

**A Legal, Scientific and Phenomenological  
Enquiry into the Reliability of Bitemark Analysis**

**Mark Andrew Page**

BDS<sub>Sc</sub>(Hons), GradDipClinDent, GradCertTeach

A thesis submitted in fulfilment of the requirements for the degree of  
**Doctor of Philosophy**

School of Health Sciences  
Faculty of Health  
University of Newcastle, Australia

March 2012

## **Statement of Originality**

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

.....

Mark Andrew Page

Date:

## **Acknowledgements**

This thesis would not have been possible without the support of many people, and I must firstly thank my supervisors, Associate Professor Jane Taylor and Dr Matt Blenkin, whose encouragement and guidance have made this journey all the more easier.

I would also like to thank those people who agreed to spend time discussing various aspects of this work and their significance, both on and off the record. I acknowledge their time and patience in educating someone who must have seemed no more than a rank amateur in the finer points of their own complex academic realms. In particular, I would like to thank Associate Professor Simon Cole, Judge Stephanie Domitrovich, Robert Epstein, Professor David Faigman, Professor Ian Freckelton SC, Associate Professor Gary Edmond, Justice Michael Kirby AC CMG, Judge Barbara Rothstein, and Professor Michael Saks in this regard. Thanks also go to Dr Mike Bowers and Dr David Sweet for the time they spent discussing various aspects of this research with me.

I must also thank each of the Australian odontologists who participated in this research, for their time but also their frank, open and honest discourse during interviews. Without their participation, this project would certainly not have been possible. Thanks also go to the directors of those odontology centres who were approached regarding access to their bitemark case records, for their overt willingness to provide data in this manner.

Finally, my heartfelt thanks go to Mark, who never seemed to doubt that I could do it.

*To the innocent*



## Abstract

Bitemark analysis has anecdotally been used for centuries to identify the perpetrators of violent crimes. Despite its long-standing use in courts across the Western world, the advent of new technologies in forensic science such as DNA analysis has drawn to light numerous examples where innocent men have been convicted of such crimes on the basis of bitemark interpretation by one or more experts. The legal sphere saw the introduction of arguably the most restrictive limitations on expert testimony in the United States in 1993, via the US Supreme Court case *Daubert v Merrell Dow Pharmaceuticals Inc.* Yet the extent to which this precedent is adhered to, which calls for assessment of a discipline's *reliability* prior to admission as evidence, even in the United States appears minimal. Judges are generally loath to exclude most long-standing forensic identification techniques, including that of bitemark analysis, despite criticism from numerous fronts that many if not all of these disciplines fail to meet any standards articulated by the *Daubert* precedent.

Despite the urging of legal and scientific academics, and despite the obvious similarity to those forensic techniques used in the United States, Australian courts have categorically rejected the relevance of *Daubert* to expert evidence in this country. The liberal admission of expert testimony appears to be a most jealously guarded facet of our legal system. Australia does not appear likely to adopt precedents that in any way give judges the authority to exclude expert testimony on the grounds of a failure to meet a reliability threshold, particularly when the witness currently meets the requirements for admission as an expert under the relevant sections of the Evidence Act.

Yet the answer to whether bitemark analysis is justified, as both a science and tool by which the courts can use for forensic identification purposes is not as straightforward as it seems. The literature supporting the ability of individuals to claim identity from marks made by teeth on human skin is very weak, and combined with a growing history of wrongful convictions from analysis in this manner seems to clearly suggest that the answer to this question is 'no'. Yet there is also need to assess whether these wrongful conviction cases are indeed reflective of the practice that most odontologists necessarily engage in this country. Anecdotal evidence suggests that these forms of conclusions are comparatively rare in Australia, and retrospective analysis of casework supports the notion that only a relatively small percentage of such cases end in conclusions regarding individual identity.

Bitemarks potentially reveal more information regarding the nature of the perpetrator than simply identity, and so their interpretation still plays a useful role in forensic evidence

investigations even if not used for that ultimate purpose. Yet even when interpreting injuries without specific regard to identification of the perpetrator, odontologists in Australia have demonstrably drawn conclusions that remain unsupported by any evidence that they are indeed justified. This is perhaps partly due to the lack of objective standards by which bitemarks are assessed, and is further fuelled by the liberal acceptance of expert opinion in Australian courts.

Yet expert opinion can be no substitute for logical conclusion. Patterns of admission of expert evidence, at least in Australia, are unlikely to change in the near future, and so we are left with little option but to try and modify the conduct of bitemark analysis so that it remains within the bounds of credible science. Despite the legal fraternity being reluctant to restrict the scope of forensic identification science testimony, including that of bitemark analysis, the odontology profession cannot remain blind to the fact that there are severe flaws in the practice of bitemark analysis and its subsequent interpretation that need to be addressed. Many of the problems that lead to the inherent ‘unreliability’ of bitemark analysis can only be addressed by long-term research projects, and so defining these boundaries of credibility is of prime importance at the current time. Odontology is best served in the immediate term by recognising the limitations associated with the practice of bitemark analysis as a united professional body, before other agencies make potentially damning decisions that we will have little influence over, and which may ultimately lead to the demise of our role as forensic investigators.



## Table of Contents

<b>Introduction .....</b>	<b>1</b>
Consequences of Invalid Forensic Science .....	4
Forensic Odontology .....	6
Problem Statement.....	7
<b>1. Research Methodology.....</b>	<b>8</b>
Researching ‘Reliability’ .....	8
Architecture.....	10
Legal Enquiry .....	10
Scientific Enquiry .....	10
Phenomenological Enquiry .....	11
Phenomenological Analysis – Grounded Theory.....	12
Phenomenological Data Collection.....	15
Expert Interviews .....	15
Retrospective Casework Analysis.....	20
<b>2. The Admissibility of Expert Evidence: Frye, Daubert and the US Federal Rules of Evidence .....</b>	<b>23</b>
General Acceptance and the <i>Frye</i> Standard .....	24
The US Federal Rules of Evidence.....	26
<i>Daubert v Merrell Dow Pharmaceuticals Inc.</i> ....	27
Responses to <i>Daubert</i> .....	29
Relevancy, Review and <i>Joiner</i> .....	33
Scope of Applicability and <i>Kumho Tire</i> .....	35
Expert Evidence in the United States Today .....	37
<b>3 Forensic Identification Evidence Since <i>Daubert</i>.....</b>	<b>43</b>
The Admission of Forensic Identification Evidence .....	45
Empirical Data on the Admission <i>versus</i> Exclusion of Forensic Identification Science Evidence .....	51
The Exclusion of Forensic Identification Evidence.....	53
Methodology.....	54
Results.....	55
Quantitative Analysis.....	58
Qualitative Analysis: Reasons for Failing to Meet the Reliability Threshold .....	61
Discussion.....	71
Conclusion .....	76
<b>4. Australian Evidence Law and the Admissibility of Scientific Expert Evidence.....</b>	<b>78</b>
Australian Evidence Law .....	78
The Admission of Scientific Evidence in Australia .....	82
Commentary on Expert Scientific Evidence in Australia.....	88
The <i>Daubert</i> Trilogy in Australian Law.....	92
Conclusion .....	100
<b>5. Uniqueness and Individualisation in Forensic Science .....</b>	<b>102</b>
The Uniqueness Fallacy.....	102
The Ideological Origins of the Uniqueness Principle .....	104
The Evidence for Uniqueness .....	106
Anecdotal Evidence and Experience .....	106

Knowledge about the process of formation. ....	108
Studies.....	109
Quantifying Uniqueness: The problem of numbers.....	119
<b>The (II)logic of the Uniqueness Literature.....</b>	<b>123</b>
<b>The Irrelevance of Uniqueness .....</b>	<b>127</b>
<b>Conclusion .....</b>	<b>132</b>
<b>6. A Critical and Contextual Evaluation of Bitemark Analysis .....</b>	<b>134</b>
Literature Concerning the Collection and Analysis of Bitemark Evidence.....	140
Interpretation of a Suspected Bitemark Injury .....	145
Effect of Quality of Injury on Assessment.....	148
Reporting – Agreement in Language and Semantics .....	150
Error and Accuracy Rates in Bitemark Analysis.....	158
Conclusion .....	167
<b>7. Bitemark Analysis in Contemporary Australia .....</b>	<b>168</b>
<b>Results.....</b>	<b>169</b>
Experience in Forensic Odontology.....	169
Post-graduate qualifications in Forensic Odontology (at time of interview) .....	170
Courtroom Experience in Presenting Bitemark Evidence .....	170
Case Type .....	171
Quality.....	172
Method of Analysis of Bitemarks .....	173
Consistency of Opinions .....	175
Assessment of the Forensic Value of Bitemarks .....	188
Agreement on Origin of Injury .....	190
Reporting – Conclusionary Language and Semantics.....	191
<b>Discussion.....</b>	<b>194</b>
Experience and Qualifications.....	194
Quality.....	195
Methods of Analysis.....	196
Odontologist Agreement.....	198
Severity and Significance Scoring: ‘Forensic Value’ Analysis.....	203
Bitemark Reporting .....	205
<b>Conclusion .....</b>	<b>209</b>
<b>8. Contextual Effects and Observer Bias In Forensic Odontology .....</b>	<b>210</b>
<b>The Bias Blind Spot .....</b>	<b>211</b>
<b>Context Effects .....</b>	<b>215</b>
Emotional effects .....	215
Confirmation bias.....	215
Target shifting .....	216
<b>Other Sources of Cognitive Bias .....</b>	<b>217</b>
Observer Effects.....	217
The Contrast Effect.....	218
The Overconfidence Effect.....	221
<b>Minimising Cognitive Bias .....</b>	<b>223</b>
<b>Conclusion .....</b>	<b>229</b>
<b>9. Conclusion.....</b>	<b>230</b>
<b>Is <i>Daubert</i> Relevant?.....</b>	<b>230</b>
<b>The Scientific Status of Bitemark Analysis .....</b>	<b>232</b>

<b>Bitemark Analysis – Contemporary Australian Practice .....</b>	<b>233</b>
Nature of casework .....	234
Quality of Evidence .....	234
Interpretation of Bitemark Injuries .....	235
Method of Analysis .....	236
Standard of Reporting .....	237
Language .....	238
Potential for Bias .....	240
<b>The Future .....</b>	<b>241</b>
<b>References .....</b>	<b>243</b>
<b>Appendix: Guidelines for Bitemark Analysis .....</b>	<b>268</b>
<b>Publications:</b>	
1. Page, M., Taylor, J.A. & Blenkin, M. (2011) Forensic Identification Evidence Since Daubert Part I - A Quantitative Analysis of the Exclusion of Forensic Identification Science Evidence. <i>J Forensic Sci</i> , 56(5), 1180-1184	
2. Page, M., Taylor, J.A. & Blenkin, M. (2011) Forensic Identification Evidence Since Daubert Part II - Judicial Reasoning in Decisions to Exclude Forensic Identification Science Evidence on Grounds of Reliability. <i>J Forensic Sci</i> , 56(4), 913-917.	
3. Page, M., Taylor, J.A. & Blenkin, M. (2011) Uniqueness in the Forensic Identification Sciences - Fact or Fiction? <i>Forensic Science International</i> , 206, 12-8	
4. Page, M., Taylor, J.A. & Blenkin, M. (2012) Reality Bites — A ten-year retrospective analysis of bitemark casework in Australia. <i>Forensic Sci Int</i> , [In Press], Available online [Early view].	
5. Page, M., Taylor, J.A. & Blenkin, M. (2012) Context Effects and Observer Bias - Implications for Forensic Odontology. <i>J Forensic Sci</i> , 57(1) 108-112	
6. Page, M., Taylor, J.A. & Blenkin, M. Expert Interpretation of Bitemark Injuries – A Contemporary Qualitative Study [Accepted for publication, April 2012]. <i>J Forensic Sci</i> .	

# Introduction

---

The practice of forensic science has come under increasing scrutiny in the last twenty years, largely due to a combination of seminal court rulings and major advances in science. Forensic science encompasses a variety of sub-specialty practices, including analysis of fingerprint comparison, hair and fibre comparison, DNA analysis, firearms and toolmarks analysis, handwriting analysis and forensic odontology to name but a few. Others such as forensic pathology, anthropology, chemistry and toxicology can also be considered ‘forensic science’, and all of these share commonality in that they seek to aid justice in our society through application of their particular theories and techniques to answer questions posed by the justice system. This places these disciplines in a unique position at the intersection of two great doctrines; science and law.

