case that forensic scientists, particularly those working directly for the police, may have experienced some discomfort with such comments as the following:

> 'The ****** police own and operate the labs. I was head of the police laboratory in *******, that's why I decided that it really should sit separately from the police. For everyone's benefit it should sit separately from the police. I know the police don't believe that...unfortunately the police even made it worse...the detective chief superintendent, when asked about impartiality, said 'what impartiality? Look at your salary slip'. That didn't go down well with the scientists. (Interviewee 3, 2008)'

Hence police suspicions may have functioned as a notable obstacle preventing the CAI from being taken up by the forensic science community. For the CAI to function effectively, and for evidence to exert optimal impact on the course of a case if considered within this 'Bayesian' approach, forensic scientists need potentially high levels of information from the police. However, the CAI also emphasises the desire for scientists to enjoy a greater degree of influence with regard to which pieces of evidence are seen as most pertinent to a case. This potential re-positioning of forensic scientists within the decision-making hierarchy of criminal investigative frameworks may not necessarily be in keeping with traditional police views on the place of forensic scientists in criminal investigation.

With regard to consultations with senior figures in the FSS, another set of issues presented themselves, related to commercial pressures:

> '...whilst the leaders and managers of the FSS could see some benefit...in the approach we were promulgating...I think they felt some commercial problems as well...certainly at the time...the way we were working in terms of earning our incomes was items examined, the more items we examined the more income we got. So there was almost a counter-pressure...not to apply CAI, because CAI in some ways, said 'lets just look at the items that are gonna be really effective, really efficient, in addressing this question, and particularly upfront, if you decide with the customer these are the key issues in the case, the strategy to address these key issues is this, this and this, it's not these items or these tests, so there was almost a counter-pressure not to apply it. And I think therein lay some of the difficulties from the managers and leaders, because you could see the natural consequences, if we apply CAI...we're gonna lose a lot of income, potentially. And there was, to be honest, that was probably a real risk...there had to be a real change in the commercial basis of charging... so instead of charging per item, we charged per service.' (Interviewee 5, 2007)

This the manner in which the FSS organise the way in which forensic science is conducted within the organisation clearly has some bearing with regard to the uptake of Bayesian methods. Interviewee 3 cited a change in the way in which forensic scientific practice was organised by the FSS:

> 'Particularly in the FSS now, they've gone away from experienced case scientists looking at multiple evidence types and synthesising...all of the

> evidence together to...neat little boxes y'know of 'products' where 'you want the hair analysed? We'll do that' does the hair analyst ever speak to the DNA person, speak to the paint person, speak to the glass person, that aspect has probably been reducing, and that's a consequence of the more commercial approach.' (Interviewee 5, 2007).

Here we see an increased concern with commercialism leading to an organisational approach which attempts to render forensic scientific 'products' as neatly packaged and clearly differentiated techniques and specialisms:

> 'Most scientists who've been involved in forensic science for any length of time know the pros and cons of each, and if you like you know the limits of what you know...knowing what you know and knowing what you don't know is vital and so its that but...to compartmentalise the evidence in the way that seems to be happening, I think is compromising the scientific evaluation of the whole case.' (Interviewee 3, 2008).

It must be emphasised that senior FSS scientists will often assist Senior Investigating Officers (SIOs), in a 'special advisor' role in a serious criminal investigation, providing guidance on how the course of forensic investigation should proceed, and possibly also liasing with FSS scientists. Nonetheless, whilst there was some indication that the CAI may have been used by such 'special advisors' (Interviewee 7, 2006), it was not clear as to the extent it had been used successfully in these instances. If the results of the study are considered, it is far from certain that one individual, acting as a sole special advisor, would be able to successfully address the ambiguities experienced in the examples given herein. Furthermore, the point made by Interviewee 3, about the compartmentalisation of specialisms, is still relevant if one considers the investigation of less serious volume crime, for which the police are unable to devote the same amount of resources (Interviewee 7, 2006). Here there is far less likely to be a co-ordinating forensic scientist involved, forensic investigation may proceed in a more piecemeal fashion, in a manner antithetical to the demands of the CAI.

8.11 Conclusion

This case study has focused upon attempts to formulate a theoretical intervention in what may be regarded as a highly practical subject. At the heart of this intervention, as in the previous case study, lies Bayesian reasoning, regarded as 'the only logical way to reason about evidence' by its developers. This case study has sought to explore just how this logic is enacted in the course of criminal investigations. It has also sought to engage with the critics of the CAI to gain a broader idea of the perceived challenges in applying such a mode of reasoning to the area of forensic science. The challenges in this case are somewhat different to those experienced in the previous case study, and in this chapter I have sought to highlight this issue.

In the previous chapter, practices of black-boxing were able to identified in attempting to construct and use Bayesian technologies. The CAI is no less a Bayesian technology, but the practices identified in this case are of a different character to those encountered in the case of Automated DNA Interpretation Systems. The CAI represents an ambitious attempt to produce a framework which not only encompasses evidence such as DNA, but extends to encompass other forms of evidence, and serves to amount to a technology of reasoning for use in criminal investigation. It is therefore a significant extension of the scope of the Bayesian project in forensic science. Whilst it shares the same fundamental formula found in LoComatioN and PLS, the CAI can be seen to project a very different form of Bayes.

As before though, the CAI is not immune to having to contend with dealing with sources of ambiguity and uncertainty. Some of the most notable in this case involve the construction of propositions, an activity which itself may be considered one of the key aspects of what it means to reason in a Bayesian fashion. In the course of this case study however, I have attempted to show that the construction of propositions is complicated by certain difficulties. These may relate to problems concerning language itself, namely constructing phrases which exhibit a suitable degree of precision. As the CAI literature itself points out, it may be difficult to avoid constructing propositions which contain within them ambiguity. A further problem may concern the matching of suitable mutually exclusive propositions in order to explain an activity. Here, the problem becomes further complicated; not only is there the problem of trying to accurately correspond with an activity of which knowledge may be limited, but there is also the problem of attempting to producing a double explanation, both of which need to logically reflect the activity, but which also need to exhibit logical symmetry with one another. In order to account for the activity, the CAI obligates scientists to base their propositions on data wherever possible. Yet in many cases, the most suitable data is not in existence. This is perhaps a key problem with the CAI. Whilst certain cases may resemble others, it is likely that the precise circumstances of each case will vary so much to mean that any data will represent a mere best fit.

Whilst it could be thought that these difficulties are anterior to Bayes Theorem itself, overcoming them is necessary for any version of Bayes to come into being; if they aren't overcome, the CAI cannot function. Thus, the issues can be easily framed as pragmatic in nature; language must correspond with object, otherwise the CAI does not function, for language is the only resource. In order to overcome these challenges then it is necessary for scientists to generate distinctly inter-subjective understandings, not only of the propositions they wish to construct, but also of the circumstances of each case to enable them to proceed. Decisions over what constitute appropriate propositions is then very much a localised phenomenon. Whilst the CAI literature may provide some guidance in this regard, the high level of ambiguity serves to block any direct translation.

A related but no less interesting phenomenon was also observed in relation to the problem of fitting statistical data to the construction of propositions. Due to the lack of datasets, it appears that investigators using the CAI often constructed propositions in relation to the data that was available to them, rather than progressing from proposition to dataset in the manner intimated by the literature. Often, it appeared that the process of constructing propositions to data was in fact an iterative, oscillating process. This is in itself possibly reflects the localised nature of this form of Bayesian reasoning, in that the investigators had to use more practical forms of reasoning in order to overcome these environmental limitations.

Aside from the problem of formulating propositions, it is clear that experiential judgement contributes to the execution of the CAI. Although the CAI literature warns against the use of personal intuition, it appears that some for experience is still an ineradicable component of crime investigation. This is most clearly expressed via the concept of 'soft data'. Here, the comprehension of such soft data appeared to constitute a vital initial stimulus in comprehending a crime. Soft data cannot be quantified, and it is unclear to what extent it can be separated from the experience of the individual perceiver; it is a distinctly embodied phenomenon. Another clear example of the contribution of embodied in the reasoning process involves the physical re-enactment of incidents, as described in the case of the murder with scissors; once again it is hard to determine how such knowledge, vital as it is, can be quantified. Yet it can be seen that such knowledge was instrumental in providing an accounting for that particular incident.

Some of the oppositions to the use of CAI can be seen to come from other forces involved in forensic investigation. This chapter has also shown how the authors of the CAI have faced opposition from police officers and members of the FSS, the same organisation for which the CAI was developed. Here, extraneous interests can be seen to impinge upon the implementation of the framework. These examples emphasise the fact that forensic investigation proceeds in a heterogeneous network of actors and institutions. Attempts to perform the CAI in such an environment may be hampered by competing interests which are also embedded in this network. In this way, Bayesian reasoning can also be shaped by external social forces.

The CAI can be seen to involve a heterogeneous array of reasoning practices. Due to the practical limitations caused in turn by the inability to satisfy the informational demands of the Bayesian framework, these practices are vital for the perpetuation of this particular form of Bayesian reasoning. These findings therefore point to the possible diversity of practices involved in the enactment of Bayesian reasoning. In the next chapter I consider this issue further.

CHAPTER NINE - DISCUSSION

9.1 Introduction

In this chapter I examine further the issues raised by the preceding case studies. I attempt to show how both examples - of Automated DNA Interpretation Systems and the CAI - highlight a certain division between efforts to posit unified and logical principles of evidential interpretation, and how interpretation happens in the context of investigative casework. I consider how the transition from representation to intervention is bridged by a range of practices which in themselves exhibit an autonomous existence independent of theoretical representations, even though these practices facilitate attempts to consciously apply the self same principles. In such a way, these practices can be seen to effectively constitute a process of representation themselves. I discuss these phenomena in relation to approaches that have been identified as forming part of the 'semiotic turn' in science studies (Lenoir 1994), in particular the work of Callon (1998, 2006) who has advanced the notion of performativity. I critically examine the way in which the appropriation of Bayes in the context of forensic investigation may be considered as a performative phenomenon.

Although the performativity thesis may serve to act as a useful preliminary framework to investigate the way in which Bayes is constructed in such a context, I argue that it is unable to account for every aspect of these processes. I emphasise the merits of moving away from the symmetric treatment of human and nonhuman agency, which characterises the semiotic approach to science studies, in favour of a renewed emphasis on human practice. I show how the two case studies display a diversity of practices that are necessary for the construction of representations of Bayesian reasoning, and how these representations (which I refer to as 'forms of Bayes') are dependent on distinctly social practices to generate information in a 'Bayesian' manner.

I discuss the findings of my research in relation to the concept of reasoning by abduction, which is a source of great interest amongst Bayesian scholars in forensic science. Developed initially by the pragmatist philosopher Charles Sanders Peirce, abduction was advanced as a means of explaining how scientific discovery proceeded via the development and testing of hypotheses, but parallels have also been drawn between abduction and the mode of inquiry employed in criminal detection. I discuss the extent to which forms of abduction can be identified to be occurring in the examples studied in Chapters 7 and 8, and how these problematise the notion of abduction as a readily definable concept. I argue that the case studies indicate that, at least in the context of forensic investigation, the process of hypothesis formation is considerably more practice-led than the view portrayed by philosophers of science. I use this discussion to explore the possibility that the process of hypothesis formation is socially constructed. I argue that these cases demonstrate a form of performativity in which the modes of hypothesis generation and assessment become conflated. It is this conflation which is a notable aspect in the construction of forms of Bayes.

I continue by briefly discussing the consequences for previous sociological studies of expertise, and, finally, I show how the findings of this research hold consequences for evidence scholarship. I argue that attempts to render systematic frameworks for the interpretation of evidence may lead to impoverished accounts of how evidence is interpreted in practice, and that, in order to ameliorate this, closer attention should be paid to the manner in which Bayes is actually used, as opposed to merely advancing theoretical models for evidential interpretation. By advancing an understanding of the manner in which actors bridge the divide between representation and intervention. I hope to demonstrate how a suitable sociological approach may contribute to the field of evidence studies.

9.2 Representation and Intervention

As I have demonstrated in a previous chapter, a major theme prevalent in the science studies literature has been the challenging of accounts of scientific knowledge production which privilege the role of theory (Lenoir 1994). Contending approaches have instead sought to provide accounts more attuned with the primacy of practice in the production of scientific knowledge. Sociological studies of science have been instrumental in challenging the assumption, previously often promulgated in the philosophy of science, that the derivation and formation of theory is given precedence over the role of testing by experiment. These have been inspired by works such as Ian Hacking's *Representing and Intervening* which argue that, far from acting as mere testing grounds for scientific theories, practices of experimentation can yield knowledge of the world in a manner that is not necessarily dependent upon pre-existing theories, nor through any recourse to a singular 'scientific method'. Hacking demonstrates how theory and experiment 'have different relationships in different sciences at different stages of development' (Hacking 1983, p.xii), and argues that experimentation therefore 'has a life of its own' (Hacking 1983, p.150).

Hacking's insights, in particular the distinction between representation and intervention, act as a useful starting point upon which to consider issues relating to the relationship between the theoretical principles developed in the fields of probability and evidence scholarship on one hand, and the practice of forensic investigation on the other. Theoretical projects can be regarded as sets of practices related to *representation*. These may include practices of *formalisation*, *systemisation*, *idealisation*, *depiction* etc. On the other hand, forensic investigation may be regarded as a set of practices related to *action*, or *intervention*. Such practices may include the *identification*, *isolation*, *interpretation*, and the *construction* of evidence.

Translation between these two sets of practices occurs via the presence of relevant actors. In the forensic scientific context, this process of translation first involves the construction of formalised reasoning systems which are more closely tailored to the demands of a particular context or situation. These systems aim to draw upon and apply the principles of evidential interpretation; and can be considered as *secondary representative* practices, or partial translation. The formulae/algorithms which form the foundation of the automated systems for the interpretation of complex DNA profiles, and the Case Assessment and Interpretation (CAI) model, can be viewed as a process of partial translation, in that they are the result of the application of some of the principles outlined in the literature about evidence and probability (in this case those pertaining to Bayes Theorem), to a more specific set of circumstances.

The production of such technologies of secondary representation themselves require a human intervention, and it is here that the divide between representation and intervention becomes at once apparent, but at the same time also becomes susceptible to conflation. Practical interventions of evidential interpretation, which self-consciously draw upon theoretical principles, may actually proceed in a manner underdetermined by the latter. Mathematical approaches *qua* inscriptions, may, in themselves, provide little or no guidance in terms of how to proceed with their execution. As the work of Livingstone (1986) has indicated, mathematical work involves a potentially more diverse series of processes than the purely mentalistic ratiocination which may be implied by certain forms of mathematical realism. Any perceived ontological gap therefore necessitates a conflation of practice and representation.