Several areas of forensic science thought for many years to be of sound basis have now been criticised as being based on false assumptions, poor science, inaccurate techniques, and erroneous interpretations. In particular, these criticisms have been levelled squarely at the so-called ‘identification’ forensic sciences, which includes fingerprint analysis, firearms and toolmark analysis, and forensic odontology. Unfortunately, some of these criticisms have turned out to be well founded. It has now become widely recognised that bullet lead analysis was flawed science from the beginning, with irrefutable proof that its fundamental scientific hypothesis that the chemical composition of lead impurities was unique to particular manufactured batches was erroneous (National Research Council, 2004; Ungvarsky, 2006). The FBI has now discontinued the use of bullet lead analysis due to these revelations, however, there have still been countless convictions based on this flawed technique. The science of arson investigation, based on analysis of fire-consumed property in order to determine the origin of the fire, its behaviour, and whether it had been deliberately lit, has also very recently come under severe criticism. When experiments were conducted in the early 1990’s seeking to verify the basic assumptions held about the detection of arson, the results disproved many of the commonly held beliefs about the behaviour of fires (Russell, 2006).

Unfortunately, many of these results have only come to light via unfavourable media exposure, and most disturbingly, events such as the execution of a man who, after the conduct of several arson experiments attempting to prove its validity, was revealed to be innocent (Grann, 2009). The rise of DNA as a forensic identification tool has also shed

unfavourable light on other areas of forensic science. The use of DNA to re-examine evidence in past cases via *The Innocence Project* (2009)<sup>1</sup> has resulted in over two hundred exonerations so far, with many of the erroneous original convictions made on the basis of evidence given by forensic scientists, including forensic odontologists.

In 2009, the United States Congress directed the National Academy of Sciences (NAS) to prepare a report on the status of forensic science in the United States. Many witnesses called to testify before the United States Senate Committee on the Judiciary provided damning evidence against the discipline of forensic science. As one witness stated<sup>2</sup>, forensic science represents ‘a system plagued by a paucity of good research, fragmentation, inconsistent practices and weak governance’ (Edwards, 2009). The subsequent report, released as a publication entitled *Strengthening Forensic Science in the United States: A Path Forward* (National Research Council, 2009)<sup>3</sup> noted that ‘the forensic science system, encompassing both research and practice, has serious problems that can only be addressed by a national commitment to overhaul the current structure that supports the forensic science community.’ The report has made thirteen specific recommendations in order to address these issues. While some of these issues relate to funding, and to the particular institutional and organisational problems relevant to the judicial and governmental systems in the United States, there are nonetheless observations and recommendations that would apply to the theory and practice of forensic science in any country.

Problems relating to the lack of peer-reviewed studies establishing the scientific basis for a particular discipline, the need for research in order to establish the limits of practitioner performance, the impact of observer error, variability and bias, the need for standardised terminology, and the lack of mandatory standardisation, certification, and accreditation are all criticisms that have been levelled at the field of forensic odontology, as well as the forensic sciences in general. The report also notes that ‘the adversarial process is not suited to the task of finding scientific truth’, noting that lone judges without the benefit of jurors or colleagues have little time for personal research.<sup>4</sup> This stands in stark contradiction to a long and widely held belief in both Australia and the United States that examination and cross-

---

<sup>1</sup> The *Innocence Project* is a non-profit legal organisation dedicated to exonerating potentially wrongly convicted people via the use of DNA testing.

<sup>2</sup> The Honourable Harry T. Edwards, a Senior Circuit Judge and Chief Judge Emeritus of the U.S. Court of Appeals for the D.C. Circuit, and Visiting Professor of Law at the New York University School of Law.

<sup>3</sup> Hereafter referred to as the NAS Report.

<sup>4</sup> Another reason the law is unsuited to finding scientific truth is because its review processes, i.e. the appellate court system, is highly deferential to the decisions of trial judges. Appellate courts in the United States may only undertake reviews of trial judges decisions to admit evidence in very particular circumstances. This is discussed in more detail in chapter 2 in the section on *Relevancy, Review and Joiner*.



examination in court are effective tools against improper or unreliable evidence (McCormick, 1954; Cross on Evidence 2008). Importantly, the NAS Report places the ball squarely in the forensic disciplines' court to help resolve these issues, noting that 'judicial review, by itself, will not cure the infirmities of the forensic science community'.

It is not merely the 'fringe' areas of forensic science that fail in this respect. Many of the traditional and formerly unquestioned areas of forensic science, such as fingerprints, have come under criticism for their lack of scientific underpinning (Begley, 2004; Benedict, 2004; Epstein, 2001; Giannelli, 2002). Forensic science in Australia has not been immune to this criticism, with the recognition of at least eleven critical areas predicted to soon impact on traditional forensic practitioners, including that of the 'coming paradigm shift' induced by the *Daubert* decision in the United States (Raymond, 2006).

*Daubert v Merrell Dow Pharmaceuticals Inc.* (1993) was a United States (US) Supreme Court case involving scientific evidence in a civil toxic tort, however, the resulting Supreme Court decision is technically applicable to all forms of expert evidence in all case contexts, including forensic science in the criminal sphere. Yet the extent to which the courts accurately apply the *Daubert* decision to forensic science evidence appears to be highly variable. Saks (2002) has criticised the judiciary for their lack of rigor in applying the *Daubert* standard to forensic science, noting that they appear to favour 'public relations at the neglect of empirical testing'. He is but one of many authors who have levelled criticism at both the judiciary and the forensic science community.

Given the judiciary's own reluctance to deny long-standing forensic techniques their current status, undoubtedly due to intense political, social and peer pressures (among others noted by Harris, 2008), it is clear that the law by itself has not been able to rectify the poor scientific standing of some forensic sciences. Other authors have proposed law reform in order to allow defendants wrongly convicted certain avenues of appeal, (Gable and Wilkinson, 2007) however, law reform in general, let alone that suggested by legal academics in the forum of a legal journal, takes many years to evolve. Even proponents of such dramatic solutions also recognise that it is the shortcomings of the sciences themselves that ultimately need to be addressed. Practitioners and professors of such 'sciences' will need to take the ultimate responsibility in order to ensure the scientific basis of these disciplines.

## Consequences of Invalid Forensic Science

Garret and Neufeld (2009) examined a series of wrongful convictions achieved through erroneous forensic testimony, where the truth came to light only after application of modern DNA analysis techniques to the original evidence. They cite figures of up to 60% of wrongful convictions due to invalid testimony by forensic scientists. Most of these wrongful convictions were based on serology and hair comparison results, neither of which play a large role in forensic identification today, however, bite marks represented 6 of the 137 cases cited, of which four cases were found to involve erroneous conclusions by the forensic odontologist. Fingerprints represented 13 of the cases cited, with one case involving invalid testimony. They noted that due to the limitations of their study, which included cases from over twenty years ago (where technology, not to mention the rules of evidence were significantly different to those of today), these wrongful convictions cited might only be the tip of a much larger iceberg. Saks and Faigman (2008) also claimed that a leading cause of erroneous convictions in the United States is flawed forensic science, however this analysis itself has been considered flawed by other authors (See *Crime Lab Report*, 2009).

The analysis of the ‘causes’ of wrongful conviction is difficult to refine. As an example, the conviction of Ray Krone involved testimony by a dentist who stated that the bite mark on the victim ‘matched’ that of the defendant<sup>5</sup>. While this may seem to be purely a failure of the science of forensic odontology, the prosecutor had already approached another dentist, Dr John Parks, who had denied that the bite mark matched the dentition of Mr Krone. He, in turn, also consulted with Dr Steven Sperber, a well-known forensic odontologist, who agreed with his findings. Unsatisfied with this, the prosecutor approached a third odontologist, who testified that the bite mark was indeed Krone’s. The case went forward and Krone was convicted, despite a lack of any other physical evidence. Is this a failure of forensic science, or is this prosecutorial misconduct, expert witness misconduct, or simply poor expert witness selection? Saks and Koehler (2005), using data from the Innocence Project, noted that in their analysis of the factors leading to wrongful convictions, the percentages assigned added up to more than 100 because more than one factor was found in each case. Rudin and Inman (2006) have criticised the data analysis methods used by Saks and Koehler in representing errors associated with wrongful convictions, comparing it to a similar chart on the Innocence Project’s website, but noting differences in the categorisations of ‘sources of error’ that have potentially lead to different interpretations.

---

<sup>5</sup> Ray Krone was found guilty and initially sentenced to death. He spent ten years in prison, including two years on death row before being exonerated by DNA testing in 2002.

Gershman (2003) noted that in general, the frequency and nature of prosecutorial misuses of scientific evidence have not yet been studied in any detail, despite their ability to shape the trier of fact's evaluation of the evidence. Data from the source used by Saks and Koehler (2005) was later analysed and the results published in more detail by Garrett (2008). Collins and Jarvis (2009) have refuted *both* of these studies' conclusions, and claim that a far lesser percentage of wrongful convictions has been due to 'forensic science malpractice'.<sup>6</sup> The variability in analysis and classification of what is essentially the same set of data demonstrates both the difficulty of ascertaining the true leading causes of erroneous convictions, as well as the subjectivity involved in deriving these numerical representations.

In England, the House of Commons Science and Technology Committee report *Forensic Science on Trial* suggests that the number of miscarriages of justice due to flawed expert evidence is 'unlikely to be high, and the legal system as it stands should enable miscarriages in serious cases to come to light eventually.' (House of Commons Science and Technology Committee, 2005). Despite this, the committee articulates that steps could still be taken to reduce the potential for such miscarriages to occur. Several prominent cases in Australia have resulted in criticisms of the lack of standardised procedures in forensic analysis that have lead to miscarriages of justice (Wilson, 1994). The American system has been accused of the same failings (Giannelli, 2003), most recently in a government-sponsored inquiry conducted by the National Research Council that culminated in the NAS report (National Research Council, 2009).

Another, perhaps less considered consequence of poor forensic science is the cost to society of lawsuits based on patently flawed scientific reasoning. David Faigman, in defence of the *Daubert* decision, characterised this by using the example of Bendectin, the drug in issue in the *Daubert* case, which was pulled from sale as the litigation progressed.

*'If you allow ten thousand lawsuits to go forward, almost any defendant is going to go bankrupt. Not because they lost the law suits, but simply because the plaintiff got beyond summary judgement and the case went to trial....whatever the merits of causation. Pharmaceutical companies do create drugs that are good for society, and it's not such a good thing if they go out of business or pull it off the market. Bendectin is a good example of that ... there is no epidemiologist today that thinks it caused birth defects, yet it was taken off the market. Perhaps women, because they didn't have access to Bendectin, might end up*

---

<sup>6</sup> Although Collins and Jarvis' earlier attempt at this article, published on the *Crime Lab Report* website in 2008, received scathing criticism in the *News of the California Association of Criminalists*, where the authors noted several fundamental flaws with the re-interpretation of original *Innocence Project* data and described Collins and Jarvis' work as 'unnecessarily provocative' (Rudin and Inman, 2008).

*causing themselves more harm by looking for other drugs to help the severe morning sickness they were suffering'* (Faigman, 2009).

While this case refers to causation and civil litigation, it is easy to draw parallels with forensic science in the criminal context<sup>7</sup>. The cost to society in criminal cases involving flawed forensic science include the issues of innocent members of the community being persecuted and ultimately incarcerated, in addition to guilty persons remaining free, not to mention the financial costs associated with these trials. Faigman also noted that these sorts of cases invariably 'tie up' the already over-worked legal system, allowing other, perhaps more important, cases to go unresolved for longer.

## Forensic Odontology

The field of forensic odontology has been singled out by several authors as being one of these forensic science disciplines that fails to meet even the most basic of scientific criteria, in particular the sub discipline of bitemark analysis. Most textbooks describe the theory of bitemark analysis as being a three-pronged tenet, in that:

- a) the dentition is unique;
- b) a pattern associated with the unique spatial arrangement of anterior teeth is faithfully represented on a medium when bitten; and
- c) a forensic odontologist can accurately match this pattern on medium made by these teeth to a suspect's dentition, thus inferring identity.

A similar theory is propounded for most pattern-analysis 'identification sciences', including fingerprint analysis, firearms and toolmarks analysis and handwriting analysis. Unfortunately, there are some concerns with these theoretical assumptions, and emerging research data along with historical precedent are suggestive that none of these fundamental concepts hold.

Bitemark analysis has been the subject of criticism from numerous sources, the most high profile being that of the recent NAS Report. The report noted that the field of bitemark analysis has no supporting studies to suggest that bitemarks can contain sufficient detail for positive identification of suspects; there is no agreement between odontologists on standards for comparison or analysis; and that experts diverge widely in their evaluations of the same evidence (National Research Council, 2009). These concerns should indicate that there are potentially severe limitations in using bitemark analysis as a forensic identification tool.

---

<sup>7</sup> Not forgetting that civil cases can also depend on forensic identification science testimony.

Yet the NAS report only summarises criticisms that have already been levelled at bitemark analysis for several years (See Bowers, 2001; Beecher-Monas, 2008; Deitch, 2009; Fisher, 2008; Giannelli, 2007a; Saks and Faigman, 2008; Saks, 1995, and 2009b). Fuelled by the controversy surrounding high-profile cases where defendants were wrongfully convicted on the basis of bitemark evidence such as Ray Krone (1992), Calvin Washington (1997), James O'Donnell (1997), Dan Young (1993) and Willie Jackson (1989), the legal system has also already started to question the validity of bitemark analysis, particularly since the introduction of DNA techniques that are unquestionably grounded in valid scientific theory, thanks to the plethora of research that has been conducted in this area.

## **Problem Statement**

Bitemark analysis is now the subject of criticism from a new group of contemporaneous agencies such as the National Academy of Sciences. But perhaps more concerning still is that these critics appear to be ultimately backed by the frightening form of direction from no less than the US Supreme Court, via the *Daubert* precedent. Yet before embarking on potentially misdirected research, overt criticism, or the abandonment of bitemark analysis altogether, there is a need to understand firstly how the legal precedent that gave rise to this criticism is interpreted and applied by the legal system *day-to-day*. Secondly, there is a need to understand how bitemark analysis is conducted in Australia, in order to verify that criticisms raised largely by those outside the odontology profession are valid. Furthermore, it is vitally important that both of these aspects are explored without being viewed through the prism of a purely academic analysis. Once the real-time mechanics of the application and interpretation of *Daubert* and other expert evidence legislation, as well as the practice of bitemark analysis itself are understood, the discipline of forensic odontology in Australia will then have a more robust concept of what this criticism truly means, and whether it is applicable to its practice in this country. It can then begin to more realistically shape its own internal reviews in order to firstly address the concerns of its principle client – the law – in addition to those of its critics, so as to ensure its longevity as a forensic investigative tool. This thesis marks the beginning of this process by presenting a framed enquiry of the reliability of bitemark analysis from legal, scientific and phenomenological perspectives.

## CHAPTER 1

# Research Methodology

---

## Researching ‘Reliability’

The principle issue with forensic evidence that is being challenged in US courts today is that of its ‘reliability’. The controversy surrounding DNA analysis in the United States began with a discussion of its reliability and general acceptance in *Spencer v Commonwealth* (1989), and Harris (2008) noted that the decision to admit DNA evidence only occurred once the courts were satisfied that; the theory of DNA was valid and reliable; techniques and methodology used in analysis of the DNA sample were also reliable, and; that the statistics offered accurately represented the true measurement of occurrences in the population.

Yet the concept of ‘reliability’ has been the source of some confusion since the US Supreme Court’s construction of the term in *Daubert v Merrell Dow Pharmaceuticals Inc.* (1993). Justice Blackmun then used the term ‘reliable’ in order to distinguish science based on supporting facts, data and reasoning, from that of ‘unreliable’, or so-called ‘junk’ or ‘pseudo-science’. Unfortunately, Justice Blackmun’s use of ‘reliable’ in this manner is not consistent with definitions of the term as understood by scientists. Faigman (in Pyrek, 2007,) gives the example of a thermometer that always reads ten degrees high – that thermometer is one hundred percent reliable, because it always gives the same reading, but ultimately is also 100% wrong.<sup>1</sup> *Reliability* in a strictly scientific sense refers to the consistency of results in any given test. Many forensic areas pass muster in this particular aspect, such as polygraphy – the results of one examiner’s test generally agree with another examiners test. The key indicator in science that is considered equally, if not more important, is that of *validity*, which refers to the accuracy of a particular test; that is, it asks the question; ‘do the results give an answer that means what is claimed?’ Despite the fact that two terms are technically distinct from one another, much of the critical literature in forensic science uses the term *reliability* to include the meaning of both ‘validity’ and ‘reliability’.

A plain-meaning approach is the most obvious one to take when interpreting these terms. Most legal scholars agree that this was the intent of the US Supreme Court and is the current intent of the US Evidence statutes. As one commented when asked about the inconsistency

---

<sup>1</sup> This is because its accuracy, or *validity*, is zero. Therefore, the fact that it reads high means that its *reliability* of 100% also means that it is *wrong* 100% of the time.

between the US Supreme Court's interpretation of 'reliability' and the scientifically correct definition:

*'It is a relatively simple matter to resolve. I'm usually able to dispose of that with a footnote, where I say 'this is what scientists mean by the two terms, but the Supreme Court couldn't have possibly meant the technical meaning of 'reliability' when they said it'. Judges were using a perfectly plain vernacular meaning of 'reliability'. We all know what it means, and [scientists shouldn't insist] that reliability only means consistency and validity in this context, when that's not everybody's common-sense notion of it.'* (Anonymous)

So while scientists may articulate more specific, pointed definitions of these terms when using them in a scientific sense, most interpretations of the term 'reliable' when used in the legal sense are taken to mean essentially both. What the courts and critics ultimately want is some form of evidence that the expert 'can do what they say they can do' (Edmond, 2009). The *Daubert* decision suggested indicia that could be considered useful in carrying out this assessment for the expert's methodology, but these were non-exclusive, and by the Supreme Court's own admission in other cases, not necessarily applicable to all forms of evidence. Unfortunately, courts attempting reliability analysis of forensic evidence have repeatedly invoked the *Daubert* 'criteria' as benchmarks for this assessment. This has fuelled criticism from legal and scientific academics, who have not only reached profoundly different conclusions in conducting their own analyses using these guidelines, but have noted that these criteria do not wholly constitute what it means for a discipline to be 'reliable'. Additionally, most courts appear reluctant to challenge the underlying assumptions of long-standing forensic identification science disciplines altogether, and avoid grappling with the issue. This grandfathers much of the generally accepted forensic identification sciences, particularly when judges use the powerful concept of judicial notice<sup>2</sup> to avoid an in-depth analysis of a discipline's underlying theory. The nature of this debate leads one to the conclusion that a meaningful assessment of the scientific reliability of forensic odontology must consider multiple approaches, and not be limited purely to a *Daubert*-endorsed checklist.

---

<sup>2</sup> A judicially noticed fact must be one not subject to reasonable dispute in that it is either (1) generally known within the territorial jurisdiction of the trial court or (2) capable of accurate and ready determination by resort to sources whose accuracy cannot reasonably be questioned (Federal Rules of Evidence, Rule 202). Critics have argued that neither criterion is met for much of forensic science theory, however application of the rule is entirely discretionary.

## Architecture

This thesis explores the reliability of bitemark analysis by examining the discipline from several angles. The first explores the legal concepts of forensic identification evidence in both the United States — where much of this controversy has originated — as well as in Australia. The second involves an exploration of the reliability of the discipline from a theoretical perspective, by assessing the underlying scientific foundation for the theories and practices of bitemark analysis that have raised the concern of critics. The third, interwoven, aspect of enquiry into the reliability of bitemark analysis involves assessing the application of the human element to the legal and scientific aspects via a phenomenological approach. This is achieved by assessing both how judicial actors actually interpret and apply their expert evidence doctrine in forensic cases brought to trial, and how odontologists interpret and apply their theories and methods of bitemark analysis to casework.

### Legal Enquiry

Chapters 2 and 4 explore legal concepts related to the acceptance of forensic science evidence, in particular since the controversial *Daubert* decision. Chapter 2 focuses on this decision and its critical reception in the United States, in addition to examining the current legal basis for admission of expert scientific testimony in US courts. Chapter 4 considers the Australian legal system, and explores the basis for admission of forensic identification evidence in this country. A comparison of American and Australian judicial attitudes towards forensic science is conducted, by comparing and contrasting the judicial climate, legislation and seminal case law regarding the admission of expert evidence in both countries, in order to provide context for the practise of bitemark analysis in this country. Several academics, lawyers and judges in both Australia and the United States were consulted during this aspect of the enquiry, and these chapters incorporate information gleaned from interviews conducted with these individuals in order to emphasise the nature of current thinking in the legal sphere.

### Scientific Enquiry

Chapter 5 begins with an exploration of one of the fundamental concepts of the forensic identification science, including bitemark analysis, by assessing the scientific as well as philosophical basis for the concept that forensic markers can be considered ‘unique’, and that subsequent identifications can be made on this tenet. Chapter 6 continues with an exploration of some of the fundamental literature in the field of bitemark analysis, and engages in a critical analysis of endorsed methodologies in the field. The literature regarding accuracy and error rates associated with bitemark identification techniques is also considered. This



exploration is not meant to address the entirety of the literature regarding bitemarks, but is focussed on the concerns raised by the legal and academic criticism as related to the reliability of opinions espoused by odontologists concerning bitemark interpretation.

### **Phenomenological Enquiry**

While chapters 2, 4, 5 and 6, based on legal and scientific literature (including both peer-reviewed research, legal and scientific commentary, and historical case law), provide important grounds for assessment of the reliability bitemark analysis, chapters 3 and 7 focus on an approach to assessing reliability from an *experiential* perspective. The phenomenological component of this inquiry is thus framed by the relevant doctrine, and focuses on how these legal and scientific principles are implemented in the real world by the collection of both quantitative and qualitative data.

Chapter 3 characterises patterns of admission and exclusion of forensic identification evidence in the jurisdiction that currently has the most conservative precedent for admission of expert evidence by requiring an assessment of ‘reliability’ – that of the United States. In doing so, this chapter explores judicial attitudes towards forensic identification science evidence, which can then be distilled into consequences for forensic science practice. Chapter 7 parallels this line of evidence by exploring the nature of bitemark analysis as practised by odontologists in Australia through a retrospective analysis of bitemark casework since the year 2000, and engagement of practitioners in the field via interviews. This assessment of the real-world application of this discipline then allows one to analyse the nature of the relationship between scientific literature, contemporary practice, and current concerns. By establishing a baseline concept of where bitemark analysis currently sits we can then contrast the practice against scientifically established norms, and one is hence better placed to make recommendations for improving its overall reliability. Continuing to focus on the intersection of scientific theory and human experience, psychological influences on the practices of bitemark analysis that have hitherto been largely ignored by forensic science community are also explored in chapter 8.

The conclusionary chapter then ties these aspects together and summarises the current status of bitemark analysis, in terms of what it currently is, and what it perhaps should be in terms of legal and scientific expectations. Appendix 1 formalises the results of this enquiry through the draft of a series of guidelines for bitemark analysis in Australia.

## Phenomenological Analysis – Grounded Theory

Analysis of qualitative data obtained from case reports, judicial opinions, and interviews conducted during the course of this research proceeded via a phenomenological approach known as *grounded theory*. Grounded theory was a research model first proposed by Glaser and Strauss in 1967 (See Glaser et al., 1968), who attempted to integrate the strengths of quantitative methods of logic, rigor, and systematic analysis into a qualitative approach (Walker and Myrick, 2006). Described somewhat generally as ‘an inductive strategy for generating and confirming a theory that emerges from close involvement and direct contact with the empirical world’ (Patton, 2002), grounded theory involves a constant comparative method whereby data is simultaneously gathered and analysed via a technique known as ‘coding’, which attempts to categorise the data into recurring themes (Bryant and Charmaz, 2007). This then allows the development of a theory from the data, rather than attempting to ‘prove’ a preconceived hypothesis. A feature of this method is that it allows generation of explanatory propositions that correspond to real-world situations. A specific advantage of this method in health and medical practice research is that it enables data derived from practice to drive the research itself, and is not reliant on a hypothesis drawn from the literature review. A theory can be generated from this data that is drawn from actual practice, and this can then be used to inform *further* practice (Babchuk, 2009).