Hence the process of translation may reveal alternative strategies to those outlined in the conventions stipulated in the representative schema. Different forms of reasoning may be involved the process of intervention. These principles may be more informal and not as easily identifiable but no less necessary to the enactment of Bayesian reasoning by investigators. What the work outlined here indicates is that representations of theory, and practices of theorising, can not possibly be regarded as purely distinct. In what follows I draw upon examples from Chapters 7 and 8 to show how this may be the case, and explore in more detail the idea of differing enactments of forms of Bayesian reasoning.

9.3 Forms of Bayes: Automated Interpretation Systems

One recurring theme encountered throughout the research as a whole was the repeated reference, in both case studies, to 'Bayesian reasoning', or reasoning, 'using a Bayesian framework'. On the basis of the literature and the testimony of some of the respondents, certain ideas concerning 'Bayesian reasoning' in the context of forensic science seemed to have gained a degree of acceptance. These seemed to centre around a set of principles which were cited as constituting reasoning in the Bayesian mode. These included: the need to formulate explicit probabilistic estimates of belief concerning specific events, the requirement to think in terms of formulating multiple alternative propositions (often involving, but not exclusively, propositions for prosecution and defence arguments), the need to take into account as much relevant information as possible when making calculations, in turn leading to the desideratum to establish comprehensive datasets upon which prior probability estimates can be based, and a sceptical approach to judgements based on personal intuition and/or subjective experience. These kinds of principles were repeatedly cited as being indissolubly linked with Bayes Theorem. They were viewed as directly crucial elements in what it meant to reason in a 'Bayesian' manner, notwithstanding the conscious use of Bayes Theorem to calculate likelihood ratios. Furthermore, amongst its proponents, Bayesianism was seen as occupying a certain logical higher ground, particularly when compared to other probabilistic positions, most notably frequentism.

Critics of Bayes saw the situation in a different way. Rather than being seen as a whole philosophy, critics regarded Bayes as a mere tool. According to this view, statistical and probabilistic methods were perceived as aids for helping to assess scientific inferences, and were not to be regarded as wholesale blueprints for reasoning. Other criticisms included the high informational demands that Bayes placed on inquirers, to the extent that any attempt to apply Bayes in any kind of suitably rigorous manner was virtually impossible. More seriously still, the use of Bayes in forensic science was seen in some quarters as providing a dangerously superficial façade of scientificity, in which the use of likelihood ratios merely provided a quantitative gloss for judgements which concealed a high degree of subjectivity.

Certainly as a mere mathematical form, Bayes Theorem seems to invite a significant degree of interpretation, and the original rationale behind Bayes Theorem appears to have been open to interpretation since it was first published in 1763. In fact, as I have attempted to indicate in Chapter 4, the construction of notions of Bayesian reasoning appears to owe a certain amount to other ostensibly competing probabilistic approaches, most notably frequentism. If this can be taken to be the case, then it is appropriate to begin to consider how forms of 'Bayesian reasoning' constitute social constructions. Instead of becoming entities which have been derived through an inevitable process of logical proof, modes of probability can begin to be regarded as dependent on social practices for their continued generation. Moreover from the case study chapters it is possible to identify differing *forms of Bayes*, characterised in each case by the practices which constitute them. The two case studies investigated in this study can be seen to constitute two distinctive forms of Bayes. These practices are varied, and can be seen to act anterior to the idealised series of procedures and representations associated with Bayes and the technologies, but are nonetheless necessary for a particular interpretation of Bayes to come into being. In what follows I provide an outline of a selection of these practices, with reference to each case study.

The first case study demonstrates that the interpretation of DNA profiles requires a series of negotiations. The recovery and isolation of DNA samples from biological material represents only the initial stages of the process. As I have sought to demonstrate in this case study, a series of *transformations* are required to convert DNA samples into information comprehendable to jurors and lawyers. These transformations are not only material in nature, as a range of calculative processes are also required to construct data. However, the inscripted formulae, representing the algorithms which drive technologies such as LoComatioN and PLS do not alone account for the resolution of conclusions made about these profiles.

A cursory examination may leave these technologies rendered as heavily 'black-boxed' forms. Yet these systems there are the product of certain constructions which reflect contestable ontological assumptions. One example involves the use of PLS for the resolution of mixed DNA profiles. This technology relies on a number of assumptions regarding relationships between loci, and, more seriously, untested assumptions regarding the relationship between peak shape and area and quantity of DNA. These assumptions have been questioned within the forensic scientific community itself, yet it does not appear to be the case that the FSS have rigorously tested these assumptions, for they do not, as a matter of policy, quantify their DNA samples, the key means of doing so. Yet these same assumptions form an instrumental part of the foundations of PLS. Similar concerns can be raised for the seemingly tautologous use of the M_x parameter, and the assumption that M_x remains invariant across loci,

This has further consequences when we consider the example of LoComatioN, and the incorporation of variables representing 'drop in', 'drop out' and so forth. These appear to have been no less evident in the visual form of interpretation, in which they acted as incorrigibles to justify a match. In the case of a Bayesian assessment, they are required to be quantified. However, the precise means by which such parameters appear to be estimated has not yet been agreed upon, and the results of Chapter 7 suggest that not all of the assumptions associated with 'drop in' and 'drop out' have been taken into full account. For example, the assumption is held in the systems used to model drop in and drop out that these parameters are constant across all loci. The means by which they are modelled however, do not appear to reflect the high degree of environmental challenge which could affect the quality of profile recovered. The model for 'drop in' models contamination found in laboratory conditions, which are the most far removed from the kind of contaminating conditions potentially found at crime scenes. The model for 'drop out, meanwhile, assumes equal chance of drop out when this has not been assessed with crime scene samples, and, furthermore, ignores the possibility of preferential degradation of alleles at large-MW loci. The estimation of the effect of environmental challenge is a difficult task, and it is unlikely that definitive results could be returned concerning the effects of specific forms of challenge, but it is still notable that little consideration is even alluded to in the technical literature.

The latter examples show that the notions of 'drop in' and 'drop out' used in the LCN literature are constructed concepts, created in order to bypass the intractability of fully accounting for actual 'real world' effects. These constructed notions are crucial for the construction of a Bayesian technology such as LoComatioN. With the presence of the 'drop in' and 'drop out' constructs, LoComatioN itself represents another form of Bayes. It is built upon a certain set of questionable assumptions that have been made largely to circumvent an awkward gap between theory and world.

Meanwhile, the existence of a technology such as DNABoost shows the potential for a technology to appeal to a wider set of intersubjective understandings, brought about partly by the structural changes to forensic science provision in the UK. The example of DNABoost also shows how the success of a technology, brought into an environment subject to a certain degree of marketisation, may require a set of novel practices. In this case, no scientific literature on the system was released upon which other scientists could scrutinise; no equations or formulae on which the technology was based were made available. Instead, media channels were relied upon in order to reach a wider audience of police, politicians and publics. For it to have any chance of acceptance, a new network needed to be constructed. What is ironic in this case however, is that to the majority at least, DNABoost existed in name only. This is perhaps a more questionable form of Bayes, for it is impossible, at time of writing, to confirm the algorithmic basis of this technology. If it is, as suspected, a Bayesian technology however, then the means by which audiences are convinced, itself represents a new form of Bayes, in that radically different techniques have been employed in order to convince audiences of its scientific worth, and hence epistemological validity.

9.4 Forms of Bayes: The CAI

With regard to the CAI, recourse to so-called 'Bayesian reasoning' appeared very often throughout the relevant literature and was mentioned a number of times during interviews, even though it was not always clear precisely what it was being used to refer to. Regardless of the possible semantic fluidity of such a concept however, it was clear that actors took the possibility of such a form of reasoning very seriously indeed. However, the fact that such difficulty exists with regard to tracing a clear link between the concept of 'Bayesian reasoning', and Bayes Theorem itself, acts as an indicator of the level of active performance involved in constructing the former notion.

The CAI study highlighted a rich multiplicity of forms of Bayes. As discussed above, so-called soft data was cited as a playing an important part in the criminal investigation process, and it appears that it played an important role in forming propositions. Soft data however can be regarded as highly dependent on individual experience. Whilst Bayesian reasoning is intended to be able to account for subjective beliefs, the fact that soft data may be virtually unquantifiable presents an awkward case. Furthermore, the kind of soft data possessed by each individual may vary considerably, and this does not appear to be accounted for in Bayesian calculations. Nonetheless, the use of soft data, at least as an initial means of comprehending situations and informing the preliminary reasoning process, appears to be instrumental in allowing the CAI to function.

It was also apparent that the precise nature of the propositions constructed in the course of applying the CAI were often dependent in part on the availability of data against which prosecution and defence hypotheses could be compared. The data that were used did not always fit perfectly with the circumstances of the case under consideration, and could often only provide a highly approximate probabilistic estimate. In other instances, it was apparent that the construction of propositions often seemed to be guided by what kind of data was available. Here the process of construction appeared to be guided by a certain degree of pragmatism, to use the wider sense of the term. It may have been the case that investigators wished to consider certain propositions, yet often the sufficient data was not available. They were often constrained by data that they could find, and this appears to have influenced the construction of ensuing propositions. Hence a certain degree of recursivity can be demonstrated in this process. In searching for relevant data, it also appears that investigators may have had to rely on a certain amount of personal initiative in order to determine what data could be considered pertinent to a particular case. The precise nature of these data could have reflected the unique circumstances of the case. This personal (and by extension, intersubjective) initiative could have been based on any number of factors – local knowledge, experience and specialisms of the investigator, degree of knowledge in a particular scientific discipline etc.

It is quite clear then, that proposition formation occurs not in a unidirectional manner, but instead involves an iterative, recursive process in which the construction of propositions is informed by discoveries concerning the precise nature of data sets available to investigators. Constrained by the data available to them, investigators may have to modulate the propositions they intend to test. They may formulate hypotheses viewed as particularly relevant to the case, yet find that the data required to generate the required LR estimates is not available, hence they may be forced to consider other, more testable propositions.

The problems associated with constructing adequate datasets does perhaps reflect a certain contribution from the material realm, in that this issue reflects how the possible orderings of material objects through a series of events may serve to affect how investigators are able to establish numerical representations. However, what is more notable is how discussions over language predominate in these episodes. Hence concerns appear to be centred more on finding the appropriate ways of articulating reconstructive propositions which facilitated a valid correspondence between reference and world. Issues concerning the construction of propositions became notable sites of discussion and intersubjective activity. These kind of concerns place language, a uniquely human activity, at the source of the controversy, rather than as the solution. This is highlighted by the numerous forms of ambiguity that were identified in the study of the formation of propositions. Differences in the interpretation of specific phrasings, ambiguities experienced in attempting to create fully mutually exclusive hypotheses, and difficulties in reconciling linguistic constructs and statistical data were only overcome through discussion and deliberation. Hence the negotiation of such difficulties, through intersubjective agreement, represents another form of Bayes, and shows clearly how Bayesian reasoning may be founded on social practices.

Finally, it became clear that forms of embodied reasoning played an important role in the investigation of incidents. It appears that reconstructive inferences often rely on investigators physically 'acting out' possible causal scenarios that may explain the precise origins of evidence found at an incident scene. The example of the elderly Dutch woman's death clearly illustrates this. One of the most notable aspects of this instance of practical reasoning was that it appeared to contribute as much to the eventual solution of the incident as much as the employment of an avowedly 'Bayesian' approach, so much so that it is difficult to determine which form of reasoning made the most significant contribution. It was a comparison of hypotheses in the Bayesian manner, which led investigators to believe that the case was not a murder as originally thought. The use of Bayes however, appeared to provide little guidance in establishing a suitably credible alternative account of the event. Instead, a series of mock 're-enactments' of the way in which the woman may have died in an accident involving the scissors, made an important contribution to resolving the case to the investigator's satisfaction.

From this discussion it can be seen that a variety of reasoning practices appear to be involved in propagating 'Bayesian reasoning'. Whilst the exact nature of these practices may be highly contextualised, it seems clear that they appear to be necessary in order to circumvent intractable liminal spaces to allow Bayesian reasoning to proceed. This may due to an inability to translate phenomena into quantitative forms, as shown by the issue of soft data, and the ambiguities faced in constructing propositions in the CAI, or it may be due to the inability to reconcile the high information demands Bayes places on reasoners. This in turn may be due to an inability to account for every real-world instance faced by investigators, as in the use of 'drop in' and 'drop out' assumptions in the case of LoComatioN. Regardless of the nature of each form of Bayes however, it can be seen that they are required in order for Bayesian reasoning to relate in some way to the worlds which are under investigation.

9.5 Forensic Investigation and Abduction

In what follows next I seek to briefly comment upon what the findings of the case studies implicate for the Peircian model of abductive reasoning, and particularly the consequences for the adoption of technologies which employ Bayesian inference. A focus on this area has relevance for two reasons. First, the subject of Peircian abduction has been of interest to both those working in the field of forensic science (Nordby 2000; Jackson et al 2006), and to an audience of semioticians and philosophers, who have sought to consider how the logic of criminal detection might serve to act as a model for abduction. In a celebrated volume, the methods employed by the fictional 'detective', Sherlock Holmes, have come under scrutiny from the latter group (Eco and Sebeok 1983). The concept of abduction has also been taken very seriously by figures such as Ian Evett (Jackson et al 2006), and models such as the CAI can be regarded as attempts to consciously adduce a form of abductive reasoning to a forensic context. Second, recourse to the performativity thesis naturally brings forth its pragmatist roots. Although Callon cites Charles Morris over C.S. Peirce as an influence, the two nevertheless exhibit a considerable degree of common philosophical ground, and indeed, Peirce's work was cited as a key starting point by Morris (Petrilli 2004).

Peirce himself differentiated between three different forms of reasoning, namely deduction, induction and abduction. The first of these is generally taken to involve arguments in which the premises are claimed to support a conclusion, in such a way that it is impossible for the premises to be true and the conclusion false (Hurley 2000, p.33). Induction, on the other hand, involves an argument in which the premises are claimed to support a conclusion in such a way that it is improbable for the premises to be true and yet the conclusion false. Abduction is taken to be a less clearly defined notion, but Walton (2004, p.9) cites three key characteristics. First, it can be regarded as a technique for narrowing down a multiplicity of alternative explanations for an event by selecting one or a few particular hypotheses. Second, it can be conceived of as a process of guessing, or choosing the right guess; a fallible process in which wrong hypotheses can be chosen as often as correct ones. Third, it is often involved when a new phenomenon is observed which is unable to be explained by current scientific understandings. On this basis it can be summarised that abduction generally involves the assessment, on the basis of observable signs, of an inferential hypothesis which can account for the observations, which may then be tested in some way to determine the extent of its explanatory power.