The original method as developed by Glaser and Straus has undergone several refinements since 1967, and has resulted in several variations of the original grounded theory method being realised. Glaser originally proposed a strict, 18-family coding that he has maintained allows the data to ‘speak for itself’, and allows the theory to ‘emerge’ from the data without undue influence from the researcher. Glaser maintains that this results in a more explanatory theory, rather than a descriptive one. Strauss split from Glaser in 1992, by proposing a more flexible coding protocol involving three hierarchies: open, axial and selective coding.

*Open coding* involves the selection and naming of categories from analysis of the data and is the first stage in analysis. It looks for variables in the data that can be labelled and then related together. *Axial coding* follows open coding, and involves relating the categories identified in open coding to one another. *Selective coding* occurs when a core category or theme is identified from the axial codings, and then analysis can proceed with categorisation within this core theme.

The coding process and data collection occur simultaneously. Once some axial coding has been completed, coding may proceed along the ‘selective’ coding lines rather than open or

axial. The use of selective coding allows analysis of the data to occur much quicker, and it is this particular feature of Strauss' approach that Glaser criticises. There is great divide between the Glaser and Straussian approaches to grounded theory data analysis, Glaser maintains that Strauss' approach 'forces' the data into categories, rather than allowing the data to speak for itself and thus generating a theory that is then entirely objective (Kelle, 2005; See Glaser, 1992). Strauss maintains that true objectivity is impossible in this approach, and so more rigor in the analysis phases is warranted. Strauss and Corbin also place less emphasis on 'memo-ing', a concept introduced by Glaser to assist with theory development that can involve writing a few words, sentences or pages that will assist in conceptualising that which may have only been descriptive at this stage (Boychuk Duchscher and Morgan, 2004).

Grounded theory continues to evolve, with other researchers commenting on positivist *versus* post-positivist approaches (based on the original Glaser *versus* Strauss debate), constructionist *versus* objectivist (Charmaz, 2000; Glaser, 2002; Mills et al., 2006; see Charmaz, 2007), and a variety of post-modern approaches to the grounded theory method (Annells, 1996; MacDonald and Schreiber, 2001; Robrecht, 1995; See also Clarke, 2005). Most of these engage in heady philosophical discussion about the advantages and disadvantages of knowledge acquired through various data collection and analysis methods applied during grounded theory-based research. Despite these and other practical differences in grounded theory processes as refined by others, the general concept nevertheless remains that grounded theory involves a coding process that progresses from descriptive through to higher levels of analytic.

Grounded theory is an extraordinarily flexible form of enquiry can be applied to several types of research questions including those of an explanatory, descriptive and predictive nature across numerous fields (Ellram, 1986). While grounded theory, by its nature, cannot be considered completely objective, authors have argued that qualitative, inductive approaches such as grounded theory are no more subjective than quantitative, deductive approaches. Subjectivity is merely introduced at a later, more visible stage of the research cycle, during the data analysis stages in qualitative approaches, whereas the weakness of quantitative approaches occurs in the research and data selection stages (Gasson, 2003).

The concept of validity in quantitative methods is often more appropriately phrased as the quest for the concept of internal consistency, thus substituting the criteria of credibility and authenticity for qualitative inquiries rather than pure validity (Miles and Huberman 1994). Credibility itself depends less on sample size and more on the richness of the data gathered

and the subsequent detail of the analysis (Patton, 2002). While the notion of ‘test–retest’ approaches to qualitative data collection such as questionnaires has long been thought to establish reliability (Charles, 1995), more recent thinking has noted that sensitization of the participants during this process can unduly influence the nature of responses given during either of these phases, rendering this approach less reliable for qualitative measures of social phenomena (see Golafshani, 2003). In grounded theory approaches, validity is best achieved via accurate description of the sources of data, using multiple sources (Eisenhart and Graebner, 2007; Ellram, 1986), reflexive approaches to subsequent analysis, and avoiding automatic hierarchical assignation of codes so as to permit a more open-minded approach to data interpretation (Gasson, 2003). Triangulation of data is also perceived as a particularly powerful tool for improving the validity of qualitative research (Golafshani, 2003; Patton 2002). Reliability of qualitative research is viewed as a product of validity, and thus demonstration of validity is usually considered sufficient to establish reliability (Lincoln and Guba, 1995; Hoepfl, 1997; Patton, 2002).

The central issue in qualitative approaches is often deemed to be generalizability, yet phenomenological inquiry uses a naturalistic approach that seeks to understand phenomena in context-specific settings (Hoepfl, 1997). As such, the generated theory cannot necessarily be seen to be applicable to situations outside the context in which they were gathered. However, the meaning of generalizability in the qualitative framework is to generate *theories* that are general (Yin, 1994), and not necessarily theories that are applicable across general populations. This concept is enhanced by the use of multiple sources of data. While theories generated by qualitative methods may not be applicable to other social contexts, they are almost certainly *transferable* between similar constructs (Gasson, 2003; Lincoln and Guber, 1985), a concept that is further enhanced by triangulation.

Despite its perceived phenomenological pitfalls, grounded theory has remained popular with researchers in health and the social sciences. It has been described as ‘the most influential paradigm for qualitative research today’ (Denzin and Lincoln, 2011). Its advantages include its capacity for detailed study of a micro issue of a larger reality within a particular setting, and thus it can develop detailed information about a particular phenomenon; and that it is progressive in that data collection and analysis are tightly interwoven in order to benefit from one another, advancing the growth of insight. The results of grounded theory approaches are thus not deduced from some general theory, but are discovered during the research process (Laws and McLeod, 2004).

Grounded theory was seen as particularly powerful tool to use in this enquiry into the reliability of bitemark analysis, in particular those aspects dealing with the phenomenological enquiry into the legal perceptions of forensic identification evidence, the nature of bitemark casework and the actual practice of bitemark analysis. The nature of debate concerning these topics is highly emotive, with vocal critics, an emotive context, media-sensationalised cases and equally vocal proponents, and it was important to consider these issues in the context of their practice while not being unduly influenced by these biases. Grounded theory was seen as enabling the data to present itself as a truer reflection of the status quo while avoiding the highly influential sources that have coloured so much of the literature in recent years. While Glaser's original concept is idealistic, the approach offered by Strauss and Corbin was seen as more applicable to this research, having a more logical and focussed application to a research question. This approach is acknowledged by several authors as being the most appropriate for beginning research in qualitative enquiries (Patton, 2002).

## **Phenomenological Data Collection**

Data collection for this thesis proceeded via two main methods: face-to-face interaction with experts via interviews; and analysis of historical documents such as court records and case files.

### **Expert Interviews**

Expert interviews are a common methodological approach in political science and sociology, and have also been used in clinical nursing and medicine (Benner, 1982; Marsh and Stoker, 1995; Severinsson and Borgenhammar, 2003). Grounded theory proponents advocate the use of interviewing as a data collection method (Charmaz, 2003; Corbin and Strauss, 2008), however, as a qualitative, rather than a quantitative method of research, it often faces criticism from those grounded in the 'hard' sciences as being biased, subjective, invalid, non-formalised, and person dependant. Despite the fact that it now appears to be a popular method of choice for qualitative data gathering, it still invokes controversy as a 'pure' research method. Kvale (1994) published a comprehensive reply to these criticisms of qualitative interviews in the mid 1990's. He argued that these critiques usually involve a prejudgment, based on a conception of [social] science where qualitative research is expelled or relegated to a secondary position; an approach to science and data collection that no longer enjoys mainstream philosophical acceptance. Elsewhere, he has commented that there is much misconception regarding the purpose of the qualitative interview (Kvale, 1983).

Kvale expanded on this concept by noting twelve aspects of qualitative interviews that are often ignored by those who critique the technique. Some of the more salient points he has raised are that the qualitative interview allows: the interviewee to describe and understand the meaning of central themes in their own 'life-world'; the researcher to obtain specific, nuanced information from the interviewee – corresponding to exactness in quantitative measurements; the researcher to ignore his own presuppositions regarding a topic in hearing the interviewee's personal, uninhibited responses, and vice-versa – it also allows the interviewee to experience alternative viewpoints. At an even more fundamental level, it also allows the researcher to clarify and contextualise instances where ambiguity and contradiction are apparent in the interviewee's responses.

Richards (2007) has also cited other advantages in using interviews as a research methodology, such as: they can help in interpreting other literature, such as documents, articles or reports; they can help in interpreting the personalities involved in the relevant decisions and help explain the outcome of events; they can provide information not recorded elsewhere, or available for release; they can help establish networks by providing links to other individuals, and; they can help the researcher to understand the context of the research topic. In-depth interviews with experts have been used with great success as an alternative to questionnaires when applying Delphi techniques as they provide a flexibility that is absent in questionnaires. This can provide the researcher with the opportunity to probe the reasons behind decisions or viewpoints, and follow up on unexpected hints dropped by the interviewee (Glenn and Gordon, 2003).

One concern with interviewing is the potential for interaction effects (Vogel, 1995). These are commonly seen in conversational interaction, and result in the information content of such conversations being minimised. Such effects include the 'iceberg effect', where the interviewee appears to be disinterested and reluctant to give information; the 'profile effect', where the interviewee seeks to 'show off' in front of the interviewer; or the 'feedback effect', where the 'questioning' role of the interviewer and 'answering' role of the interviewee become reversed. Some of these effects, such as the feedback effect, generally only manifest in particular contexts, for example when discussing conflict-laden topics. Careful approaches by the interviewer to remain open-minded, and relatively neutral, minimise the risks of these effects exerting significant influence on the quality of the information obtained during the interviews. Pfadenhauer (2009) has admitted that regarding performance routines, skills and automatic behaviour patterns, interviews can produce deficient or misleading results, however, she maintained that interviews have proven to be

suitable tools for the reconstruction of knowledge that has been ‘learned’ and can be thematically distinguished and explained by an expert.

Meuser and Nagle (See Pfadenhauer, 2009) have argued that the negative perception of the expert interview as a research methodology is due to the persistence of the view that interviews are primarily exploratory in nature, and thus do not fulfil the requirements of a methodologically significant contributor to data gathering. Accordingly, they and others have defined several interview models that are differentiated by the quality and usefulness of data they generate.

The *exploratory* interview embodies this traditional concept of the expert interview. In this context, the interview primarily serves to establish an initial orientation in a field that is either substantively new, or poorly defined (Bogner and Menz, 2001, in Bogner et al., 2009). Its purpose is thus to help the researcher develop a clearer idea of the issue being studied. It does not necessarily involve the systemic gathering, coding or standardising of data, and thus differs from the two other types of interview.

The *systematising* expert interview is oriented towards gaining access to exclusive knowledge possessed by the expert (Bogner and Menz, 2001, in Bogner et al., 2009). The type of knowledge sought by the interviewer is derived from the practice of the expert. This is considered the most widespread form of interview method in research practice. Other recognised forms of research methodology, not necessarily open and purely qualitative, also fit into this category, such as the Delphi method.

The third type of expert interview is termed *theory-generating*. Originally developed by Mueser and Nagle (see Witzel, 2000), this methodology relies on the expert interview as a means of sampling and comparative analysis, which the researcher then formulates into a hypothesis. This hypothesis is gradually refined by the acquisition of knowledge from other experts, leading to the development of a formal theory.

In practice, it is difficult to separate these interview methodologies from one another when pursuing a broad research goal. The overall project often ends up being a hybrid of these methods, beginning with exploratory interviews, before progressing to one or both of the other methodologies. Elements of all three interview methodologies ultimately play a part in data gathering and analysis by interview methods. The particular strategy used by the researcher depends on the relationship established between the interviewer and the interviewee. Bogner and Menz (2001) advocate that where the interviewer could be

considered an expert but is from a different knowledge culture than the interviewee, the exploratory and systematising interview paradigms should dominate the research method.