Much philosophical debate has centred on the precise nature of abduction. Eco (1983) advances a four-part model. First, *overcoded abduction* involves the formation of a hypothesis about a sense-datum which depends on some knowledge of the context in which the latter is comprehended. For example, the form of a hypothesis concerning the utterance 'man' (/man/) may be different depending on whether it occurs in an English-speaking environment, or a multi-lingual one. Overcoded abductions may often be virtually instinctive, as the context of experience may condition the abduction, and indeed the act of simple perception can be seen to largely equate with the concept of overcoded abduction. The second form, *undercoded abduction*, involves a greater degree of choice. Undercoded abductions relate to scenarios in which there may be an opportunity to consider a selection of possible alternative options, depending on the experience of similar cases, and upon the reasoners knowledge of the world at that particular time. Undercoded abduction therefore involves the imperative to form a hypothesis based on one of these plausible options. Eco cites the example of the astronomer Kepler in this regard, who established the elliptical nature of the orbit of Mars. Given the observations, and his previous knowledge about the behaviour of planets, Kepler had a finite number of choices between various geometrical curves that could account for his data. Abduction in this case therefore required him to formulate and assess hypotheses to help decide which one fitted the observations the best. The choice of hypothesis may also depend on whether it tallies with previous observations of similar phenomena, such as observations of other planets etc.

A third element, *creative abduction*, refers to situations where recourse to a discrete range of options may not be possible. Unlike the previous two parts, creative abduction does not necessarily have to rely on contextual knowledge, or choosing between a limited number of possible hypotheses informed by previous observations. Creative abduction involves situations in which no experience-based generalisations may exist which could serve to account for observed evidence, and thus explanations are generated *de novo* (Schum 2000, p.1660). Creative abduction, does, however, require a fourth stage, *meta-abduction* (Eco 1983, p.207). This is due to the highly speculative nature of creative abduction. Over- and undercoded forms of abduction utilise previously held knowledge of the world, and, within that, assumptions about that knowledge. Creative abduction does not operate in the same way, and may reflect not just a change about the cause of a result, but subsequently wholesale changes in worldview. Meta-abduction refers to the process of determining whether the possible world outlined in first-order abductions based on personal experience, does actually correspond to the external environment.

In his discussion of abductive reasoning in relation to evidence, Schum (2000) however concludes by cautioning that the term 'abduction' is susceptible to over-appropriation (Schum 2000, p.1680). Schum argues that philosophical attempts to further clarify abduction 'do not seem to capture the true complexity of this vital reasoning activity when it is examined in contexts as rich as fact investigation in law' (Schum 2000, p.1680). In focusing on aspects of forensic investigation, the research presented in this thesis corroborates Schum's view, and serves to problematise the notion that abduction can be readily taxonomised. In what follows I argue, with consideration given to Eco's thesis and my own case studies, that a variety of practices that can be construed as 'abduction' are involved in forensic investigation. In doing so I further call into question the portrayal of Bayesianism as a uniform approach to forensic science.

Overcoded abduction can be seen to be present in examples from the two case studies. For example, in order for Bayesian analysis of DNA profiles to be initiated, one must first be able to recognise allele peaks. Although modern capillary electrophoresis techniques can use software to attempt to recognise peaks, their shape is not invariant, to the extent that the recognition of peaks may still lie somewhat heavily in the domain of instinctive personal experience. Furthermore, the recognition of 'soft data' at crime scenes, which may play an instrumental role in shaping the kind of questions to be addressed in the initial stage of an investigation, may occur instinctively, but is also strongly influenced by experience and context. In both cases, the recognition of these forms of sense-data are essential for Bayesian analysis to proceed.

Such identifications are instrumental in enabling actors to determine not only which elements of sense-data are relevant to the investigation, but they also constitute data which will inform the construction of hypotheses for assessment by a Bayesian form of reasoning. This latter practice, of assessing hypotheses, most closely resembles Eco's concept of undercoded abduction. Here however it is possible to see how undercoded abductions rely on a series of overcoded ones, and in doing so we can begin to question Eco's schema. To return to Eco's example of the study of Mars, Kepler would have had to rely on observed sense-data on which to form hypotheses (the observed positions of celestial objects). In Eco's example however, Kepler can be viewed as operating under the conditions of 'normal science' as described by Kuhn; he operates via an established set of methods involving commonly used instruments, and perceives the world in a certain way, where the problems for inquiry are well-established (Kuhn 1964). Given this conditioning, Kepler's observations can actually be said to be overcoded: he knows what to look for on the basis of his previous training and experience. Under normal scientific conditions, the question of what hypotheses should be constructed is a relatively unproblematic affair, due to the specificity of the scientific problems of interest.

To compare this example to some of those encountered in the case studies, it may appear that a similar process is occurring: the overcoded abductions that are the observations of forensic scientists are used as a basis upon which hypotheses are constructed. In these cases however, the nature in which these hypotheses are produced may vary. In the example of DNA profile interpretation technology, the recognition of peaks in capillary electropherograms is vital for determining which alleles may be present in a profile. Yet this is not necessarily a straightforward process. Without the ability to recognise allele peaks as such, it would not be possible to generate a list of alleles at each loci, which could then be subjected to Bayesian analysis to compare defence and prosecution hypotheses in order to determine the probability of a match. Undercoded abductions are therefore dependent on a reasoner to first make overcoded abductions in order to recognise sense-data as meaningful scientific data, despite it possibly being mediated by technology.

In this example, it is still possible to maintain a certain distinction between over- and undercoded abductions, despite the seeming dependency of the latter on the former. The uncertainty over abductive practices is highlighted even further however, in the study of the CAI. Here, soft data, clearly based on overcoded abductions, may be used to directly inform the construction of propositions The Bayesian format would depict this as an undercoded choice, yet whilst it might suggest a simple comparison of hypotheses, the fact such assessments could, at least potentially, have little or no objective basis, leads one to wonder whether a form of creative abduction is taking place instead. Hence the use of Bayesian reasoning in this manner cannot necessarily be taken to represent undercoded abduction, despite the guise of a quantitative 'scientific' mode of investigation. This also, once again, shows the lack of clear means of establishing a *fully* corresponding relationship between thought and world.

What this brief discussion demonstrates then is that the processes which can be taken as forms of 'abduction' are heterogeneous and interwoven. An attempt, such as that put forward by Eco, to make sense of abduction, is subject to struggle if compared to empirical cases such as those described here.

What these cases demonstrate is a liminality in terms of abductive practices. Not even meta-abduction can be seen to be occurring, for it appears that it is impossible to fully establish an appropriate social or material world into which one's inferential claims can make the leap. Instead, the above examples show how the different kinds of abduction postulated by Eco may co-exist, merge between each other, or even remain unable to be characterised, even in a context which is considered a natural home to 'abduction' by both philosophers, and, increasingly, forensic scientists themselves. Hence, the case studies indicate that the concept of abduction appears to be somewhat resistant to straightforward characterisation. Abductive practices appear to be entangled, and possibly interdependent with, the complex world which they seek to explain. If this is taken to be the case, then it reinforces the possibility that 'Bayesian reasoning', itself often taken as an exemplar of abduction, is merely a label for a whole series of different practices. The term 'Bayes' therefore becomes another term for making sense of what abduction is, and is hence a constructed epithet. The world which Bayes requires ultimately has to be modelled and assumed. In this way the tensions between the objective and subjective facets of Bayes become apparent.

9.6 Modes of Hypothesis Generation and Assessment

One further issue which I wish to discuss further concerns the extent to which Bayesian reasoning may be concerned with the domain of the assessment of hypotheses, rather than their generation. Although much philosophy of science has tended to by-pass the issue of a 'logic of discovery', debate continues in other quarters as to whether such a logic of discovery can be identified, and – as discussed earlier in this chapter - abduction has been portrayed as possibly leading the way to showing what such a logic might entail. Creative reasoning in science has been attributed to a number of processes, such as radical ontological change, conceptual combination, analogical and visual thinking, anomaly resolution etc (Magnani 2005), which might function as the catalyst for abductive thinking. The suggestion of such modalities, does, however, beg a series of questions. There is no sense of *how*, or by what means, these suggested processes may be made to come into play.

Are such modalities innate to human reasoners, might they be evolutionary as Peirce himself thought? What kind of stimuli trigger these kinds of responses? Can such stimuli be seen to exist in the first place? Moreover, do they fully reflect the complexity of the sociomaterial worlds in which reasoners inhabit, one which may gradually reveal further layers of complication, presented as much by social factors as material ones?

The suggestions which are summarised by Magnani suggest a conception of the world as a relatively simple puzzle; once a particularly modality is used the world may be re-arranged, or be made to be viewed in a different light. The assumption here is clearly one of an ontologically separate world. Yet even in these accounts there is no explanation for *why* a creative strategy is chosen *at a particular moment in time*. In this schema, the resolution of these puzzles is simply a matter of (perhaps unconsciously) selecting a strategy which then leads to the resolution of the issue. The precise moment at which this strategy is chosen could be regarded as a virtually instantaneous, and perhaps instinctive phenomenon; redolent of the metaphor of the lightbulb appearing above a cartoon character's head. No account is given for what stimuli might lead to such a phenomenon, and it is difficult to envisage how such an account may proceed.

The results in this study do not, however, suggest such instantaneous occurrences. What should also be apparent from these studies is that the formation of inference is a decidedly sociomaterial affair. This must be contrasted with the above, which seems to assume an isolated reasoner, capable of forming instantaneous solutions. Therefore we must consider how one might proceed with an alternative explanation for creative inference. A major problem in relation to Bayes concerns the construction of worlds through which prior probabilities can be measured. The first case study shows how the assumptions needed to generate a Bayesian formula may be elaborately constructed in terms of their supposed correspondence to the material realm for which they seek to account. The second case study also highlighted this, showing how data required to construct a prior probability to test a proposition may not be in existence, and how this might lead to a coproductive process between proposition and dataset. In both instances, the generation of a Bayesian hypothesis involves a need to overcome an intractable correspondence gap between proposition and materiality, which occurs through strictly sociomaterial practices. This thesis indicates therefore that hypothesis creation comes about through a need to fill this gap. The means by which this occurs may be contingent, and may require a plethora of resources and actors. It is a complicated, and possibly gradual, endeavour, and cannot, therefore, be accounted for by instantaneous thought processes alone, however they may be thought of.

These studies on Bayesian reasoning have instead shown how hypothesis formation may be a complex practice, involving both contingent and emergent social factors. In order to be able to form hypotheses, a series of constructions need to be enacted. In the first case study, this required elaborate constructed concepts, and recourse to questionable assumptions about the material realm, in order to enable the technology to explain it. The second case study showed even more clearly how ambiguity had to be managed and debated intersubjectively in the course of constructing propositions, and how hypothesis formation was co-produced with the datasets it was meant to draw upon. It also showed, through the existence of 'soft data', how previous experience played an important part in helping forensic scientists postulate hypotheses.

What needs to be reiterated is that all these behaviours which led to hypothesis formation were invoked in direct response to the desire to use Bayes in the course of investigations. The performance of Bayes therefore, had a direct effect on the kind of hypotheses constructed. This contrasts with the view of its acolytes in forensic science and evidence scholarship, who in some ways view Bayes as a means for the assessment of hypotheses. The examples above, however, show how the mode of generation and the mode of assessment become conflated.

Both case studies therefore reveal an interesting tendency of Bayesian reasoning. The research indicates this conflation of hypothesis generation and assessment is perhaps even acknowledged and accepted by staunch Bayesians themselves, as signified by Ian Evett's comment in Chapter 5 concerning the way in which Bayes 'guides the reasoner in asking the right questions'. In itself, this tendency could indicate a strongly performative phenomenon; forms of Bayes emerge which reinforce and prescribe a particular set of behaviours which come to resemble more closely the representative construct from which they originally draw upon. If the application of Bayes can be taken to be a form of performativity, then this essentially pragmatic process shapes its own form of reasoning.

9.7 Performativity

The notion that reasoning practices might arise in an interdependent manner with the assemblages which they attempt to make sense of, may appear to be consistent with the idea of performativity as put forward by Callon, Pickering et al. With forms of reasoning being undefined, it may well be the case that assemblages, or agencements, control their own way of making sense about themselves. In what follows however, I investigate this claim further by reference to my two case studies.

In Mackenzies, and Mackenzie and Millo's study of performativity, the Black-Scholes-Merton equation took some time to be accepted, as, for a while, the extent of its accuracy and fit with reality was initially doubted (Mackenzie and Millo 2003). The eventual acceptance of the formula appears to have been dependent in part on the development of the derivatives market as a whole (Mackenzie and Millo 2003). Over time, as derivatives were demonstrated to perform more in line with the prediction of the equation, it gained greater acceptance in the field. Following the 1987 stock market crash however, the equation was perceived to make inaccurate predictions of implied volatility of derivatives, which was compensated for by the introduction of 'skew' estimates. The introduction of the skew estimate was cited by Mackenzie and Millo as an instance of 'counterperformativity', in that it reflected a lack of correspondence with the predictive warrant of the equation.

In the case studies outlined in this thesis, the focus has not been on a mathematical device that enables prediction. Instead, Bayes Theorem, in the context of forensic science, has been utilised for a different purpose. In this case, Bayes has generally been used as a means of testing and assessing one's own inferences. It acts as a means of assembling a set of representations about variables relating to a piece of evidence, or various pieces of evidence, into a coherent whole. Rather than serving as an *a priori* means of prediction, Bayes Theorem acts a *post hoc* means of testing inferences. There is a clear contrast then, between these studies of the Black-Scholes-Merton equation and the use of Bayes Theorem as outlined in this study. Furthermore, Bayes can be regarded as an all-encompassing probabilistic theory, and indeed, forms the basis of what some regard as an all-encompassing philosophy. The form of Bayesian theory used in forensic science is itself a modulation of the original theorem, but not such that it deviates significantly from the original derivation. This version of Bayes enables investigators to compare prosecution and defence hypotheses, and is common to virtually all applications to forensic science, including those technologies featured in the case studies.

Furthermore, the two examples presented here demonstrate how the performative renderings of Bayes enable actors to contend with the liminal spaces which arise when reference and world are attempted to be brought into correspondence. A study of Bayes, in such a way, shows how performativity can be used to fully account for the relationship between language and sociomateriality. This contrasts somewhat with previous performative studies of mathematical constructions, in which a certain division between the two is still apparent. For example, in Mackenzie (2003), and Mackenzie and Millo (2003), there still appears to be a certain dichotomy between sociomateriality on one hand, and reference on the other. The latter is not as closely constitutive of the former as is perhaps suggested by the argument of Callon. Whilst the performances of the sociomaterial assemblage are linked to the Black-Scholes-Merton equation in ways which either suggest a correspondence to the latter, the equation in itself is regarded as being interpreted in an unproblematic way. The equation has a set of clearly definable variables, and appears to exist for a particular purpose, even if that predictive purpose has not always reflected reality over time. This is perhaps not a fault of the performativity thesis as a whole, but the question remains regarding the range of its application.

One further possible shortcoming of the performativity thesis, as it has been regarded thus far, is that relatively little emphasis is exercised on studying the *nature of performance* itself, or what kinds of performances are facilitated via the construction of agencements. The only issue that tends to be highlighted by performativity concerns that of whether sociomaterial assemblages begin to reflect the representative constructs which are themselves constituted by a particular assemblage. However, the models advanced in the work of Callon (2006), and Mackenzie and Millo (2003), do not begin to reflect the diversity of forms of performance that might be necessary in order to constitute these assemblages. Indeed, a consideration of the fluidity of meaning of Bayes Theorem, as focused upon in the research outlined here, highlights a key agential element, namely that of *interpretation*. This is not a process which occurs in passive receivers of data, nor is it a Cartesian process occurring in isolated instances. Interpretation is an active phenomenon, dependent upon the presence of embodied and experienced actors. This however, poses a problem for the performativity thesis. It is these aspects which the Bayesian framework has difficulties with accommodating, and indeed, checks on personal intuition form part of the motivation behind the use of such a framework. However, it seems to be the case that a framework which attempts to overcome the inaccuracies of personal intuition is itself dependent on subjective data.

A discussion of performativity prompts a short digression, whilst I raise some methodological questions: For example, does a successful example of a performative assemblage necessitate the total removal of phenomenological aspects, and hence aspects more commonly identified as peculiar to human agency? Can performative systems only operate in conditions in which human agency is not *a priori* given, and agency becomes malleable, an independent variable of the constructed network? These questions raise the issue of possible regress, namely the concern that performative studies, by their very existence, may themselves act as legitimating forces, contributing to the construction of a certain ontology. Performativity emphasises the role of actions in facilitating reason; however, on these terms the *act of writing about performativity* may in itself be seen as a way of creating a position, and leads one to doubt its basis as a suitably reflexive approach.

Another related methodological issue concerns the temporal aspect of performativity studies. Many of the performativity studies have tended to focus on relatively short spaces of time. A discussion of the place of time in semiotic studies of science points to a set of issues which possibly hold more profound consequences with regard to methodological symmetry. If we take into account the longer history that Bayes has enjoyed, it is possible to see a process which transcends the kind of assemblages conceived in performativity studies which have tended to be more short-term in character. A key question is whether the version, or versions, of Bayes adopted by forensic scientists represent a break with those of the past.

What semiotic studies overlook is the fact that elements in networks experience and are shaped by time differentially. Elements in a network are subject to relative temporalities of development; different organisms, be they plants, animals and humans, will grow, live and die along different temporal trajectories. This is something however, that appears to be ignored in semiotic accounts, whereby the existence of each element is somehow posited as being dependent on the network as such. Furthermore, each element in a network experiences time differently. This is none more so than in the case of humans, who are not only able to retain and code experiences along a temporal vein in a whole series of embodied, cognitive and reflexive ways, but are also able to fully express this experience in a reflexive manner via language. This is a unique property, and once more points to a more privileged role for human agency than is given in semiotic accounts.

9.8 The Problem of Agency

The requirement to shape speculations, conjectures and inferences into a formalised system of comparable propositions, as is the case with the Bayesian approach, merely serves to highlight the difficulties in applying such a mode of probability to an area like forensic science. The kind of practices outlined in this chapter can be seen to act as methods to circumvent the discordancies that arise through attempts to perform Bayes in such a context. The existence of these practices shows that the mutual construction of reference, materiality and social ordering, is neither an unproblematic nor uncomplicated process. Whilst many STS studies have sought to emphasise this point, questions remain as to whether such approaches capture a sufficiently appropriate level of requisite detail. Indeed in this case some of the discordancies appear to be almost intractable, and it is unclear whether semiotic approaches are able to account for the practices that arise to circumvent these problems.

As the CAI study in particular demonstrated, language and its meaning, the most human of interactions, is the source of contestation, rather than a locus around which understandings converged. Whilst epithets like 'Bayesian reasoning' might indicate the existence of a performative construct, they hide these sources that occur at a localised level. Furthermore, it is difficult to consider how any such controversies can be considered as anything but intersubjective in character, and indeed this study showed how the clarification of propositions was a pragmatic process, which involved a necessary amount of intersubjective discussion. Furthermore, this case study also showed how experiential and embodied aspects of reasoning continued to be incorporated into the process of investigation, even when Bayes was actively invoked. It also showed how such embodied practices had the potential to transcend the Bayesian process, in a manner for which the latter could not account.

At issue is not just *what* kind of reasoning forms are being generated in each case, but also *where* this reasoning may be located. What lies at the heart of each technology are the algorithms that drive each system. It is these algorithms which form the crucial link between the way in which Bayes may potentially be used, and the world required for this potential to be realised. However, there is one important difference between the two which should be noted. In the case of Automated DNA Interpretation Systems, the algorithms are supposedly realised by computerized means; although conceived by humans, they are run automatically, supposedly 'objectively', by computer systems. By contrast, the algorithms involved in the CAI are run in human minds. The CAI itself only provides a set of representations, which may be thought of as broad guidelines, or operating procedures, for reasoning about evidence in a supposedly 'Bayesian' manner. The actual operation of these algorithms, or the *actualization* of them, is performed by human reasoners. This difference already indicates the potential for a certain diversity in the way Bayes is actualised, whereby there are differences in which key elements of the performative process are localised. More importantly however, whilst it could be argued from a performative standpoint that it is the assemblage, (which constitutes each form of Bayes), that determines the kind of agency which ensues, the example of the CAI raises particular problems for this position. In this case, the way in which these algorithms are put into practice by humans is neither a purely mental phenomenon, nor is each individual actualisation an isolated instance, unaffected by previous actualizations. As this case study demonstrates, 'Bayesian reasoning' using the CAI has the potential to involve a high degree of both embodiment and experiential judgement, regardless of the quantitative trappings the CAI seeks to adduce to the process

At the heart of each supposedly performative process therefore, lie different aspects of agency. The processes involved in the case of Automated Interpretation Systems are, automatic, impersonal, disembodied and non-subjective. Those involved in the CAI are, however, fallible, contextualised, personalized, embodied and highly situated in terms of an individual's personal history and location. If different reasoning processes lie at the heart of these apparent examples of performativity, one may not only ask whether methodological symmetry is justified, but also whether each of these different algorithmic forms may hold implications for the success of each of these projects. In considering this latter question, one issue to be kept in mind is the kind of 'Bayes' that is emerging from each agencement. Furthermore, what is the nature of the relationship between each depiction of 'Bayesian reasoning' and the manner in which this reasoning proceeds? Can each reasoning process, or algorithmic enactment, be said to be constructed within each agencement, as a conventional reading of the performativity thesis might have it?

The studies outlined in this thesis indicate that explanations utilising a largely semiotic approach are insufficient to fully account for the formation and consequences of the application of Bayesian technologies in forensic science. Hence, one is led to question just to what extent ANT-based studies are able to fully account for the totality of agential interventions that the material domain is supposed to contribute to the co-production process (in the Latourian sense), and to what extent it can account for the totality of corresponding social practices. This points to the concern that to be truly convincing, the semiotic approach to STS, as it is currently conceived, may require more fine-grained explanations of the processes involved in creating reference. It is difficult to envisage how this may occur, for this would suggest a more rigorous account of practice, which might expose, counter-productively from the interests of many semiotic accounts, the continuing predominance of human agency. However, it is difficult to envisage how such a position might provide such finely-grained accounts, unless there is some form of reconciliation and synthesis with another ethnographic method which introduces a phenomenological aspect.

The tendency to place nonhuman agency on an equal footing thus results in an incomplete account of the process by which actors are able to make meaningful interventions in linking theoretical representations with actual practices of evidential interpretation. The kinds of technologies and infrastructures developed in order to aid the process of evidence interpretation are primarily the result of human activity in this regard, rather than a series of drivers in a wider, unfolding process of co-construction of theoretical representation and material shaping[1]. Furthermore as these contingent practices come about due to the ontological liminality induced by the application of Bayesian theory, it can be seen that they result from intractable semiotic uncertainty, in contrast to accounts which portray actors as capable of making sense of their external environment through the unproblematic linkage of signs and inscriptions.

9.9 Does Bayesianism Reflect, Model or Construct Reasoning?

Bayesian reasoning involves the consideration of relevant questions in a highly process-led way. Questions are considered via an iterative, serialised procedure, and in the forensic application of Bayes, involve the consideration of two competing hypothetical propositions. As I have described, the use of such a mode of reasoning has been repeatedly cited, at least in certain quarters, as the most logical way in which to reason under uncertainty. Aside from the nuanced objections this argument might elicit from the community of probabilists, not to mention scholars of formal logic, one most also pose another question. Despite the apparent logicality of Bayesian reasoning, is it an epistemological form which comes readily to human reasoners? Does it correspond with the way in which human actors make sense of events, particularly those of which they are uncertain? Or do actors comprehend events in ways which reflect a significant departure from the Bayesian portrayal?

As the results of the two case studies demonstrate, the means by which hypotheses are assessed may be interdependent with the manner in which they are formed. This casts doubt on the portrayal of Bayes as a procedure entirely distinct from the generation of scientific hypotheses. This is further compounded by the fact that Bayesian technologies can be recognised as

[1] One question which arises is whether this invariant domain can be classified as exerting a form of agency, in that it provides the means for co-production 'on the top of it' as it were. But then if one were to pursue this approach, the question arises of whether such permanence can be classed as 'agency' at all; it is difficult to see how a domain of permanence could be characterised in a manner which accrues it with agential power, for to bestow it with mechanisms of change seems hopelessly contradictory.

emergent social constructions. Bayes Theorem cannot therefore be regarded as a simple tool-like device into which data can be unproblematically plugged in and processed. The complexity of the forms involved in bringing Bayesianism into being contradict the latter representation, and instead show how Bayesian technologies may co-produce hypotheses and data, as opposed to the latter being exogenous entities.

One alternative concept worthy of attention is the notion that actors reason in a manner more akin to an act of narrative. This differs from the Bayesian mode in that it may entail an inclination to consider a series of events as somehow already linked; the reasoner puts events in some form of order, and uses various strategies to account for gaps in the story, or contradictory elements. These strategies may use creative accounts, or they may involve personal experience, recalling the 'commonsense' form of reasoning supposedly used by jurors in Lynch and McNally (2003). Scholars have, in the past, applied the narrative concept to a wide variety of areas, including economics, psychoanalysis, and legal discourses and science (Nash 1990). The work of Hutto (2007) is instructive in this regard (Hutto 2007). His studies on 'folk psychology' emphasise the innate ability of humans to comprehensively account for the behaviour of others by recognising the 'story' of the third party. This of course, involves the capacity to fully comprehend the set of beliefs, desires, emotions that a third party might have been to subject to in their actions (Hutto 2007, p.43). According to Hutto, this is gradually acquired in childhood, through 'socially scaffolded means' (Hutto 2007, p.47).

This research demonstrates that a narratorial aspect may be present even in the technological practices of forensic scientists. This is particularly so in the case of the interpretation of complex DNA profiles. Such an activity should be entirely removed from the host of other concerns which might play on the minds of investigators involved in a particular case; to invoke just one of the Mertonian norms of science, they should be 'disinterested' observers of phenomena-data. This particular case study, however, highlighted how the process of story construction played a vital role in the reasoning processes of scientists. If true, the claim that LCN, in particular, can *only* inculpate rather than exculpate a suspect throws a considerable amount of doubt on the portrayal of DNA evidence as a neutral indicator. LCN interpretation most often occurs when conventional profiling techniques have failed to provide sufficient intelligence on which to provide the basis for a convincing case. The example of 'cold cases' show quite strongly how scientists, who will work together with police on such cases, may be informed by a wider narrative framework in which the DNA evidence forms an important thread.

This extends through to the finer points of the reasoning practices used by forensic scientists in their interpretation of such profiles. As I have shown, concepts such as 'drop in', 'drop out' etc are invoked to justify the claims made regarding the kinds of hypotheses constructed. I have demonstrated how such concepts are actively converted into quantitative signifiers, a necessary step for their incorporation into the Bayesian algorithms used to drive interpretation systems. However, as I have also shown, the realisation of 'drop in', 'drop out' and 'contamination' is open to a considerable amount of question, with regard to how accurately such constructions correspond to the forensic environment from which the original DNA samples may be derived. Although it may not be quite appropriate to view these key elements of LCN interpretation as fictions, they nonetheless act as important constructions, based on questionable assumptions, around which 'stories', to use the vernacular of one respondent (Interviewee 3), are created within which the incomplete and confusing LCN data can be rendered as sensible and meaningful. These constructions therefore act as important links between the seemingly random array of numbers and letters that represent LCN DNA profiles, and the wider *mise en scene* that constitutes a criminal case. The lack of consideration given to the environmental conditionings of DNA is shown in the use of such assumptions, which instead of reflecting this appear to be based in another space where certain scientific beliefs do not appear to hold.

This example also demonstrates the rapid tendency of a Bayesian schema to break down in temporal terms. As Interviewee 9 stated, a fully Bayesian approach as applied to forensic evidence might consider the entire process, from the immediate recovery of evidence:

> 'What would be useful is if you went away and did a Bayesian analysis of the whole process right from the recovery of evidence, as there are so many factors to consider even from there' (Interviewee 9, 2008)

This is not apparent in the above example however, where assumptions are made about how the data came to be. A more advanced Bayesian form might concern itself fully with the process from recovery to interpretation, and thus the way in which the data came to exist could be considered in an even more rigorous manner. The factoring of other issues pertaining to recovery of DNA could even go further to consider further issues, such as the prior probability of preferential degradation at a given locus, such that this could be incorporated into the P(D) calculation. Even this brief discussion however, is possibly sufficient to reiterate the fact that a potentially vast amount of information could be factored into a comprehensive Bayesian analysis of LCN DNA.

The fact that such information is omitted from this Bayesian technology draws doubt upon whether a supposedly highly automated system can be seen to fully adhere to Bayesian reasoning, as least as it depicted in theoretical accounts. Instead of complying with a formalised system of 'scientific' reasoning, the use of LCN in a criminal case is highly contextual, and its apparent success may be often due to the kind of stories in which it plays a role, rather than acting as a standalone source of objective evidence. Furthermore, devices such as 'drop in', 'drop out' can be seen to act as devices linking the data to the circumstances of the case. LCN interpretation is therefore not necessarily a 'logical', iterative procedure, but one which is embedded in a series of assumptions (which may correspond to a different reality), and contexts. Whether this is folk psychology may still be up for debate, but it certainly shows that the 'scientific' reasoning practices employed are not held in isolation; they may require a series of contexts for data to acquire meaning. More interestingly still, elaborate links may need to be constructed in order maintain a technology in the juridico-scientific framework. The irony here is that these links actually enable scientists to justify their stories, and further still, these links are actually a vital part of the Bayesian architecture itself.

9.10 Bayes and Expertise

A discussion concerning the location of Bayesian reasoning has implications for the manner in which expertise may be conceptualised. Automated Systems for DNA Profile Interpretation could be regarded as 'expert' systems, but it is perhaps more appropriate to regard them, as one interviewee insisted, as 'intelligent systems' (Interviewee 2). Certainly, it seems that the construction of systems such as LoComatioN and PLS reflect the limitations of their makers. This in a way provides support for the Latourian notion of expertise being distributed throughout an assemblage, as no privilege can be accorded to either machine or human, and each become interdependent on one another.