A key question regarding the feasibility of interviews as a method of data collection was related to the required 'sample size' in order to ensure that all possible factors were given the opportunity to be explored. While mathematical formulae exist for determining sample size for quantitative analytical methods, qualitative methods require a different approach. Patton (2002) cites several sources advocating 'redundancy' as the primary criterion for determining sample size – the sampling process is terminated when no new information is forthcoming from sampled units (see Lincoln and Guba, 1985). Owing to the relatively small number of scholars and practitioners in the relevant fields associated with this research, redundancy for the early interviews was achieved after what would ordinarily be considered relatively modest sample sizes. This is typical in narrow qualitative enquiries, which often achieve redundancy after no more than 15 or 20 participants.

### ***Academic Interviews***

The researcher, with experience in forensic odontology but little experience in law, found it necessary to conduct interviews with legal academics, lawyers, judges and sociologists of science. Following a literature review, it was possible to identify the key players relevant to the research question. Interviews conducted with these candidates were often contextually broad, and had the primary focus of defining problem boundaries. The objective of these early interviews was not to generate codifiable data, but to narrow the field of issues to those relevant to the primary research topic. Consequently, these interviews tended to be exploratory in nature, progressing to theory-generating as the researcher gathered more and more information from the interviewees. Interviews with various judges, lawyers and academics were conducted in both Australia and the United States. Both written notes and an audio recording of each interview were made. All subjects agreed to participate on this basis and were consented in accordance with the University of Newcastle Ethics Committee policies (Approval No H-2009-0204). These interviews proceed more in line with the exploratory and systematising approaches, in order to gain orientation with legal concepts and attitudes. Interviewee participants were selected on the basis of their standing within the legal community, rather than using a random sampling method, as the subject matter was considered to be too specific for a randomised approach.

These interviews were most often used to clarify points that had been raised in the literature by the participant, or to deepen the understanding of broader concepts regarding the law. While there was general intent to quote from these interviews as necessary during the



production of this thesis, some of these participants wished to remain anonymous, and others preferred not to be interviewed in any ‘formal’ sense. Quotations from these interviews are used throughout this thesis in order to highlight important points that are perhaps not evident through the published literature.

### ***Odontologist Interviews***

Interviews with practising Australian odontologists formed part of the second aspect of the phenomenological enquiry. The sample size for this aspect was considered in light of the total number of practising odontologists in Australia, as members of the Australian Society of Forensic Odontology (AuSFO) – currently the peak body for odontology practice in Australia. While membership of the society is not a pre-requisite for practice as an odontologist in this country, the vast majority of practising odontologists are members of this organisation. Currently the society has 54 members, of which less than twenty are engaged in permanent, paid forensic work of one day a week or greater. The remainder either participate on an on-call roster, or are called on sporadically, as needed during large-scale disaster efforts. Consequently, the majority of the members of AuSFO would have had little, if any, experience in bite mark analysis, with the majority of experience and expertise lying with those 20 or so members engaged in either full or part-time work.

15 of these ‘permanent’ members agreed to be interviewed. Anonymity was protected and only the author, who conducted these interviews, knew the identities of the participants. Random numbers were assigned to each participant as a means of identification using the random-number generator formula in Microsoft Excel. Again, both written notes and an audio recording of each interview were made. All subjects agreed to participate on this basis and were consented in accordance with the University of Newcastle Ethics Committee policies (Approval No H-2009-0204).

Interviews generally proceeded along a semi-structured approach, in order to allow maximum engagement with the participant, although in order to generate more focussed data this required the use of a more formalised interview rubric, which was developed to enable easier coding and analysis of the subsequent data. All odontologist interviewees were therefore asked similar questions, in contrast to the approach used in the earlier interviews, which were tailored specifically to each participant in order to draw on their particular areas of expertise.

All odontologists were guaranteed anonymity, although some expressed willingness to be identified.<sup>3</sup> The structured part of the interview involved asking odontologists their opinions regarding a series of six images purported to be bitemarks. These electronic images were from contemporary cases presented to one institution in Australia. None of the cases from which the images were associated were *sub judice* and none of the interview participants had the opportunity to view or analyse the bitemarks before being shown them during the interview. These images were loaded into a Microsoft PowerPoint presentation, using a plain black background, labelled A to F, and were suitably anonymised so as to prevent the participants knowing their origin. The resolution of each photograph was not changed or enhanced from the original image presented to the odontology department. Participants were shown the images using the full-screen 'presentation' feature on the author's personal laptop computer. They were asked to comment generally on the bitemark, and to include as much or as little information as they liked regarding its quality, orientation, origin, or potential use as an identification tool. There was no time limit set and participants could take as long as they wanted to assess each mark.

Once all six bitemarks had been commented upon, they were then asked to rate each bitemark in accordance with the severity and significance scale developed by Pretty (2007). At first, only the written descriptions of each score were given to use as a reference guide (first pass). After all six images had been rated, the participants were then given a laminated colour copy of the exemplar marks originally published by Pretty to demonstrate each of the scores. They were then asked to again rate the bitemark images using the visual exemplar provided as a guide to scoring (second pass).

The comments made by each participant were transcribed and then analysed using the grounded theory approach. Axial and theoretical coding allowed linkage of these themes in order to generate an overall perspective on the nature of bitemark analysis from practising odontologists.

### **Retrospective Casework Analysis**

The second method of data collection for the phenomenological enquiry focussed on what legal and forensic practitioners are actually *doing*, rather than that what they *say* they are doing.

---

<sup>3</sup> Despite this, none are associated by name with any of the data contained in this thesis.

***Court Records***

The source of data that was deemed best to explore this issue in the legal setting was in published legal opinion from cases specifically where the issue of admissibility of identification evidence had been considered. Legal databases are the obvious source for these opinions, however these only selectively return unreported judgements. Searching for case law specifically regarding forensic identification science evidence issues was initially difficult, as so much of criminal law refers to forensic evidence in some way. Vast amounts of cases were returned in early searches, with very little of them specifically involving a legal challenge to the evidence itself.

An alternative to the use of legal databases such as Westlaw<sup>®</sup> and Lexis<sup>®</sup>, was offered in the form of an expert witness database, the Daubert Tracker<sup>™</sup>. This database is accessible via the World Wide Web on a fee-for-service basis, and was primarily designed to assist lawyers in research for a witness with specific expertise. This database claims to be comprised of all reported and numerous unreported cases from both US state and federal jurisdictions where an admissibility standard has been cited or mentioned in a decision and where a testifying expert's methodology or qualification has been challenged. Daubert Tracker also has a repository of gatekeeping documents including opinions, briefs, docket sheets and transcripts linked from the case record. It derives its information primarily from these sources, as well as others such as the PACER [Public Access to Electronic Court Records] database, Courtlink<sup>™</sup>, court websites and via submission from individual legal authorities. This tool provided a much more focussed database with which to base an empirical enquiry into legal decision making in forensic science evidence challenges.

The written judicial opinion for each case where forensic identification evidence was excluded was examined in line with the grounded theory approach adopted by Strauss and Corbin. Following several readings of each opinion, comments that related to the reasons why forensic identification evidence were identified. Open, axial and theoretical coding of these comments was then conducted, allowing theories regarding the judicial exclusion of forensic identification evidence to be generated.

***Bitemark Casework***

With regard to forensic odontology practice in bitemark analysis, it was decided to undertake a retrospective review of casework conducted in the last ten years in Australia. Such a review provides useful information for critics and proponents alike, including the recognition of strengths and weaknesses in the current practice of bitemark analysis in this country, and provides yet another perspective on the nature and conduct of bitemark analysis that is

distinct from expert opinions on the matter. Five major odontology centres were contacted in Australia to ask whether it would be possible to conduct a review of case files for bitemark analyses conducted at that centre in the last ten years. While all centres were willing to participate, the different filing protocols for cases caused insurmountable issues in the collection of data from three of these centres. Consequently, data was collected from only two.

Case files from January 2000 to July 2010 were reviewed on site by the author of this thesis. Cases were initially coded according to the type of crime involved, the material quality of the evidence presented, and the method of analysis in order to generate data on the nature of bitemark casework in Australia. As part of this coding, all of the photographic records of bitemarks that were available were rated for severity and significance in accordance with the scale developed by Pretty (2007). The bitemarks were rated using both a copy of the descriptor and a colour copy of the scale exemplars (as published by Pretty in his original article) for direct reference and comparison. The odontologists' conclusionary statements were analysed on a semantic basis for each case, using a grounded theory approach involving open and axial coding methods. Data from this analysis allowed a theory regarding the nature of bitemark casework in Australia to be generated, in addition to one of the overall nature of practice and reporting in bitemark casework by Australian odontologists.

## CHAPTER 2

# The Admissibility of Expert Evidence: Frye, Daubert and the US Federal Rules of Evidence

---

Forensic identification science stands apart from other scientific disciplines in that this specialty grew out of collaboration between two distinct philosophies, science and law. The law remains forensic science's biggest customer, and in order to maintain its position as a valued tool in the law's armamentarium against crime and injustice, it is expected that forensic science will meet basic thresholds and principles as laid down in legal doctrine. When considering questions regarding the 'reliability' of any particular scientific process, the question is often asked: 'reliable enough for whom?' The answer for forensic science is not a simple one: without acceptance from the legal sphere it is unlikely to survive as little other than an anecdotal or academic discipline – its applicability outside legal contexts is relatively small, particularly for disciplines such as bite mark analysis, firearms and toolmark analysis and handwriting analysis. This and the following two chapters attempt to clarify the law's expectation of expert witness testimony, in order that we may understand the legal philosophies that guide the practice of forensic science. This in turn, will ensure greater longevity as a customer of the courts, and ultimately, as a scientific discipline.

This chapter specifically focuses on legal standards in the United States. The applicability of these standards is discussed further in chapter 4, where an exploration of similarly-based Australian admissibility standards ensues. There are several reasons for conducting this seemingly peripheral discussion to the core focus of this thesis, that of the reliability of bite mark analysis in Australia. Firstly, the US court system is, like Australia's, based on an adversarial system where two parties are required, as advocates for opposing viewpoints, to present their cases to an impartial third party (i.e. a judge or jury). Legal doctrine in Australia and the United States was distilled from the English model, and as such, shares a common history and philosophy. The United States also has one of the most well-developed legal and judicial sectors in the world, and combined with one of the highest populations and one of the highest crime rates in the Western world, it provides opportunity for studying the

interaction between forensic science and law that is difficult to observe elsewhere simply because of relative incidence.

Much of the forensic science literature emanates from the United States, and the practice of forensic science, as already discussed, is largely driven by the legal sector. While there are differences in legal doctrine between the United States and Australia, the practices of forensic science tend not to be so compartmentalised. Much forensic science research is conducted and published in the United States, and via the *scientific* literature, quickly gains acceptance in other parts of the world, at least with the practitioners of the discipline. Understanding baseline legal models in the United States helps explain the nature and direction of this research, and ultimately, forensic practice across the world.

Perhaps one of the main reasons for focussing on legal doctrine in the United States is to clarify what has become one of the most controversial topics in forensic science since the mid-1990's — the *Daubert* standard for admissibility of expert evidence in courts of law. The implications of this decision have been discussed in both Australian and English scientific and legal literature due to the perceived nature of its consequences for admissibility standards for expert evidence, with forensic science evidence gaining particular attention. The literature in forensic science continues to cite this legal decision, and its progeny, nearly some twenty years after it first captured the attention of forensic scientists. Yet relatively little of the literature in forensic science, in stark contrast to the legal literature, has actually conducted a meaningful analysis of what the *Daubert* standard means. Anecdotally, it appears that the forensic science literature compresses *Daubert* and the admissibility of scientific evidence in general into a few lines of text, as a cursory introduction to the justification for forensic science theory, practice and research. A more detailed understanding of the context of this decision, as well as its place in 21<sup>st</sup> century rules regarding admissibility of expert evidence, is surely warranted in order to ensure that we are not perpetuating misunderstandings, or out-dated and idealistic views of legal requisites that are subsequently influencing our practice and research directions.

## **General Acceptance and the *Frye* Standard**

It was initially a murder case, *Frye v United States* (1923) that set the standard for the admissibility of scientific expert testimony in the United States for nearly seventy years. In *Frye*, an expert for the defendant offered testimony based on the results of an early form of

polygraph testing<sup>1</sup> in support of his innocence. The trial judge found this evidence inadmissible, citing the fact that there was no ‘general acceptance’ of this technique in any relevant community of experts. The appellate court upheld the trial judge’s ruling, and ‘general acceptance’ became the standard to which other expert testimony was subsequently held in all other US District courts from that time on.<sup>2</sup> Proponents of the *Frye* standard claimed that its use promoted judicial efficiency (*Reed v State*, 1978) and uniformity of decisions through the courts (*People v Kelley*, 1976), as well as preventing unproven hypotheses from having a potentially prejudicial effect on the trier of fact (*United States v Brown*, 1978).