The forms of Bayes identified in the CAI study however, may suggest otherwise. Here, experiential and embodied components played a greater role in propagating Bayesian reasoning. Furthermore, the issues experienced in producing propositions reflected a strong intersubjective component. Moreover, the fact that individuals were keen to cite 'Bayesian reasoning' in justifying their behaviours suggests that it has important role to play in constructing expert identities. This is also bourne out by the discussions in Chapter 5 concerning the history of the incorporation of Bayes into forensic science. Bayes appears to be playing a major role in shaping a disciplinary identity for forensic science, and with it, representing a set of behaviours to which 'expert' forensic scientists are meant to adhere.

In terms of the accounts of expertise discussed in Chapter 2 though, there is little to support any of the conceptualisations put forward by Dreyfus. His account does indeed to suffer from a rather solipsistic view of expertise and skill. For this particular form of expertise to occur, 'experts' need to be recognised as such (by self-asserting a commitment to Bayes). Furthermore, whilst the operation of expert reasoning may encompass an experiential aspect, that alone is not enough to account for the high level of intersubjectivity apparent in other propagations of Bayes. His criticisms of AI, are, however, echoed in the clear limitations apparent in the technologies featured in the case studies, and the need for a strong element of human involvement. Yet whilst it highlights the limitations of computerised technologies in abductive reasoning, his work perhaps has little to say about the CAI, and a limited conception of the term 'technology' is apparent here. Bayes is a curious technology in that it does require a strong human contribution for it to work, and it is perhaps that a more nuanced conception of technology is required.

Furthermore, the fact that language become a site of contestation and ambiguity in the case of the CAI, casts doubt on Collins and Evan's particular version of expertise. It appears that in their case, something more is required which perhaps takes fuller account of the problems a communication-led mode of expertise contains. The main problem identified within this study was the problem of correspondence between language and world. This problem is the source of issues which lead to uncertainties about how language may be used. Whilst Collins and Evans position might require common language, it is not always certain how language itself refers to the environment which encompasses this expertise. This study indicates that language may be a source of contestation for experts, rather than a locus around which expertise can be recognised. There also consequences for Fuller's conception of expertise; whilst a dramaturgical element is present in terms of the self-assertions made in the literature and by forensic scientists themselves, the ontological issues which are raised in the course of the application of Bayes play a considerable role in the behaviours displayed. What is shown is the need to somehow make sense of a complex and often intractable world. Hence the 'expert' behaviours pertaining to Bayes should be regarded as contingent and highly localised in character, and therefore not always readily rehearsable.

On the other hand, the fact that the CAI has been vulnerable to opposition from those actors it was designed to benefit shows how the success of an 'expert' technology of reasoning can be shaped by external actors. In this way, the 'expert' behaviours of forensic scientists is still influenced by a variety of institutional forces beyond the control of the developers of the CAI. These include relatively well-entrenched sets of normative attitudes with regard to the role of the police in criminal investigations, and how forensic scientists stand in relation to them. The opposition from within the FSS itself may also reflect a certain tension apparent between the need to satisfy commercial interests, and the desire to produce a 'rational' system that satisfies law enforcement needs. Aside from questions of how Bayes can be shaped to make sense of the material domain, an equally complex set of relationships exist between these wider forces and the CAI. Whilst recourse to exogenous forces may not be sufficient to explain the ontological issues that lead to forms of Bayes, they cannot be discounted in playing some form of role in shaping 'expert' technologies, and more research may be desirable to explore the precise nature of the relationship further.

9.11 Implications for Evidence Scholarship

I conclude this chapter with some remarks regarding the implications that my research may hold for the growing field of evidence scholarship. This work is of particular relevance given the relative prominence that Bayesian theory appears to currently occupy in this area. One aspect I have sought to draw attention to, and which evidence scholars and statisticians may wish to consider further, are the practical issues associated with applying Bayes in 'real-world' contexts. These are not simply inconveniences which can be disregarded in theoretical summaries; as this study has sought to show, the theoretical representations which emerge from disciplines such as statistics and evidence studies will inexorably be bound up in localised sociomaterial realms. It is through such domains that representations are made meaningful. As this study of Bayes has attempted to show, this may often occur in ways markedly contingent to the original formulations of equations and formulae, despite what the users of such formulations may themselves claim. Furthermore, the results of the study indicate not only a possible need to somehow consider how mathematical representations, or 'inscriptions' to borrow Latour's term, are intertwined with sociomaterial domains, they also show the limitations of statistical methods. Experiential and embodied aspects continue to play a strong role in criminal investigation, in a way which transcends the capabilities of Bayes.

The field of evidence studies perhaps needs to investigative the purpose, and the role, of approaches such as Bayes. However, that is not to say that attempts should need to prescribe any form of limitation or all-encompassing method. Instead, the work outlined in this thesis can be seen to tentatively point towards the need to understand further how statistics and probability *is* incorporated into the realm of evidence interpretation, rather than how it *ought* to be. Simply put, what this set of studies shows is that evidence scholarship needs to concern itself more with descriptive accounts concerning how people reason about evidence, and how different contexts, technologies and situations affect these processes. It is here that sociological approaches may be of considerable benefit. Of course, to make a truly valuable contribution, such approaches would benefit themselves from an informed understanding of the statistical issues, and hence there is a clear need for open communication if there is to be any *rapprochement* in this regard.

A significant proportion of research activity within evidence scholarship is currently devoted to developing the Bayes Networks (BNs) introduced in Chapter 4, and indeed these forms can be regarded as somewhat emblematic of the field itself[2]. Whilst BNs are not employed in casework contexts as yet, practitioners have noted the potential of such a technique in certain aspects of forensic science (Gill et al 2006). It is therefore worth briefly considering BNs in further detail, as relevant discussions by evidence scholars echo some of the issues raised here.

The discussion of these two difficulties, contrasts with other discussions of BNs in the evidence literature, which tend to present them in an unproblematic fashion (Taroni et al 2005). Much of the literature from leading figures in the field has tended to view this technology as a means of remedying errors associated with evidential reasoning, particularly those involving human

[2] (http://www.evidencescience.org, accessed 9 October 2008)

judgement. As well as solving problems such as bias or perception error, technologies such as BNs are also viewed as contributing to overcoming supposedly logical shortcomings, such as the prosecutor's fallacy (Evett and Dawid 1997). In these discussions, BNs are presented in simple and perhaps idealised forms. It is also worth noting that much of the BN research presented in forensic and legal literature demonstrates the technology via the consideration of problems concerning DNA profiling. As interviewee 1 pointed out, application of BNs to such issues is relatively unproblematic given the generally ordered and predictable nature of Mendelian inheritance (Interviewee 1, 2007).

More recent developments have involved attempts to apply BNs on a more holistic basis to criminal casework, following in the path first charted by Kadane and Schum (1996). These attempts however, have tended to elicit more critical discussions within the field. Leucari (2006), in a comparison of Bayes Networks and Wigmore charts, views the latter as devices for depicting and hence constructing arguments. This, she argues, is not a property apparent in Bayes Networks, which force users to make assumptions about the events related to the nature of the problem under consideration. This, it seems, may strongly influence the way in which Bayes Networks may be constructed to depict a particular event. In the case of complex events, such as, say, an armed robbery, there may be considerable scope for subjective judgement. As Schum (2005) notes, in his lucid discussion of the application of Bayes Networks to such situations, 'someone can always come along and identify a node you left out' (Schum 2005, p.14). Another problem, this time common to both BNs and Wigmore charts, concerns the potentially limitless amount of information that could arguably be required in order for a chart-based to provide an accurate depiction of an event: 'Wigmorean methods seem to be the epitome of compulsive behaviour' (Schum 2005, p.15). This issue is not unique to Wigmore charts, for the same criticism is cited as one of the most common shortcomings of the Bayesian approach (Rawling 1999).

According to this view, these newer evidential technologies provide a limited contribution to the problem of reasoning about evidence, and, despite the claims of some researchers, continue to place a considerable burden on the human user. Far from solving the problems associated with evidential reasoning, recent research in evidence studies appear to have exposed them, a phenomenon which partially correlates with the findings of this thesis. The leads one to consider what position this leaves the field of evidence scholarship, given the strong influence it has exerted upon developments in forensic science.

This thesis has highlighted the complexity of the practices involved not only in reasoning about evidence, but also in applying a supposedly monolithic procedure to this aim. In my studies I have not sought to ask why these practices have emerged in the way that they have, for that would contradict my broad position of these practices being autochthonous and part of emergent sociomaterial assemblages. It is necessary however to pose the issue of what the implications the apparent conflation of the modes of hypothesis generation and assessment, as represented by these practices, holds for the field of evidence scholarship. It may appear that the kind of critical position adopted here may be totally incommensurable with the approach taken by evidence scholars, but in what follows I attempt to provide an initial indication as to how the kind of approach outlined in this thesis may yet be able to be reconciled with probabilistic approaches to evidence.

The results of this study may not necessarily be viewed therefore as a deathknell for the evidence studies project, and it is not my intention to suggest so here. On the contrary, the fact that Bayes is a performed phenomenon (albeit with a strong agential component) allows one to consider what factor might play a key role in this conflated process. Recall that the first case study showed how Bayesian reasoning processes involved the instrumental contribution of constructed concepts in stabilising this particular sociomaterial assemblage. Consider also the problems highlighted by the second case study in matching linguistic constructions with the material realm, and the use of soft data. In both cases then, background assumptions have played an important role in the construction of Bayesian reasoning. Yet in evidence studies literature, such background knowledge is seemingly ignored.

Often, background knowledge will be at the very most acknowledged by its inclusion in a Bayesian formula the following way:

$$P(Hp| E, I) = P(Hp)\, P(E|Hp, I)$$

This is similar to the equations presented in previous chapters, but note the inclusion of 'I'. This is commonly used in the theoretical literature on Bayes to denote the consideration 'background knowledge' in a Bayesian formula. Yet often the employment of the letter 'I' is as far as the treatment of background knowledge proceeds. This may be due to the fact that 'knowledge' may be simply too problematic to comprehend in a way amenable to a Bayesian way of thinking. If we recall again the criticism by Schum regarding the informational demands that Bayes places, it quickly becomes apparent how the complexity of the world might quickly break down the formula. There is also the problem of defining what actually constitutes knowledge. Bayesians appear to prefer a conception of knowledge defined as quantitative measures of belief, yet the example of soft data shows how knowledge relevant to a Bayesian calculation may be involved in a form which is not readily quantifiable. Background knowledge might also take the form of embodied knowledge, such as the possibilities of physical reconstructing an incident. A further issue to consider is that the type of background knowledge will vary greatly, influenced by factors such as an individual's education, training and other personal experiences.

What this thesis has shown is that greater understanding is needed on the part of evidence scholars concerning the way that Bayes itself is appropriated by human users, rather than viewing it as a simple means to an end. It appears from the research presented herein that the manner of appropriation may both depend on certain ontological assumptions, and on the precise type of knowledges which are deployed. What this research indicates therefore, is a possible need to consider the nature of such background knowledge in a far more rigorous and critical fashion. This in turn points to a continued role for a form of sociological inquiry which could aim to identify and highlight the practices informed by these knowledges, which in turn propagate the kind of technologies prescribed by evidence scholars.[3]

Such an approach would however, need to take each case on its own merits, for the kind of practices involved may well vary from case to case, yet if considered in combination it may be possible to identify certain patterns which might aid further consideration of the challenges involved in applying evidential technologies. It is not my intention however, to claim that sociological inquiry can alone lead to the kind of technological outcomes that evidence scholars might desire. What it does highlight is the potential worth of an interdisciplinary approach to problems concerning evidence. Such an approach would require fully open discourse. Through this it may be possible to explore in more detail the potential issues involved. The kind of study I do propose through this research is not one which seeks to 'deconstruct' science, in the loose 'postmodern' sense of the term. Instead, I propose a more critical approach which seeks to question the science on its own terms.

Methods such as qualitative interviewing could have an important role to play in this process, being used to ask innocent-sounding, but possibly scientifically awkward questions, in order to get scientists to think more critically about their work. In the course of this research the answers put forward by interviewees have often contrasted sharply with published material in terms of both their candour about the study of evidence, and the complexity of the problems under discussion.

This would however, require an openness on the part of those who wish to advance these technologies. Doubts can be raised as to whether this might be the case. The belief that Bayes, and statistics in general, and what they stand for, should occupy the most privileged ground, appears from observation to sometimes become a blinkered view. It still seems that disciplinary tensions still abound, and particularly towards a sociological approach, which may

[3] Some Bayesians have made recourse to psychology, but this only raises a limited number of epistemological questions. The kind of approach here, which has directly engaged with the subject, albeit in a critical fashion, might enable further unknown issues to be uncovered.

possibly be a hangover from the so-called 'Science wars' of the 1990s. A further problem however, concerns the increasing lack of transparency displayed by certain key actors in this study, most notably the FSS. This is highlighted both in the account given in Chapter 7 concerning both the Omagh trial and DNABoost, but also in the difficulties in gaining access to members for interview. It is my suspicion that such opacity may reflect certain doubts in the technology, which is understandable, but it is hard to see how future mistakes may not occur. Wedded to this is the increasingly commercialised nature of forensic science provision, which may increase the reluctance of companies to be open to approaches, albeit at the possible expense of proper scientific scrutiny.

9.12 Conclusion

The results of this thesis hold mixed implications for the various strands of STS. Whilst the application of Bayes to forensic science provides a relatively strong measure of support for the performativity thesis, this is tempered by some other aspects of the 'forms of Bayes' identified in the two case studies. Most notably, many of the practices involved in bringing 'Bayesian reasoning' into being show a decidedly human contribution. The practices involved in constructing 'Bayesian reasoning' are more consistent with STS positions which emphasise human agency, and hence there are parallels with the ethnomethodological position advocated by Lynch et al. Forms of Bayesian reasoning could be identified as emergent phenomena, and the diversity of forms suggests that they are closely linked with the locale and context of application. In these instances, the application of an ostensibly theoretical construct may therefore be regarded as a highly practical achievement. Furthermore, the interdependency of forms of hypothesis generation and assessment informed by both scientific and legal presuppositions, indicates parallels with phenomenon of 'co-production', in the sense of Jasanoff's use of the term (Jasanoff 2004).

My studies have sought to highlight not just the practices involved in propagating Bayesian reasoning, but also to locate the origins of them. Such practices can be located in the sources of liminality that these practices represent attempts to circumvent. As the findings, these liminalities may exhibit different forms in relation to the context in which is Bayes is being performed, but what is common to them is that their existence is realised by the invocation of Bayes itself. This occurs through the intractable informational demands placed on reasoners and reasoning systems by the Bayesian framework, and commits to formulating a series of means of negotiating these liminalities. In this way then, ontological orderings are both a result of the Bayesian schema being put into practice, and the reason for these practices emerging.

The existence of these liminalities indicates the manner in which further inquiry may proceed. Studies looking to focus on Bayes should begin to identify these sources of liminality, as they may well prove instructive for both identifying and accounting for the kinds of practices required to bring Bayesian reasoning into being.

CHAPTER TEN – CONCLUDING REMARKS

10.1 The Consequences of Paradigm Shift

> 'Legal and scientific forces are converging to drive an emerging skepticism about the claims of the traditional forensic individualization sciences. As a result, these sciences are moving toward a new scientific paradigm.' (Saks and Koehler 2005, p.892)

> 'You don't *learn* maths, you *do* it' (My Sixth Form Pure Maths teacher, 1994)

Forensic science has always occupied an awkward niche, buffered on one side by the interests of law and justice, to whom it must ultimately be answerable, and on the other by the realm of 'science' to which it has always looked to in order to provide credibility and a sense of objectivity. Tensions inherent in attempting to serve both these domains have often resulted in questions being raised about forensic science. Often these questions have been raised in relation to its apparent failings, exposed in cases such as the Birmingham Six, and more recently, Barry George. This sceptical attitude to the claims of certain branches of forensic science has continued in recent years, and in their article, published in *Science* in 2005, Saks and Koehler describe the concerted increase in both legal and scientific challenges to the claims of the so-called 'individualisation sciences'. This, they claim, exposes some serious shortcomings: 'A dispassionate scientist or judge reviewing the current state of the traditional forensic sciences would likely regard their claims as implausible, underresearched and oversold' (Saks and Koehler 2005, p.892).

In fact, Saks and Koehler claim that forensic science is increasingly moving away from the traditional tenet of discernible uniqueness, namely the assumption that the identification of an individual may occur through the presence of supposedly unique marks, such as fingerprints. This so-called 'positivity doctrine' (Broeders 2005), they claim, is giving way to a more conditional stance to identification evidence, in which probabilistic reasoning is the norm, accelerated in part by the advent of DNA profiling, often referred to as 'the gold standard' in forensic evidence (Lynch 2003, Aronson 2005, p.3).

This thesis has sought to address this trend towards the application of probability theory in forensic science, and has focused in particular on the use of Bayesian reasoning. In identifying the application of Bayes as a series of practical, contingent and contextualised performances, I suggest that a form of reasoning intended only for the assessment of hypothesis probabilities, is able to exert a considerable degree of influence on the way in which investigations may proceed, by seemingly shaping the generation of hypotheses as well. In the realm of forensic science at least, this apparently neutral mathematical construct appears to exert a great deal of influence on the behaviours of its supposed users, and leads to a rich diversity of practices.

In guiding the behaviours of forensic scientists, Bayesianism may therefore be considered as an important way in which modern forensic science is constituted and ordered. Moreover, through its ability to guide inquiry, Bayesianism can be regarded as playing an instrumental role in shaping the disciplinary identity of forensic science as a whole. Forensic science has long sought to carve out such an identity for itself. This is readily apparent in the enduring invocation of locutions such as Locard's principle, and the proclamations of other noted figures such as Paul Kirk, and his claim for 'criminalistics' to be the 'science of individualisation'. These arguments highlight a certain need on the part of forensic scientists to lay claim to a set of epistemic principles and practices which can render their practices as distinct from other scientific disciplines, rather than being parasitic on them. Some of the findings in this thesis strongly suggest that Bayesian reasoning may also have a role to play in this regard, as supported by the repeated insistence that Bayes 'is the *only logical* way to reason about evidence'.

10.2 Summary of Findings

The view of Bayes expressed above seems to presume a monolithic set of reasoning practices, which is contradicted by the findings of this thesis. Rather than suggesting a monolithic model for processing into which evidential data can be processed, the results show how the practices which produce Bayesian reasoning actually occur in response to the ontological liminalities which are exposed by its use. Considered in this way, Bayesian theory cannot alone be regarded as providing an all-encompassing procedure for reasoning about evidence. Hence neither can it be considered as a unifying procedure to be prescribed for all forensic scientific conduct, despite the claims made over the association between Bayes and 'best practice' as exemplified in Chapter 8.

This thesis shows how the manner in which Bayes constitutes 'expert' behaviours is organisationally and socially constructed. In demonstrating this, my thesis raises another set of issues that extend beyond the consideration of the immediate concerns of forensic science. Chapter 5 discussed academic debates in legal circles which centred on the use of statistics and probability theory in courtroom proceedings. Whilst objections were raised in the course of these discussions, the advocation of Bayes clearly played a role in introducing this form of reasoning into investigative uses of forensic science. Furthermore, in Chapter 6 I described the unsuccessful experience of attempts to apply a Bayesian model to the defence case in a particular criminal hearing. What these examples highlight is the different attitudes exhibited towards Bayesianism in courtroom and forensic scientific contexts. Despite the efforts of evidence scholarship, Bayes has struggled to find support as a replacement for traditional methods of juridical reasoning; the criticisms of Tribe (1971) appear to have had a lasting impact on the perceived role of Bayesianism a law, a situation significantly compounded by the judgements delivered in *R vs Dennis John Adams*.

10.3 Implications of Research

Whilst Bayes may be regarded as unsuitable for courtroom use, the fact that it does play an increasingly significant role in forensic investigations points to a curious dichotomy of epistemological labour. In the course of this thesis I have aimed to display aspects of an area of juridico-scientific life which has, so far, received relatively little attention, but plays a vital instrumental role in the delivery of justice. Despite their importance, the kinds of practices highlighted herein are more often kept concealed from the more public face of justice visible in courtroom proceedings. Only on these kind of occasions may the techniques used by forensic scientists receive any significant scientific scrutiny via cross-examination, and this is normally restricted to cases where there is a resourceful or particularly inquisitive defence counsel, such as the Omagh trial or the cases handled by Neufeld and Scheck, as described in Chapter 6. Instead forensic evidence is more often likely to be presented in a relatively circumscribed manner, with a reporting 'expert' scientist giving his opinion on the likelihood of a piece of evidence incriminating a defendant. From this however, members of the court are unlikely to gain any comprehensive impression of the reasoning processes the scientist may have used to reach their conclusions about the evidence. As I also reported in Chapter 6, alleged failures in reasoning, have, occasionally, been exposed in court, as exemplified by the case of *R vs Keen*, and *R vs Doheny and Adams*, and, whilst unsuccessful, *R vs Dennis John Adams* represented another notable insight into the reasoning processes of forensic science. Yet even the discussions raised by these cases present a partial view of the work involved in forensic reasoning, as these discussions do not in any way reflect the makeup of the activities involved.

I have hoped that this work has gone some way towards redressing this, although I am of course limited in the scope of examples to which I can devote attention. Whilst hopefully pointing the way towards an area of inquiry which merits further attention, I also wish to raise a further issue with regard to the relationship between the practices identified in this study and courtroom deliberation. The fact that 'expert' behaviours can be identified as contingent social practices, even those which ostensibly might appear to be wholly cognitive in character, must raise some issues for the way in which the concept of scientific admissibility is conceived. This research casts doubts on the US method for assessing scientific evidence, which determines whether the production of such evidence has followed a 'falsifiable' method. Here, I have shown that recourse to a single scientific method is insufficient to determine admissibility, if such a principle were to be followed. Whilst the topic of admissibility may not readily yield any easy answers, this research has shown that the price for high standards of admissibility may be a stringent, case-sensitive approach. This research has shown how subjective assumptions may be built into advanced scientific technologies, and how these may be no more impartial than the old 'case construction' prosecutorial approach.

To return to discussions of disciplinarity, it still clear that, despite claims of unifying principles, forensic science continues to borrow heavily from other scientific disciplines, adapting scientific techniques for the purposes of identification. The field of evidence studies, from which Bayesian forensic scientists have also appropriated ideas, can itself be viewed as an interdisciplinary undertaking, bringing together researchers from the fields of statistics, law, medicine, psychology, and even history, to name but a few.

When addressing the subject matter I too have aimed to adopt an interdisciplinary posture. I have attempted to maintain an open mind on the supposed value of Bayes, keeping a sceptical view towards the supposed logicality of the approach in the spirit of another form of symmetry, this time that proposed by the Strong Programme (Bloor 1976).[1] Wherever possible however, I have attempted to address the scientific issues raised in the course of the case studies on their own terms. By this I mean that I have not sought to approach the subject with a pre-given set of assumptions to be tested, but instead I have aimed to explore the subject as I have found it, but to be informed as much by science as sociological theory.

[1] According to the Strong Programme, the 'symmetry principle' is a key tenet, and refers to the need of the inquirer to commit to a neutral position with regard to the alleged truth or falsity of the scientific practices under study.

In this study I have emphasised the performances of Bayesian reasoning as being emergent phenomena, and it is also true to say that the study itself exhibits much the same property. I did not enter into this research with any specific set of questions in mind; instead these evolved as I gained familiarity with the subject through the investigation of a wide variety of literatures and interviews. I believe that this approach, of entering into the research with few preconceived ideas, has enabled me to discover areas of forensic scientific practice that have hitherto been relatively overlooked by sociological studies, and I hope that this has led to original findings. However, this process has enabled time for reflection, and at this juncture it is appropriate to consider what this research could have additionally achieved, as well as looking toward future possibilities.

10.4 Suggestions for Further Research

To begin with, I believe there is more to be learned about the construction of some of the Bayesian technologies featured in this study, particularly those which rely more on automated forms of Bayesian reasoning, such as LoComatioN, Pendulum and also the Bayes Networks applications which have been developed to address similar forensic issues. Whilst I have been able to question these technologies in terms of some of the assumptions which are incorporated into them, there exists further scope to investigate the precise nature of the means of their construction. In the course of this study I did devote some time in attempting to capture this, as I was struck early on by how the technical literature depicted these technologies in a finished form, without describing the difficulties and false starts that are involved in their construction. For example, Interviewee 2 described the process as akin to an 'apprenticeship' and 'learning a craft'. From this it is possible to draw the conclusion that the construction of such systems is a more gradual process, and one involving a greater degree of trial and error than is suggested by the technical literature. This interviewee also suggested that I use a relevant software package, such as Matlab or HUGIN, in order to gain a first-hand feel for the construction of Bayesian systems. Whilst time prevented me from taking up this suggestion, such an approach could provide a highly informative account of the phenomenological aspects of performing Bayesian reasoning.

Ideally I would have liked to include some form of first-hand account of the construction of the systems which formed the focus of Chapter 7. As I have stated above, whilst concerted attempts were made to make contact with some of the FSS scientists involved in the construction of LoComatioN and PLS, no positive response was forthcoming. Whilst this has ultimately not affected the claims of this study, it is perhaps indicative of a certain reticence to communicate on scientific issues, which seems to be characteristic of that organisation. Unfortunately, no other forensic science provider offers equivalent products, as this would have represented a possible alternative means of gaining insight in this particular area.

Another issue which merits further attention concerns the differences in attitudes towards Bayes across differing jurisdictions. Whilst Bayesian reasoning has found considerable application within the forensic science community in the UK, and certain other jurisdictions in Europe, the attitude towards Bayes amongst the American community has been somewhat more sceptical. Indeed, as Ian Evett has admitted, there exists 'trenchant opposition' to Bayes amongst forensic science practitioners in the US (Aronson unpublished), which plays against the supposed logicality of the approach. The reasons why this might be the case are unclear, but a discussion with Interviewee 2 was instructive. This interviewee's attempts to introduce computerised Bayesian systems for DNA profile interpretation into American forensics laboratories had met with considerable opposition, and he felt that this might be due to the fact that Bayesian technologies are relatively new in the USA (Interviewee 2, 2007). In passing, Interviewee 2 cited 'cultural' reasons for this which he did not specify further, except to talk about the difficulties of introducing this technology into a 'medieval' context (Interviewee 2 2007). This latter comment, seemingly directed at the attitudes of American forensic scientists, seems to indicate a lack of enlightenment on their part. Bayesian methods have been communicated in the US however (Evett and Weir 1998), and it may have been more reflective of the attitude towards the possibility of automation doing a human's job. Whatever the case, this example shows once again how Bayesian forms do not evoke acceptance, and once again they become contested artefacts.

In terms of other research, it may be worthwhile to monitor how future developments, most notably the increasing marketisation of forensic science in the UK may affect the construction of future technologies in forensic science. The production of technologies such as LoComatioN and PLS play an important role in maintaining the FSS's position as a market leader in the field. However, there are concerns that the increased commercialisation of forensic science may not best serve the interests of British justice. Some of the criticisms expressed in the Omagh judgement reflect the alleged lack of transparency with regard to the validation of LCN techniques by the FSS. This opacity is also apparent with regard to DNABoost, where no scientific information about the technique has been placed in the public domain. This did not stop the FSS however from effectively advertising the technology via media channels orchestrated by a PR firm. This example shows a radically different means of convincing audiences of the scientific efficacy of a product, and draws further questions concerning the relationship between market forces, forensic sciences and the technologies which arise.

This study has also shown how a commercial discourse influenced the development of the CAI and the reasoning practices associated with it, and it may be fruitful to consider in more detail the manner in which Bayesianism may become appropriated, and subsequently tailored, to reflect these kinds of influences. As this study also indicated however, commercial pressures also appear to have constituted spaces of opposition to the CAI. Hence the precise manner in which these kind of factors exert influence on the development of Bayesian technologies may be highly complex, and merits further inquiry.

It is likely however, that the increasing commercialisation of forensic science will only serve to hasten the pace at which technological development proceeds in this area. This has been bourne out already by the production of automated DNA interpretation systems by the FSS. Whilst the introduction of

Bayesian reasoning has been accelerated by its application to DNA profiling technologies, interest is growing in a number of other forensic technologies being developed in areas such as computing and linguistics, and it is likely that these could incorporate Bayesian methods. If so, then these would represent highly fruitful and relevant areas in which to apply some of the concepts developed in this thesis. It is also worthwhile to reiterate that Bayes is finding increasing application in areas outside forensic science and I hope that my studies might inform others to undertake other critical studies of the application of Bayes in other fields. Bayesian methods are being developed in a variety of other fields, including economics and medicine, and are of particular interest in the field of AI. Such areas may also provide fruitful sources for furthering understanding in the application of Bayes.

It appears that, given the vehement support it receives in influential circles, Bayes will continue to play a role in forensic science, as it will in other areas of scientific and social life. However, it also seems clear that its use will continue to attract controversy and contestation in equal measure. What this thesis has attempted to achieve is an improved understanding of the practicalities of putting such a mode of reasoning into action. In doing so it is hoped that this research will benefit anyone with an interest in the debate regarding Bayesianism and forensic science, as well as those with an interest in the study of evidence in general, regardless of their personal position on the efficacy of Bayesian reasoning.

BIBLIOGRAPHY

Abraham, J. (1994). "Bias in Science and Medical Knowledge: The Opren Controversy." Sociology **28**: 717-736.