One of the most influential critics of the *Frye* standard was McCormick, who as early as 1954 declared general acceptance ‘an improper criterion for admissibility’, and argued that the judge or jury should be allowed to make their own assessment of the evidence via the traditional use of examination and cross examination, in keeping with the adversarial philosophy of American justice (McCormick, 1954). *Frye* was criticised by others for its generic nature, who noted that it failed to define what was meant by ‘general acceptance’, or ‘a particular field’s boundaries’ (Louisell and Mueller, 1977), making its application in court susceptible to manipulation. The *Frye* opinion also only referred to *novel* scientific evidence, which led to further concerns among the legal fraternity regarding how to ascertain whether something was ‘novel’, or even ‘scientific’ (Taslitz, 1995). Applying the *Frye* standard was described as ‘difficult’, and others thought its focus had a tendency to ‘obscure the issues’ rather than exemplify them (Giannelli, 1980).

Black and colleagues (Black et al., 1994) noted that the vagueness of *Frye* ultimately gave free reign to individual courts in determining fundamental issues, such as what particular scientific field should be consulted; what was meant by ‘general acceptance’; and what demonstrated that such acceptance had been achieved. In their opinion, *Frye* did not appear to address the real question at hand, which was whether the testimony was derived from valid scientific knowledge, and thus it only led to confusion and contradictory rulings in different jurisdictions concerning the same types of evidence. More recent authors have opined that this inconsistency in the application of *Frye* across cases in the US arose because some courts did little more than ‘count [expert] noses’, whereas others performed in-depth analyses of the proffered expertise (Beecher-Monas, 2007).

---

<sup>1</sup> The test was officially known as the ‘systolic blood pressure deception test’.

<sup>2</sup> Although several authors have noted that *Frye* was virtually unheard of until the mid 1970’s, when the federalisation of criminal law permitted a subsequent rise in the number of citations of the *Frye* case.

Saks and Koehler argued that the *Frye* standard forced courts to adopt the standards of the very discipline under scrutiny (Saks and Koehler, 1991), and another author summarised the situation more recently by stating that *Frye* permitted testimony to be given by what were essentially ‘self-validating’ experts (Pyrek, 2007). The *Frye* standard was also said to have shifted the burden of analysis away from both judge and jury (Black et al., 1994), and while some felt that this was acceptable, as the expert was supposedly the most qualified person in court for analysing their disciplines’ particular methodology (*United States v Addison*, 1974), others have been critical of this shift, arguing that it gives what is technically a judicial policy decision to non-judicial actors (Beecher-Monas, 2007).

*Frye* was also accused of causing unnecessary delay and conservatism in the courts in waiting for supposed ‘general acceptance’ (Boyce, 1963; Imwinkelried, 1992; Lederer, 1986). While Black and colleagues argued that such cases have been few, these authors contend that probative evidence in past cases may have been excluded unnecessarily, with the passage of time having revealed it to be admissible as more scientists warmed to a particular theory (Giannelli, 1983).

## The US Federal Rules of Evidence

In 1975, definitive legislation, as opposed to case law, pertaining to the admissibility of evidence at trial was passed in the form of the US Federal Rules of Evidence.<sup>3</sup> This statute was drafted by the Supreme Court and approved by Congress, which enacted the Rules under public law. The Rules provided a legislative framework for evidentiary standards that were intended to be applied across the federal judicial system, and therefore would provide a more standardised approach to evidence admissibility in the Federal courts.

Judges were entitled to exclude evidence if evidence was deemed to be irrelevant, with a subset of such relevance being the *reliability* of the evidence: the reasoning basically follows that if expert evidence was not found to be reliable, then this also deems it to be irrelevant, and can be excluded. However, there was no specific guidance within the Federal Rules on how to assess such evidence for relevance or reliability. Some scholars proposed evaluation of scientific evidence via a series of ‘factors’ to be considered by the trial judge. In 1985, former Iowa State Supreme Court Justice Mark McCormick (1982) proposed criteria for determining admissibility of evidence that included:

1. The potential error rate in using the technique;

---

<sup>3</sup> Referred to throughout this dissertation as the ‘Federal Rules’, or simply, the ‘Rules’.



2. The existence and maintenance of standards governing its use;
3. Presence of safeguards in the characteristics of the technique;
4. Analogy to other scientific techniques whose results are admissible;
5. The extent to which the technique has been accepted by scientists in the field involved;
6. The nature and breadth of the inference adduced;
7. The clarity and simplicity with which the techniques can be described and its results explained;
8. The extent to which the basic data are verifiable by the jury;
9. The availability of other experts to test and evaluate the technique;
10. The probative significance of the evidence in the circumstances of the case; and
11. The care with which the technique was employed in the case.

In the same year, the 3<sup>rd</sup> Circuit ruled in *United States v Downing* (1985) that in addition to factors such as error rate, testability, peer review and general acceptance,<sup>4</sup> the court should consider an assessment of whether there is a body of professional literature on the subject; whether there is a non-judicial use for the particular aspect of science in question; the relative novelty of the technique; and the qualification of the witness. Weinstein and Berger also proposed a 7-factor test similar to that proposed by McCormick in their textbook on evidence (Weinstein and Berger, 1987).

While *Frye* proponents rejected the use of such ‘factors’, continuing to hold that ‘general acceptance’ was a sufficient indicator of reliability, Jonakait, among others, argued that the adoption of the Federal Rules entailed an explicit rejection of the *Frye* test for assessing the reliability of expert witness evidence (Jonakait, 1990). As this debate raged on, the Supreme Court granted certiorari<sup>5</sup> for a case that would supposedly decide the issue for once and for all.

### ***Daubert v Merrell Dow Pharmaceuticals Inc.***

The vehicle for this ruling was the toxic tort<sup>6</sup> *Daubert v Merrell Dow Pharmaceuticals Inc.* (1993), a civil law suit brought against the manufacturer of an anti-nausea drug known as

---

<sup>4</sup> Note the striking resemblance here to the four factors mentioned in *Daubert v Merrell Dow Pharmaceuticals*, which was still eight years in the future at the time of *Downing*.

<sup>5</sup> The U.S. Supreme Court does not accept all cases. A petitioner must submit a ‘writ of certiorari’ outlining their case for consideration by the justices. If four of the nine justices agree to hear the case, the court ‘grants certiorari’, and then sets a date to hear oral argument.

<sup>6</sup> A type of personal injury lawsuit in which the plaintiff claims that exposure to a chemical caused the plaintiff's injury.

Bendectin. In the District Court, the family of Jason Daubert sued Merrell Dow Pharmaceuticals for damages based on the fact that his physical birth defect was caused by his mother's use of Bendectin during her pregnancy. They tendered an expert who provided evidence that a link between ingestion of Bendectin and birth defects existed, based on re-analysis of pre-existing in-vitro and animal studies that had already been published with the *opposite* conclusions.

The manufacturer of Bendectin submitted evidence that no published study had demonstrated a link between the popular anti-nausea drug and birth defects, and thus won a summary judgement<sup>7</sup> based on the *Frye* principle that the re-analysis methodologies used by the plaintiff's experts in reaching their conclusions had not gained general acceptance, therefore this evidence was unreliable and consequently, inadmissible. Without this evidence, the plaintiffs essentially had no case. The plaintiffs appealed to the 9<sup>th</sup> Circuit, where the decision made by the trial judge was upheld, the ruling declaring that expert opinion based on a "...methodology that diverges significantly from the procedures accepted by recognized authorities in the field ... cannot be shown to be 'generally accepted' as a reliable technique." Despite two failed attempts at litigation, the plaintiffs appealed once more to the US Supreme Court.

The appeal to the Supreme Court was based on the fact that the lower courts had applied an inappropriate standard for the admissibility of expert evidence, as *Frye* had been superseded by the Federal Rules, and hence the trial judge had erred in rejecting the scientific testimony presented in the initial case by claiming it was not 'generally accepted'. The Supreme Court held that the Federal Rules of Evidence had indeed superseded the *Frye* criteria, and suggested that there were other factors that the lower courts should consider when deciding admissibility of expert evidence, not solely that of 'general acceptance'. They noted that the Federal Rules of Evidence represented the intent of the law when it came to scrutiny of expert testimony, and using these rules as a framework, the majority opinion of the court articulated criteria that could be considered in assessing the validity of scientific evidence. They declared that scientific evidence should be subject to a 'reliability test', rather than a 'general acceptance test', with judges required to act as 'gatekeepers' so as to ensure that the reasoning and methodology of the expert testimony is scientifically valid; and that such reasoning could then be applied to the facts at issue.

---

<sup>7</sup> A *summary judgement* is a decision made by a judge on the basis of statements and evidence presented for the record without a trial. It is usually used when there is no dispute as to the facts of the case, and one party is entitled to judgement as a matter of law.

The criteria that judges may use<sup>8</sup> when performing the assessment of reasoning and methodology were summarised by the Supreme Court opinion, authored by Justice Blackmun, as:

1. Whether the theory or technique can be or has been tested;
2. Whether the theory or technique been subjected to peer review and publication;
3. Consideration of the known or potential error rate associated with the technique;
4. The existence of standards controlling the method; and
5. Whether or not the theory or technique enjoys general acceptance in the relevant scientific community.

This opinion rendered a paradigm shift in the way expert testimony was now to be considered. Under *Frye* and the Federal Rules, the admissibility of scientific testimony depended largely on the qualifications of the expert proffering it. Consideration was made of the expert themselves, by reviewing their professional qualifications, their experience, and their standing in the relevant field. This standing was then automatically transposed to the evidence, with little or no scrutiny given to its scientific basis. Now, following *Daubert*, the Supreme Court specifically directed judges to consider evaluating the expert's reasoning process and methodology, as opposed to resting their basis for admission on the expert's credentials.<sup>9</sup>

## Responses to *Daubert*

A wealth of legal literature arose following the publication of the *Daubert* decision, with much of it having predicted wide-ranging consequences for evidence law (Imwinkelried, 1995). By 1997, there were over 1000 law review articles and 1000 cases citing *Daubert*<sup>10</sup> (Thornburgh, 1997), however, as one scholar noted early on '...much of the commentary regarding the *Daubert* decision is of limited utility, dwelling excessively on relatively tangential points discussed in *Daubert*, many of which, as Chief Justice Rehnquist points out, tend to be not only general, but vague and abstract' (Chesebro, 1994). Nonetheless, some of these early articles offered practical suggestions to be incorporated into court

---

<sup>8</sup> Note that the *Daubert* opinion also uses the words '*may use*', not '*must use*' or '*should use*'. This implies that these criteria are only suggestions, a fact re-iterated by the Supreme Court in a later case, *Kumho Tire v Carmichael* (1999)

<sup>9</sup> The *Daubert* case was then remanded back to the 9<sup>th</sup> Circuit, where incidentally, *Merrell Dow* again filed for summary judgement, and again, the 9<sup>th</sup> Circuit ruled in their favour, despite the further consideration given to the defendant's testimony by the appellate judge under the Supreme Court's criteria (*Daubert v Merrell Dow Pharmaceuticals Inc. II*, 1995)

<sup>10</sup> A search of the Westlaw database on 18 June 2009 for 'Daubert' in article or case text yielded more than 5000 legal articles and cases.

proceedings so as to ensure the task of assessing the reliability of expert evidence was carried out sufficiently (see Bernstein, 1994; Black et al., 1994; Fenner, 1996; Hutchison and Ashby, 1994).

Others, however, predicted difficulties in carrying out the Supreme Court's direction due to a lack of resources in county and rural areas, and the lack of scientific training offered to judges and lawyers. One in particular commented that there appeared to be a 'chasm between state trial court reality and the commentary produced by academics and appellate judges' (Gless, 1995). Another author, who happened to have been involved in writing an *amicus curiae* brief<sup>11</sup> for the *Daubert* case, agrees that *Frye* was never an appropriate standard for admissibility of expert evidence, but criticises the 'general observations' made by the Supreme Court in the *Daubert* decision regarding the characteristics of good science, paraphrasing Chief Justice Rehnquist's dissent and pointing out that '...general observations by the Supreme Court all too quickly ossify into rigid authority' (Koukoutchos, 1994).