Aldrich, J. (1997). "R.A. Fisher And The Making of Maximum Likelihood 1912-1922." Statistical Science **12**(3): 162-176.

Aldrich, J. (2002). "How Likelihood and Identification Went Bayesian." International Statistical Review **70**(1): 79-98.

Altheide, D. L. (2004). Ethnographic Content Analysis. The Sage Encyclopedia of Social Science Research Methods. M. S. Lewis-Beck, A. Bryman and T. F. Liao. Thousand Oaks, CA, Sage.

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

Aronson, J. D. (2008). "Creating the Network and the Actors: The FBI's Role in the Standardization of Forensic DNA Profiling." BioSocieties **3**: 195-215.

Balding, D. J. (2005). Weight-of-evidence for forensic DNA profiles. Chichester, UK, John Wiley and Sons.

Ball, V. C. (1961). "The Moment of Truth: Probability Theory and Standards of Proof." Vanderbilt Law Review **14**: 807-830.

Bayes, T. (1764). "An Essay Towards Solving A Problem In The Doctrine of Chances." Philosophical Transactions of the Royal Society of London **53**: 370-418.

BBC (2007). Newsnight.

BBC News (2005). Outback handcuffs 'contaminated'

BBC News. London.

BBC News (2006a). Crime-fighting successes of DNA

BBC News Website. London, BBC.

BBC News (2006b). Falconio DNA evidence 'damning'

BBC News Website London, BBC.

BBC News (2006c). Sex attacker jailed indefinitely

BBC News Website London, BBC.
BBC News (2006d). Trial Told of Forensic 'Errors'. BBC News Website. London.
BBC News (2007). Verdict raises DNA evidence doubt
BBC News Website. London.
Brenner, C. (1997). "Whats Wrong With The Exclusion Probability?", from http://www.dna-view.com/exclusn.htm.
Broeders, A. P. A. (2005). Principles of Forensic Identification Science.
Broeders, A. P. A. (2006). "Of Earprints, Fingerprints, Scent Dogs, Cot Deaths, and Cognitive Contamination - A Brief Look at the Present State of Play in the Forensic Arena." Forensic Science International **159**: 148-157.
Bryman, A. (2004). Social Research Methods. Oxford, Oxford University Press.
Buckleton, J. S. and P. Gill (2004). Low Copy Number. Forensic DNA Evidence Interpretation. J. S. Buckleton, C. Triggs and S. J. Walsh. Boca Raton, CRC Press: 275-298.
Budowle, B., D. L. Hobson, et al. (2001). Low Copy Number - Consideration and Caution. Twelfth International Symposium on Human Identification
Washington D.C., Laboratory Division, FBI, Washington D.C.
Butler, J. M. (2006). Introduction to Low Copy Number (LCN) DNA Testing Issues. Mid-Atlantic Association of Forensic Scientists DNA Workshop.
Caddy, B., G. R. Taylor, et al. (2008). A Review of the Science of Low Template DNA Analysis, London, Home Office.
Callon, M. (1998). The Laws of the Markets. Oxford, Blackwell.
Callon, M. (2006). What does it mean to say that economics is performative? . CSI Working Paper. Paris, Centre de Sociologie de l'Innovation Ecole des Mines de Paris
Charmaz, K. (2002). Qualitative Interviewing and Grounded Theory Analysis. Handbook of Interview Research: Context and Method. J. F. Gubrium and J. A. Holstein. Thousand Oaks, Sage.
Clayton, T. (1998). "Analysis and Interpretation of Mixed Forensic Stains Using DNA STR Profiling." Forensic Science International **91**: 55-70.

Clayton, T. and J. S. Buckleton (2004). Mixtures. Forensic DNA Evidence Interpretation. J. S. Buckleton, C. Triggs and S. J. Walsh. Boca Raton, CRC Press.

Cleary, E. W. (1952). "Evidence as a Problem in Communicating." Vanderbilt Law Review **5**: 277-295.

Cohen, L. J. (1977). The Probable and the Provable. Oxford, Clarendon.

Cole, S. A. (1998). "Witnessing Identification: Latent Fingerprinting and Expert Knowledge." Social Studies of Science **28**(5-6): 687-712.

Collins, H. and R. Evans (2002). "The Third Wave of Science Studies: Studies of Expertise and Experience" Social Studies of Science **32**(2): 235-296.

Collins, H. and T. Pinch (1993). The Golem: What Everybody Should Know About Science. Cambridge, UK, Cambridge University Press.

Comte, A. (1974). Cours de philosophie positive. (English) New York, AMS Press.

Cook, R., I. Evett, et al. (1998). "A Hierarchy of Propositions: Deciding Which Level to Address in Casework." Science and Justice **38**(4): 231-239.

Cook, R., I. Evett, et al. (1998). "A Model For Case Assessment And Interpretation." Science and Justice **38**(3): 151-156.

Cook, R., I. Evett, et al. (1999). "Case Pre-Assessment And Review In A Two-Way Transfer Case." Science and Justice **39**(2): 103-111.

Court of Appeal (1996). London. Regina vs Dennis John Adams (1996).

Crown Prosecution Service (2003). Northumbria: Annual Report 2002-2003, Crown Prosecution Service.

Curran, J. M., P. Gill, et al. (2005). "Interpretation of Repeat Measurement DNA Evidence Allowing For Multiple Contributors And Population Substructure." Forensic Science International **148**: 47-53.

Daemmrich, A. (1998). "The Evidence Does Not Speak for Itself: Expert Witnesses and the Organisation of DNA-Typing Companies." Social Studies of Science **28**(5-6): 741-772.

Daston, L. (1994). "How Probabilities Came To Be Objective And Subjective." Historia Mathematica **21**: 330-344.

Davies, C. A. (1999). Reflexive Ethnography: A Guide to Researching Selves and Others. London, Routledge.

Dawid, P. and I. Evett (1997). "Using a Graphical Model to Assist the Evaluation of Complicated Patterns of Evidence." Journal of Forensic Sciences **42**: 226-231.

Dawid, P., J. Mortera, et al. (2006). "Representing And Solving Complex DNA Identification Cases Using Bayesian Networks." International Congress Series **1288**: 484-491.

de Groot, M. (2004). Optimal Statistical Decisions. London, Wiley.

Deleuze, G. and F. Guattari (1983). Anti-Oedipus: Capitalism and Schizophrenia. Minneapolis, Minnesota University Press.

Derksen, L. (2000). "Towards a Sociology of Measurement: The Meaning of Measurement Error in the Case of DNA Profiling." Social Studies of Science **30**(6): 803-845.

Dixon, G. J. (2004). Operation Phoenix: Operational Overview. Presentation.

Doak, S. and D. Assimakopoulos (2007a). "How forensic scientists learn to investigate cases in practice" R&D Management **37**(2): 113-122.

Doak, S. and D. Assimakopoulos (2007b). The Tacit Dimensions of Collaborative Network Traffic. Establishing the Foundations of Collaborative Networks. L. Camarinha-Matos, H. Afsarmanesh, P. Novais and C. Analide. Dordrecht
New York, Springer-Verlag: 425-433.

Dreyfus, H. L., S. E. Dreyfus, et al., Eds. (1986). Mind over machine:The Power of Human Intuition and Expertise in the Era of the Computer. New York, Free.

Dunlap, A. (1931). "Science vs Practical Common Sense in Crime Detection." The American Journal of Police Science **2**(4): 322-327.

Eco, U. and T. Sibeok, Eds. (1983). The Sign of Three: Dupin, Holmes, Peirce. Bloomington, IND, Indiana University Press.

Edmond, G. (2000). "Judicial Representations of Scientific Evidence." Modern Law Review **63**(2): 216-251.

Edwards, A. W. F. (1997). "What Did Fisher Mean By 'Inverse Probability in 1912-1922." Statistical Science **12**(3): 177-184.

Egelund, T. (2003). "Estimating the Number of Contributors to a DNA Profile." International Journal of Legal Medicine **117**: 271-275.

Evett, I. (1984). "A Quantitative Theory for Interpreting Transfer Evidence in Criminal Cases." Journal of the Royal Statistical Society - Series C **33**: 25-32.

Evett, I. (1986). "A Bayesian Approach to the Problem of Interpreting Glass Evidence." Journal of the Forensic Science Society **26**: 3-18.

Evett, I. (1987). "Bayesian Inference and Forensic Science: Problems and Perspectives." The Statistician **36**(2/3): 99-105.

Evett, I. (2005). Forensic inference – is the law a ass? Joint Evidence/Statistics Seminar. Evidence. Inference and Enquiry, . Department of Statistical Science, University College London.

Evett, I., C. Buffery, et al. (1991). "A guide to interpreting single locus profiles of DNA mixtures in forensic cases." Journal of the Forensic Science Society **31**(1): 41-47.

Evett, I., P. E. Cage, et al. (1987). "Evaluation of the Likelihood Ratio for Fibre Transfer Evidence in Criminal Cases." Journal of the Royal Statistical Society - Series C **36**(2): 174-180.

Evett, I., G. Jackson, et al. (2000a). "More On The Hierarchy of Propositions: Exploring The Distinction Between Explanations And Propositions." Science and Justice **40**(1): 3-10.

Evett, I., G. Jackson, et al. (2000b). "The Impact of the Principles of Evidence Interpretation on the Structure and Content of Statements." Science and Justice **40**(4): 233-239.

Evett, I. and Joyce, H. (2005). "Career Story: Consultant Forensic Statistician." Significance: 34-37.

Evett, I. W. and B. S. Weir (1998). Interpreting DNA Evidence - Statistical Genetics for Forensic Scientists. Sunderland, Sinauer Associates Ltd.

Federov, A., S. Roy, et al. (2003). "Mystery of Intron Gain." Genome Research **13**(10): 2236-2241.

Fienberg, S. E. (2003). "When Did Bayesian Inference Become 'Bayesian'?" Bayesian Analysis **1**(1): 1-41.

Findlay, I., A. Taylor, et al. (1997). "DNA Fingerprinting from Single Cells." Nature **389**: 555-556.

Finkelstein, M. O. and W. B. Fairley (1970). "A Bayesian Approach to Identification Evidence." Harvard Law Review **83**(3): 489-517.

Fisher, R. A. (1912). "On The Absolute Criterion For Fitting Frequency Curves." Messenger of Mathematics **41**: 155-160.

Fisher, R. A. (1915). "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." Biometrika **10**: 507-521.

Fisher, R. A. (1921). "On the 'probable error' of a coefficient of correlation deduced from a small sample." Metron **1**: 1-32.

Fisher, R. A. (1922). "On The Mathematical Foundations of Theoretical Statistics." Philosophical Transactions of the Royal Society of London Series A **222**: 309-368.

Fisher, R. A. (1932). Statistical Methods For Research Workers. Edinburgh, Oliver and Boyd.

Forensic Science Service. (2005a). "Forensic Science Service Case Study: Operation Phoenix." Retrieved 27 May 2008, 2008, from <http://www.forensic.gov.uk/forensic_t/inside/news/list_press_release.php?case=25&y>.

Forensic Science Service (2005b). Forensic Science Service Fact Sheet 6: DNA Low Copy Number **2008**.

Forensic Science Service (2006). National DNA Database Annual Report 2005-06.

Forensic Science Service. (2007). "Colin Pitchfork." from http://www.forensic.gov.uk/html/media/case-studies/f-18.html.

Forensic Science Society. (2008). Retrieved 29 October, 2008, from http://www.forensic-science-society.org.uk/.

Foster, J. (1995). "Informal Social Control and Community Crime Prevention." British Journal of Criminology **35**: 563-583.

Fraser, J. (2007). The Application of Forensic Science to Criminal Investigation. Handbook of Criminal Investigation. T. Newburn, T. Williamson and A. Wright. Cullompton, Willan: 381-402.

Fung, W. (1996). "10% or 5% Match Window in DNA Profiling." Forensic Science International **78**(2): 111-118.

Garfinkel, H. (1967). Studies in Ethnomethodology. Englewood Cliffs, N.J., Prentice Hall.

Genewatch (2005). The Police National DNA Database: Balancing Crime Detection, Human Rights and Privacy

Genewatch Reports. Genewatch. Buxton, Genewatch.

Gieryn, T. (1983). "Boundary Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." American Sociological Review **48**: 781-795.

Gieryn, T. (1995). Boundaries of Science. Handbook of Science and Technology Studies. S. Jasanoff, G. E. Markle, J. C. Petersen and T. Pinch. Thousand Oaks, CA, Sage Publications: 393-443.

Gigerenzer, G., Z. Switjtink, et al. (1989). The Empire of Chance: How Probability Changed Science and Everyday Life. Cambridge, Cambridge University Press.

Gilder, J., R. Koppl, et al. (2008). Comments on the Review of Low Copy Number Testing

International Journal of Legal Medicine. **Letter to the Editor**.

Gill, P. (1998). "Intepreting Simple STR Mixtures Using Peak Allele Areas." Forensic Science International **91**: 41-53.

Gill, P. and J. S. Buckleton (2004). Biological Basis for DNA Evidence. Forensic DNA Evidence Interpretation. J. S. Buckleton, C. Triggs and S. J. Walsh. Boca Raton, CRC Press: 1-26.

Gill, P., J. M. Curran, et al. (2006). "A graphical (Bayes net) molecular model of the entire DNA STR process to aid interpretation." International Congress Series **1288**: 468-470.

Gill, P., A. J. Jeffreys, et al. (1985). "Forensic Application of DNA 'Fingerprints'." Nature **318**: 577-579.

Gill, P., A. Kirkham, et al. (2007). "LoComatioN: A software tool for the analysis of low copy number DNA profiles." Forensic Science International **166**: 128-138.

Gill, P., R. Sparkes, et al. (2000). "Report of the European Network of Forensic Science Institutes (ENSFI): Formulation and Testing of Principles To Evaluate STR Multiplexes." Forensic Science International **108**: 1-29.

Gopnik, A., C. Glymour, et al. (2007). "A theory of causal learning in children: Causal maps and Bayes nets " Psychological Review **111**(1): 1-30.

Guilianotti, R. (1995). "Participant Observation and Research into Football Hooliganism: Reflections on the Problems of Entree and Everyday Risks." Sociology of Sport Journal **12**: 1-20.

Hacking, I. (1975). The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference. London, Cambridge University Press.

Hacking, I. (1983). Representing and Intervening: Introductory Topics in the Philosophy of Science. Cambridge, Cambridge University Press.

Halfon, S. (1998). "Collecting, Testing and Convincing: Forensic DNA Experts in the Courts." Social Studies of Science **28**(5-6): 801-828.

Hobbs, D. (1993). Peers, Careers, and Academic Fears: Writing as Field-Work. Interpreting the Field: Accounts of Ethnography. D. Hobbs and T. May. Oxford, Clarendon Press.

Holdaway, S. (1983). Inside the British Police: A Force at Work. Oxford, Blackwell.

Home Office (2005a). 'Cold Cases' Cracked - One Year On (Press Release).

Home Office. (2005b). "Operational Policing." Retrieved 7 November, 2008, from <http://www.denverda.org/DNA_Documents/Familial_DNA/Report%20Mahmood.pdf>, (Originally cached at <http://police.homeoffice.gov.uk/operational-policing/technologyequipment/forensic-science/>).

Home Office (2006). Consultation Document: Regulation and Standard Setting in Forensic Science. H. Office. London, Home Office.

Howie, D. (2002). Interpreting Probability: Controversies and Developments in the Early Twentieth Century. Cambridge UK, Cambridge University Press.

Hurley, P. J. (2000). A Concise Introduction to Logic. Belmont, CA, Wadsworth.

Hutto, D. (2007). The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology Narrative and Understanding Persons. D. Hutto. Cambridge, Cambridge University Press: 43-68.

Innes, M. (2003). Investigating Murder: Detective Work and the Police Response to Investigating Homicide. Oxford, New York, Oxford University Press.

Jackson, G., S. Jones, et al. (2006). "The Nature of Forensic Science Opinion - A Possible Framework to Guide Thinking and Practice in Investigation and in Court Proceedings." Science and Justice **46**(1): 33-44.

Jackson, J. D. (1996). "Analysing The New Evidence Scholarship: Towards A New Conception of Evidence." Oxford Journal of Legal Studies **16**(2): 309-328.

Jamieson, A. and S. Bader (2008). Press Release on Behalf of the Forensic Institute, Glasgow, Regarding the Caddy Review of Low Template DNA

Jasanoff, S. (1998). "The Eye of Everyman: Witnessing DNA in the Simpson Trial." Social Studies of Science **28**(5-6): 713-740.

Jasanoff, S. (2004). States of Knowledge: The Co-Production of Science and Social Order. Routledge, London.

Jeffreys, A., V. Wilson, et al. (1985). "Hypervariable 'Minisatellite' Regions in Human DNA." Nature **314**: 67-73.

Jeffreys, A. J., V. Wilson, et al. (1988). "Amplification of Human Minisatellites by Polymerase Chain Reaction: Toward DNA Fingerprinting of Single Cells " Nucleic Acids Research **314**: 10953-10971.

Jeffreys, H. (1939). Theory of Probability. London, Oxford University Press.

Kadane, J. and D. A. Schum (1996). A Probabilistic Analysis of the Sacco and Vanzetti Evidence. New York, Wiley.

Kaplan, J. (1968). "Decision Theory and the Fact-Finding Process." Stanford Law Review **20**(6): 1065-1092.

Kimpton, C., P. Gill, et al. (1995). "Report on the Second EDNAP Collaborative STR Exercise." Forensic Science International **71**: 137-152.

Kingston, C. R. (1965a). "Applications of Probability Theory in Criminalistics 1." The Journal of Criminal Law, Criminology and Police Science **60**(309): 70-80.

Kingston, C. R. (1965b). "Applications of Probability Theory in Criminalistics, 2." The Journal of Criminal Law, Criminology and Police Science **60**(312): 1028-1034.

Kingston, C. R. (1966). "Probability and Legal Proceedings." Journal of Criminal Law, Criminology, and Police Science **57**(1): 93-98.

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

Koppl, R. (2005). "How to Improve Forensic Science" European Journal of Law and Economics **20**(3): 255-286.

Kuhn, T. (1962). The Structure of Scientific Revolutions. Chicago, University of Chicago Press.

Kvale, S. (1996). InterViews: An Introduction to Qualitative Research Interviewing. Thousand Oaks, CA, Sage.

Lander, E. (1989). "DNA Fingerprinting on Trial." Nature **339**: 501-505.

Laplace, m. d. P. S. (1820). Theorie Analytique des Probabilities. Paris, Mme Ve Courcier.

Latour, B. (1987). Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge, Mass, Harvard University Press.

Latour, B. (1990). Drawing Things Together. Representation in Scientific Practice. M. Lynch and S. Woolgar. Cambridge, Mass, London, MIT Press.

Latour, B. (1999). Pandora's Hope: Essays on the reality of science studies. Cambridge, Mass, Harvard University Press.

Latour, B. and S. Woolgar (1979). Laboratory Life : The Social Construction of Scientific Facts. Los Angeles, CA, Sage.

Law, J. (2007). Actor Network Theory and Material Semiotics Lancaster, Lancaster University.

Law, J. and J. Hassard (1999). Actor Network Theory and After Oxford, Blackwell.

Leiter, B. (1997). "The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make For Good Philosophy of Evidence." Brigham Young University Law Review: 803-820.

Lempert, R. O. (1977). "Modelling Relevance." Michigan Law Review **75**(5/6): 1021-1057.

Lempert, R. O. (1986). "The New Evidence Scholarship: Analysing the Process of Proof." Boston University Law Review **66**(3/4): 439-477.

Lenoir, T. (1994). "Was the Last Turn a Right Turn? The Semiotic Turn and A.J. Greimas." Configurations **2**: 119-136.

Leucari, V. (2005). Analysis of Complex Patterns of Evidence in Legal Cases: Wigmore Charts v Bayesian Networks. London, Department of Statistical Science, University College London.

Lewontin, R. C. and D. C. Hartl (1991). "Population Genetics in Forensic DNA Typing." Science **254**: 1745-1750.

Livingston, E. (1986). The Ethnomethodological Foundations of Mathematics. London, Routledge.

Lofland, J. and L. Lofland (1995). Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. Belmont, CA, Wadsworth.

Lowe, A. L., C. Murray, et al. (2002). "The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces." Forensic Science International **129**(1): 25-34.

Lucas, P. (2001). Bayesian Networks in Medicine: A Model-Based Approach to Medical Decision Making. The EUNITE Workshop on Intelligent Systems in Patient Care, Vienna.

Lynch, M. (1993). Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science.

Lynch, M. (2003). "God's signature: DNA profiling, the new gold standard in forensic science." Endeavour **27**(2): 93-97.

Lynch, M. (2004). Circumscribing Expertise: Membership Categories in Courtroom Testimony. States of Knowledge: The Co-Production of Science and Social Order. S. Jasanoff. London, New York, Routledge: 161-180.
Lynch, M. and D. Bogen (1997). The Spectacle of History: Speech, Text and Memory at the Iran-Contra Hearing Durham, NC, Duke University Press
Lynch, M. and R. McNally (2003). ""Science", "Common sense," and DNA Evidence: A legal controversy about the public understanding of science." Public Understanding of Science **12**: 83-103.
Mackenzie, A. (2003). "Bringing Sequences to Life: How Bioinformatics Corporealises Sequence Data." New Genetics & Society **22**(3): 315-332.
Mackenzie, D. and Y. Millo (2003). "Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange" American Journal of Sociology **109**(1): 107-145.
Magnani, L. (2001). Abduction, Reason, and Science. Processes of Discovery and Explanation New York, Kluwer Academic/Plenum.
Maguire, J. M. (1947). Evidence: Common Sense and Common Law. Chicago, IL, Foundation Press.
McBaine, J. P. (1944). "Burden of Proof: Degrees of Belief." California Law Review **32**(3): 242-268.
McCartney, C. (2006). "The DNA Expansion Programme and Criminal Investigation." British Journal of Criminology **46**: 175-192.
McClure, D. (2007). Focus Group on Scientific and Forensic Evidence in the Courtroom, National Institute of Justice. Rockville, MD, National Institute of Justice.
McCormick, C. T. (1955). Handbook on the Law of Evidence. St Paul, MN, West.
Medialink. (2008). Retrieved 5 January, 2008, from http://www.medialink.tv.
Merleau-Ponty, M. (1962). The Phenomenology of Perception. London, Routledge & Kegan Paul.
Merton, R. C. (1973). Sociology of Science: Theoretical and Empirical Investigations

Chicago, IL, University of Chicago Press.
Morris, N. (2006). "British police forces to trial new DNA evidence procedures." New Criminologist.
Murdoch, L. (2007). Doubt over DNA test in Falconio case The Age. Melbourne.
Nash, C., Ed. (1990). Narrative in culture : the uses of storytelling in the sciences, philosophy, and literature. London, Routledge.
National Research Council (1992). DNA Technology in Forensic Science. Washington D.C., National Academy Press.
National Research Council (1996). The Evaluation of Forensic DNA Evidence. Washington D.C., National Academy Press.
Newton, C. R. and A. Graham (1997). PCR. Oxford, BIOS Scientific.
Nordby, J. J. (2000). Dead Reckoning: The Art of Forensic Detection. Boca Raton, FL, CRC.
Nowotny, H., P. Scott, et al. (2001). Re-Thinking Science: Knowledge and the Public in an Age of Uncertanty. Oxford, Polity.
Nuffield Council on Bioethics (2007). The Forensic Use of Bioinformation: Ethical Issues, London, Nuffield Council on Bioethics.
O'Reilly, K. (2000). The British on the Costa del Sol: Transnational Identities and Local Communities. London, Routledge.
Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA, Morgan Kaufmann.
Petricevic, S. F., J.-A. Bright, et al. (2005). "DNA Profiling of Trace DNA Recovered from Bedding." Forensic Science International **159**(1): 21-26.
Petrilli, S. (2004). "From Pragmatic Philosophy to Behavioural Semiotics: Charles W. Morris after Charles S. Peirce." Semiotica **148**(1-4): 277-315.
Pickering, A. (1994). "After Representation: Science Studies in the Performative Idiom." Proceedings of the Biennial Meeting of the Philosophy of Science Association **2**: 413-419.
Porter, T. (1986). The Rise Of Statistical Thinking 1820-1900. Princeton, NJ, Princeton University Press.

Punch, M. (1979). Policing the Inner City: A Study of Amsterdam's Warmoesstraat. London, Macmillan.

Rawling, P. (1999). "Reasonable Doubt and the Presumption of Innocence." Topoi **18**: 117-126.

Ream, W. and K. Field (1999). Molecular Biology Techniques: An Intensive Laboratory Course. San Diego, CA, London, UK, Academic Press.

Redmayne, M. (1996). "Science, Evidence and Logic." The Modern Law Review **59**: 747-760.

Redmayne, M. (2001). Expert Witness and Criminal Justice. London, Oxford University Press.

Redmayne, M. (2002). "Appeals to Reason." The Modern Law Review **65**(1): 19-35.

Redmayne, M. (2003). "Rationality, Naturalism, and Evidence Law." Michigan State Law Review: 849-884.

Rubinstein, J. (1973). City Police. New York, Ballantine.

Saks, M. J. and J. J. Koehler (2005). "The Coming Paradigm Shift in Forensic Science." Science **309**(5736): 892-895.

Sample, I. (2006). DNA advances lengthen long arm of the law. The Guardian. London.

Savage, L. J. (1972). The Foundations Of Statistics. New York, Dover Publications.

Schafer, G. (1976). A Mathematical Theory of Evidence. Princeton, Princeton University Press.

Schum, D. A. (2000). "Species of Abductive Reasoning in Fact Investigation in Law." Cardozo Law Review **22**: 1645-1681.

Schum, D. A. (2005). A Wigmorean Interpretation of the Evaluation of a Complicated Pattern of Evidence. University College London Studies of Evidence Science. London, University College London.

Selinger, E. and R. P. Crease (2006). Dreyfus On Expertise: The Limits of Phenomenological Analysis. The Philosophy of Expertise. E. Selinger and R. P. Crease. New York, Columbia University Press: 213-245.

Selinger, E. and R. P. Crease, Eds. (2006). The philosophy of expertise. New York, Columbia University Press.

Shen, Q., J. Keppens, et al. (2007). "A Scenario-Driven Support System for Serious Crime Investigation." Law Probability and Risk **Article In Press**.

Silverman, D. (1993). Interpreting Qualitative Data: Methods for Analysing Qualitative Data. London, Sage.

Smith, R. and B. Wynne (1989). Expert Evidence:Interpreting Science in the Law. London, Routledge.

Soper, H. E., W. A. Young, et al. (1917). "A cooperative study. On the distribution of the correlation coeffiecient in small samples. Appendix to the papers of 'Student' and R.A. Fisher." Biometrika **11**: 328-413.

Star, S. L. and J. R. Griesemer (1989). "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." Social Studies of Science **19**: 387-420.

Stigler, S. M. (1982). "Thomas Bayes' Bayesian Influence." Journal of the Royal Statistical Society - Series A (General) **145**(2): 250-258.

Stigler, S. M. (1999). Statistics On The Table: The History Of Statistical Concepts And Methods. Cambridge, Mass, Harvard University Press.

Taroni, F., C. Aitken, et al. (2006). Bayesian Networks and Probabilistic Inference in Forensic Science. Chichester, John Wiley and Sons

Tenenbaum, J. B. and T. L. Griffiths (2006). "Theory-based Bayesian models of inductive learning and reasoning" Trends in Cognitive Sciences **10**(7): 309-318.

The Economist (2000). Reviving Bayesian Statistics. The Economist. London.

The Economist (2006). Bayes Rules. The Economist. London.

The Journal (2007). The new key to murder riddles, The Journal. Newcastle.

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" Law and Human Behaviour **11**(3): 167-187.

Tillers, P. and D. A. Schum (1988). "Charting New Territory in Judicial Proof: Beyond Wigmore." Cardozo Law Review **24**(4): 931-1012.

Torres, Y. (2003). "DNA Mixtures in Forensic Casework: A 4-Year Retrospective Study." Forensic Science International **134**: 180-186.

Tribe, L. (1971). "Trial By Mathematics: Precision and Ritual in the Legal Process." Harvard Law Review **84**(6): 1329-1393.

Twining, W. (1984). "Taking Facts Seriously." Journal of Legal Education **34**: 22-42.

Walton, D. (2004). Abductive Reasoning. Tuscaloosa, AL, University of Alabama Press

Weir, J. (2007). The Queen vs Sean Hoey. The Crown Court Sitting in Northern Ireland.

Wight, C. (1999). "They Shoot Dead Horses Don't They? Locating Agency in the Agent-Structure Problematique." European Journal of International Studies **5**: 205-222.

Wigmore, J. H. (1913). "The Problem of Proof." Illinois Law Review **8**(2): 77-103.

Williams, R. (2007). The Problem of Dust: Forensic Investigation as Practical Action, Orders of Ordinary Action. D. Francis and S. Hester. London, Ashgate.

Williams, R., P. Johnson, et al. (2004). Genetic Information and Crime Investigation: Social, Ethical and Public Policy Aspects of the Establishment, Expansion and Police Use of the National DNA Database Wellcome Trust Report, Wellcome Trust.

Willis, P. (1977). Learning to Labour. Farnborough, Saxon House.

Wittgenstein, L. (1958). The Blue and Brown Books. New York, Harper and Row.

Xu, F. and J. B. Tenenbaum (2007). "Word Learning as Bayesian Inference." Psychological Review **114**(2): 245-272.

Zabell, S. (1989). "R.A. Fisher On The History of Inverse Probability." Statistical Science **4**(4): 247-256.