Several authors criticised *Daubert* as a representation of a clearly flawed approach to assessing scientific worth. One of the fundamental philosophical issues that arose from this line of discussion was that *Daubert* failed to recognise the distinct difference in the methodologies and expected outcomes of a scientific inquiry versus a legal one. Caudill and LaRue (2006) wrote extensively on the ideological interaction between science and law in their work *No Magic Wand*, which characterises the wide disparity that currently exists between legal and scientific ideologies. Faigman and colleagues (Faigman et al., 2007) noted that science searches for 'facts, which are broadly applicable and not adjustable or open to interpretation', whereas the law seeks 'justice, a concept that involves negotiation, particular to each individual situation'. Science and the law also both seek 'truth', however it is now largely recognised that *legal truth* is a distinct concept from *scientific truth*. Scientists also recognise that the 'truth' is mutable and may evolve over time, whereas the law requires a concept of truth be established relatively quickly in order to resolve the case before the court (Science, Technology and Law Panel, 2002). Many authors saw *Daubert's* attempt to reconcile these disciplines and their diametrically opposing philosophies as inadequate.

In *Daubert*, Blackmun J uses terms such as *falsifiability* and *falsified*, construing them as a cornerstones of modern scientific method, but defines neither<sup>12</sup>. Chief Justice Rehnquist, in

---

<sup>11</sup> Literally, 'friend of the court' brief. A person with strong interest in or views on the subject matter of an action, but not a party to the action, may petition the court for permission to file a brief, ostensibly on behalf of a party but actually to suggest a rationale consistent with its own views.

<sup>12</sup> 'Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested. "Scientific methodology today is based on generating hypotheses and testing them to see if they can be falsified;

his dissenting *Daubert* opinion, stated; ‘I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its "falsifiability," and I suspect some of them will be, too’. Scholars have mostly turned towards philosopher Karl Popper’s use of these terms when searching for their meaning.<sup>13</sup> Popper’s premise was that falsifiability, the ability to prove something false by observation or experiment rather than true, was a key delineator between science and non-science. This led to the assumption by many legal commentators, and some judges, that any hypotheses or technique that could not be proven false was, therefore, not scientific. Often ignored by such commentary was that Popper eventually appreciated that it was impossible to discriminate science from non-science on the basis of the falsifiability of a theory alone (Stanford Encyclopedia of Philosophy, 1997), and thus O’Connor (1995) has dismissed most legal scholars’ interpretation of Popper’s work as inaccurate for failing to acknowledge Popper’s eventual conclusions. O’Connor nonetheless confirms that ‘falsifiability’ is not a suitable model for distinguishing bad science from good, describing Justice Blackman’s use of such terms as a valid process for assessing science as ‘...potentially problematic’. Kaye, too, argued that commentators who believed *Daubert* represented an adoption of the philosophy of science of Karl Popper are misguided, and that courts who have used falsifiability as a determinant of a valid scientific method in their enquiries have misconstrued and misapplied the *Daubert* ruling (Kaye, 2004).

There was certainly general confusion amongst lawyers, judges, and scholars alike regarding not only its meaning, but also what the effects of the *Daubert* approach would be (Fenner, 1996, p. 955). No example better illustrates this than when both the original defendant *and* the plaintiff declared ‘victory’ following the *Daubert* ruling, each confident that the remand hearing would rule in their favour under the new evidence guidelines proposed by Blackmun (Mervis, 1993). Mansfield offered numerous interpretations of the Supreme Court’s definition of ‘scientific testimony’, ‘validity’, ‘falsifiability’ and ‘science’, in doing so highlighting the confusion surrounding the definition of these terms amongst the legal fraternity, as well as their incompatibility with scientific interpretations of the same terminology (Mansfield, 1996). Farrell (1994) predicted that application of *Daubert* logic in the lower courts would actually result in *more* inconsistent jury verdicts and judgements due to its complexity. Other authors have highlighted alternate concerns with the *Daubert* approach, such as an unnecessary ‘clogging-up’ of the judicial system with lengthy pre-trial motions and *in*

---

indeed, this methodology is what distinguishes science from other fields of human inquiry.” (*Daubert v Merrell Dow Pharmaceuticals, Inc.*, 1993, quoting; *Green*, 1992).

<sup>13</sup> For some examples, see Allen (1994) and Black (1994).

*limine*<sup>14</sup> hearings (Tamarelli, 1994), or the fear that lower courts would largely ignore the *Daubert* direction and continue to adopt *Frye* under the guise of the new language proposed by the *Daubert* opinion (Allen, 1994).

The question of whether the *Daubert* test was more permissive or more restrictive than *Frye* appeared to have no simple answer (Capra, 1997). Polentz (1996) noted that examination of *Daubert*'s workings in the courts revealed widely disparate applications; the 2<sup>nd</sup> Circuit adopted a liberal approach to admissibility of evidence, reckoning that the standard had been lowered, whereas the 8<sup>th</sup> Circuit ruled that the standard had been raised. The 9<sup>th</sup> Circuit rigidly followed the Supreme Court's suggested criteria for assessing expert testimony, notably also in the remanded *Daubert* case (*Daubert v Merrell Dow Pharmaceuticals Inc. II*, 1995), but complained about the nature of the test and expressed concern regarding the judiciary's general lack of knowledge about scientific matters, as well as their general discomfort with decision making when they themselves didn't understand the evidence. The 10<sup>th</sup> Circuit adopted a somewhat middle ground approach, between the strict enquiry conducted by the 8<sup>th</sup> Circuit and the 2<sup>nd</sup> Circuit's liberal approach. Caudill and Redding later criticised the more conservative approaches as out-of-date versions of scientific enterprise, perceiving liberal interpretations as being more consistent with contemporary insights from history, philosophy and the sociology of science (Caudill and Redding, 2000). They suggested a more centre-left approach, coining 'pragmatic constructivism' as the appropriate ideology. Some scholars initially predicted that *Daubert* would paradoxically be both more and less restrictive of expert evidence (Berger, 1993), the answer not necessarily depending on the application of the standard itself by the court, but on the standard of testimony proffered by the expert. This has proved to be the case, with Faigman and colleagues proposing the following two-axis illustration by way of explanation (adapted from Faigman et al., 2007):

		Scientific Foundation ( <i>Daubert</i> )	
		Strong	Weak
General Acceptance ( <i>Frye</i> )	High	Both admit	<i>Frye</i> admits <i>Daubert</i> excludes
	Low	<i>Frye</i> excludes <i>Daubert</i> admits	Both exclude

Figure 1.1 *Daubert* vs *Frye*

<sup>14</sup> Literally 'at the threshold'. An *in limine* hearing occurs as a separate event before the trial begins, or sometimes during the course of the trial. In this context, it refers to a hearing conducted in order to rule on the admissibility of an expert's evidence before the expert is allowed testify at the trial.

The Harvard Law Review addressed a number of other issues associated with the early implementations of *Daubert* in Federal courts (Harvard Law Review, 1994), citing cases where exclusion or otherwise of evidence has occurred through a trial judge's misapplication or misunderstanding of the *Daubert* factors. In particular, the review noted difficulties with reconciling statistics in the courtroom, concerns with addressing relevancy as a barrier to admission of evidence, the confusion surrounding reliability and scientific validity as distinct concepts, and the danger in relying on superficial criteria rather than substantive evaluation of scientific concepts.

Addressing some of these concerns, numerous articles implored that the courts made use of the privilege allowed by Rule 706, which gives judges the authority to appoint an expert for themselves in order to assist them in understanding any aspect of the evidence before them (Fenner, 1996; Maskin, 1994). Deason (1998) noted that there is always the potential for bias, among other undesirable consequences, in the use of court-appointed experts and points out that this would need to be managed carefully by the trial judge, however, Cecil and Willging (1995) cited numerous cases where this method had proven helpful, and developed potential solutions to some of the concerns that judges perceived when applying this discretion.

## Relevancy, Review and *Joiner*

Amid these criticisms, the Supreme Court soon heard another case that gave it the opportunity to comment further on the admissibility of scientific evidence, that of *General Electric Co. v Joiner* (1997). The plaintiff in this case, Robert Joiner, was an electrician who in the course of his employment was exposed to a dielectric coolant later discovered in some instances to have been contaminated with polychlorinated biphenyls (PCB's).<sup>15</sup> In 1991 he developed small cell lung cancer which he claimed was due exposure to contaminated coolant during his time working for the Water and Light Department in Thomaville.<sup>16</sup> The district court dismissed Joiner's claim, which was supported by an expert who's studies demonstrating a link between cancer and PCB's involved tests on rats using massive amounts of the chemical agent, and granted summary judgement to General Electric. However, on appeal the 11<sup>th</sup> Circuit undertook its own review of the scientific evidence presented to the district judge and reversed the decision, on the basis that the testimony of

---

<sup>15</sup> These chemical substances had been banned since 1978 due to health concerns associated with their use.

<sup>16</sup> Coincidentally, and perhaps also relevant, was the fact that Robert Joiner was also a smoker, with a history of smoking and lung cancer in his immediate family.

the plaintiff's experts had been improperly excluded. On reaching the Supreme Court, the 11<sup>th</sup> Circuit's ruling was overturned, and the decision made by the district court in favour of General Electric was re-instated.

This case illustrates two other important aspects of scientific evidence assessment relevant in the United States today. Firstly, the Supreme Court controversially commented on the consideration of expert *conclusions* in assessing proffered testimony. The court stated: 'nothing in *Daubert* or the FRE requires a court to admit opinion evidence which is connected to existing data only by the *ipse dixit* of the expert...conclusions and methodology are not entirely distinct'. This affirmed that courts are not bound to admit expert evidence if the trial judge feels that there is simply too great an analytical gap between expert's opinion and the existing data offered in evidence<sup>17</sup>. Several commentators perceived this as contrary to a section of *Daubert* opinion that had formerly embraced a clear distinction between an expert's methodology and conclusions, entitling the trial judge only to assess the former.<sup>18</sup> *Joiner*, they claimed, had now re-written *Daubert*'s original intent, a move that was perceived as misguided and unnecessary. Some cited the original methodology-conclusion distinction as the prime focus of the *Daubert* opinion, fundamentally enabling judges to carry out effective gatekeeping (Chesebro, 1994). Saks believed that *Daubert*'s distinction was far more sensible, and that the *Joiner* opinion left the way open for testimony to be excluded on the basis of simply not agreeing with the conclusions of the data, whereas it should properly be admitted if examination of the methodology in reaching the conclusions proves reliable (Saks, 2000).

Other commentators noted that judges now faced the additional hurdle of distinguishing between methodology and conclusion, before then having to choose how far to extend their reliability inquiry (Brown et al., 1999). Gottesman (1998) argued that as a result of *Joiner*, district judges now had far too much discretion, and foresaw the outcome of toxic tort cases in particular, resting 'on the personal prejudices of judges, rather than law'.

The second issue regarding expert testimony addressed in *Joiner* was the authority of appellate courts to over-rule trial judges' decisions on admissibility. The Supreme Court affirmed that a trial judge's decision on the admissibility of expert evidence was not to be reversed by an appellate court carrying out their own independent assessment, except in the

---

<sup>17</sup> In *Joiner*, the plaintiff's expert used data based on animal studies using massive amounts of PCB's injected intra-peritoneally. The Supreme Court ruled that this bore too little relevance to the facts of the case where humans were exposed to trace amounts, thus finding that the district court properly excluded the testimony.

<sup>18</sup> "...the focus, of course, must be solely on principles and methodology, not on the conclusions that they generate..." (*Daubert v Merrell Dow Pharmaceutical Inc.*, 1993 at 595)



case of abuse of discretion.<sup>19</sup> This aspect of the *Joiner* ruling also caused considerable controversy; with authors claiming that failure to allow appellate courts to re-examine admissibility determinations had the potential to result in significant discrepancies between jurisdictions on admissibility of the same evidence. Gottesman predicted a variety of other issues sustained by the *Joiner* opinion apart from inconsistent rulings on identical issues, including the potential for parties to engage in legal forum shopping; problems with obtaining impartial expert witnesses; appellate court manipulation of lower court rulings; and changes in state-based legal legislation to counter application of these rulings (Gottesman, 1998).

### Scope of Applicability and *Kumho Tire*

In most commentators' view, *Joiner* had failed to address some of the more obvious concerns with *Daubert*. The applicability of the *Daubert* decision had been under debate for some time, with various courts and scholars insisting that it only applied to 'scientific' testimony (Hrabosky, 1999). This resulted in the continued admission of evidence that would undoubtedly have failed the *Daubert* test, but because the discipline was not considered 'scientific', such scrutiny did not apply<sup>20</sup>. Even Congress endorsed this approach via a bill introduced in 1995 that proposed amending Rule 702 so as to limit the applicability of the rule to scientific testimony only. Conversely, other commentators maintained that *Daubert* was applicable to all expert testimony, whether meeting strict definitions of 'scientific' or not (Faigman, 1995). Several disciplines, particularly those in the forensic sciences, went so far as to declare themselves 'non-scientific', in order to escape the scrutiny of a *Daubert* challenge<sup>21</sup>. Inconsistencies in applying *Daubert* grew more alarming as judges and scholars wrestled with their various definitions of 'science' and, if found not to meet such definitions, what tests, if any, would then apply to assessing that evidence. Slovenko (1998) noted that even circuits who agreed on the status of a discipline as science or non-science still ended up with varying results for the same discipline by their use of differing approaches in order to establish reliability.

---

<sup>19</sup> Abuse of discretion is defined as 'A failure to take into account proper consideration of facts or law relating to a particular matter, or an arbitrary or unreasonable departure from precedent and settled judicial custom', and is considered a most serious claim. The threshold for establishing proof of abuse of discretion is generally considered to be very high. In *Joiner*, the Supreme Court ruled that there was no evidence that the district court abused its discretion, hence in their opinion the appellate court erred in conducting its own inquiry into the expert's evidence.

<sup>20</sup> One of the most often cited cases in this regard is *US v Starzecpyzel* (1995).

<sup>21</sup> The International Association of Arson Investigators petitioned the court not to apply the *Daubert* test to arson investigation testimony because, among other reasons, it did not involve 'scientific' analysis (International Association of Arson Investigators).

The Supreme Court undertook this issue in what was to become the third case in the *Daubert* trilogy: *Kumho Tire Co. v Carmichael* (1999). Following an automotive accident in which a tire blew out, injuring three passengers and killing another, the plaintiff filed suit against the tire company claiming that the tire was defective, causing the accident. The plaintiff's expert testimony was based on his visual and tactile inspection of many tires, and on his underlying theory that in the absence of at least two of four physical symptoms indicating abuse, the tire failure was due to a manufacturing or design defect. The trial court excluded this testimony, as in their opinion it did not meet the requirements set out in Rule 702 of the Federal Rules of Evidence (as interpreted under *Daubert*), having not been subject to peer review or testing, and lacking an acceptable methodology. The appellate court disagreed with this exclusion, as in their opinion *Daubert* only referred to 'scientific evidence', and given the expert's testimony in *Kumho* was skill and experience based, any application of the *Daubert* test to his evidence was erroneous. Following the reversal of the trial court's decision, the defendants appealed to the US Supreme Court requesting a review of the 11<sup>th</sup> Circuit's decision. The Supreme Court held that *Daubert* 'may apply to testimony of engineers and other experts who may not be considered, strictly speaking "scientists"', further clarifying that the language of Rule 702 applied to *all* expert testimony.<sup>22</sup>

The Supreme Court also recognised that the consequences of such a wide scope in assessing evidentiary reliability meant that the factors suggested in *Daubert* would not necessarily be applicable in every case. Their opinion stated further; '[A] trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of the expert testimony... whether *Daubert's* specific factors are, or are not reasonable measures of reliability in a particular case is a matter that the law grants the trial judge broad latitude to determine' (*Kumho Tire Co. v Carmichael*, 1999). Justice Scalia, however, tempered this with a warning in his concurrence that while these factors were not 'holy writ', failure to apply one or more of them in a particular case may constitute an abuse of discretion.

Most scholars agreed that *Kumho* had now levelled the playing field for all experts (Stilwell, 2000). Many saw *Daubert* and *Kumho* as having addressed some of the inherent flaws in the common law, acting specifically to re-focus the scrutiny of expert testimony on the two most important criteria, reliability and necessity (Imwinkelreid, 1999). Critics of the *Kumho* decision, however, cited the difficulty of applying specific factors to non-scientific testimony, which by its nature encompasses an enormous range of methodologies and philosophies (Hrabosky, 1999). The same author also criticized Imwinkelried's early attempt

---

<sup>22</sup> The district court's verdict was thus re-instated.

(Imwinkelreid, 1993) at establishing a list of factors to assist in scrutinising non-scientific expertise for, among other things, being nowhere near as specific and helpful to the trial court as the *Daubert* factors. Other commentators criticised Justice Breyer's use of the words 'broad discretion', and argued that allowing such latitude at the trial court level would continue to have a negative impact on the consistency of procedures and rulings (Goodwin, 2000).

The *Kumho* court also stated that trial judges, when performing their reliability assessments, should make certain that '...an expert, whether basing testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field'. This paragraph apparently tethers the standard for admissibility of expert testimony to the standards of professional practice in that field (Cecil, 2005), which, suffering an unfortunate resemblance to one of the main criticisms of *Frye*<sup>23</sup> served only to cause further confusion for some courts. Several jurisdictions began applying the standard of professional practice to the standard required in court as the *only* test of admissibility, altogether ignoring the reliability enquiry as proposed in *Daubert* (Faigman, 2000; Saks et al., 2009). Most courts, however, did not endorse such a literal reading of Justice Breyer's opinion, and realised the intention was to include this as a factor that may be considered among many.

## Expert Evidence in the United States Today

The Advisory Committee on the Federal Rules was established in 1992 to engage legal and public comment on the Rules, and to make recommendations that ultimately pass to the Judicial Conference for consideration in presentation to Congress.<sup>24</sup> While having no formal authority to amend the Rules themselves, the Advisory Committee has considerable influence on their form and content. Any public member is entitled to suggest reform to the Advisory Committee, and two influential lobby groups submitted proposals prior to the year 2000 amendments. The Evidence Project suggested a comprehensive amendment to both Rule 702 and 703, with the proposed Rule 702 articulating the qualifications required to

---

<sup>23</sup> A careful interpretation of this section of the *Kumho* opinion might lead one to conclude that if the discipline itself applies less than rigorous standards, then the Court may only apply that same level of rigour in its assessment of reliability. This would lead to disciplines with a high degree of scientific rigour being held to a high standard, and those with a lower degree being held only to their lower standard. This is a fundamental problem in assessing reliability of a discipline using only its own professional standards as a benchmark for the level of enquiry, and is essentially the same criticism found in the application of *Frye*.

<sup>24</sup> The Judicial Conference is formed at the behest of Congress, is chaired by the Chief Justice of the United States and consists of judges from each of the federal appellate and district circuits.

testify as an ‘expert’, and the proposed Rule 703 devoted to explicitly establishing the standard of reliability that their evidence should meet (The Evidence Project, 1998). This submission was not favored by the Advisory Committee and was largely ignored, along with similar suggestions for reform by the National Conference of Commissioners of Uniform State Laws (Reporter's Notes, 1999). Largely in response to the *Daubert* and *Joiner* cases, Rule 702 was amended in the year 2000 in accordance with the Advisory Committee’s recommendation. It now reads:

If scientific, technical or other specialised knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, expertise, training or education may testify thereto in the form of an opinion or otherwise, provided that: 1) the testimony is sufficiently based on reliable facts or data; 2) the testimony is the product of reliable principles and methods; and 3) the witness has applied the principles and methods reliably to the facts of the case. (Federal Rules of Evidence)

By this amendment, Congress have virtually nullified any further debate on the applicability of the Federal Rules to non-scientific opinion, as Rule 702 now clearly articulates that any scientific, technical or other specialised knowledge is to be subject to a reliability and relevance evaluation<sup>25</sup>. Additionally, the amendment emphasises the fact that judges have discretion in deciding *how* they perform the gatekeeping function, but not *whether to perform it* or not. Assessment of the reliability of expert testimony, at least in some form, is essential prior to allowing evidence that may have bearing on the case before them to be heard. The Advisory Committee also noted at the time of the amendment that a review of the case law after *Daubert* has demonstrated that the rejection of expert testimony is still the exception rather than the rule in the United States (Advisory Committee on the Federal Rules, 2000).

Berger noted in 2005 that data supporting the notion that the *Daubert* has led to better expert proof is still lacking (Berger, 2005). Rice is highly critical of both the Rule 702 amendment and the Advisory Committee’s general attitude towards Rule reform, arguing that the Rules need to provide judges with explicit guidance on how to conduct a reliability assessment, as well as guidance on how high the ‘standard of reliability’ should be (Rice, 2000). He and his colleague further lamented that Rule 702 ‘still offers no guidance for determining reliability of an opinion, other than the tautology that the opinion must be based on “reliable facts or data”, and “reliable principles and methodologies” that are “reliably applied.”’ (Rice and

---

<sup>25</sup> The criterion of relevance is articulated specifically by Rule 401, but is also implied to in Rule 702 by the use of the words ‘*if ...[such knowledge] will assist the trier of fact*’.

Delker, 2000). The same authors argue that the Advisory Committee's attempt to codify *Daubert* is a fundamental failing, stating that '...the Supreme Court's interpretation of a poorly written rule should not define the parameters of how expert testimony should be screened.'

It appears, however, that there is relatively little evidence to support the feared notion that courts are still misinterpreting the *Daubert* factors as 'the only' criteria for admissibility, as some were accused of doing in the mid-1990's (Saks et al., 2009). Some studies have been critical of the finding that judges appear to be referring to the *Daubert* factors, rather than technically exercising them, hypothesising that this is perhaps due to poor comprehension or lack of skills in applying them (Groscup et al., 2002). While there may be some merit in these interpretations, there has been an unmistakable realisation amongst the judiciary since *Kumho* that the *Daubert* 'factors' themselves do not constitute a definitive checklist, should not be given undue weight, and may not be appropriate for the assessment of some disciplines. The Advisory Committee also noted that '...courts both before and after *Daubert* have found other factors relevant in determining whether expert testimony is sufficiently reliable to be considered by the trier of fact. Some of these factors have included: (1) whether experts are testifying about matters that arose in his research independently of litigation;<sup>26</sup> (2) whether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion;<sup>27</sup> (3) whether the expert has adequately accounted for obvious alternative explanations;<sup>28</sup> (4) whether the expert is being as careful as he would be in his regular professional work outside his involvement in the litigation process;<sup>29</sup> and (5) whether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give.<sup>30</sup> All of these factors remain relevant to the determination of the reliability of expert testimony under the amended Rule 702, with no single factor supposedly being necessarily dispositive of the reliability of a particular expert's testimony' (Advisory Committee, 2000).

A persistent inconsistency in courts today is that some are interpreting *Kumho* as a definitive requirement for a discipline to carry out testing where it can be done, whereas others are maintaining that 'testing' is not necessarily part of the reliability criteria per se (Saks et al., 2009). However, when an expert relies on experience as a basis for his or her testimony, most courts do realise that as a bare minimum, at least some form of explanation will be

---

<sup>26</sup> *Daubert v Merrell Dow Pharmaceuticals, Inc.* (1995).

<sup>27</sup> *General Electric Co. v Joiner* (1997)

<sup>28</sup> *Claar v Burlington N.R.R.* (1994)

<sup>29</sup> *Sheehan v Daily Racing Form, Inc.* (1997), *Kumho Tire Co. v Carmichael* (1999)

<sup>30</sup> *Kumho Tire Co. v Carmichael* (1999)

required as to how that experience provides a foundation for their opinion in order for it to be admissible (Saks et al., 2009). As another author notes, the difficulty the courts will face by insisting on a testability requirement is that many theories, including those forming cornerstones of modern science today (such as evolution, and the theory of relativity) are not genuinely testable, as they have arisen through observation and agreement, rather than an empirical analysis (Beecher-Monas, 1998).

The situation where judges set their own reliability criteria, lauded by some as an indication of their ability to carry out the Supreme Court's task, is not necessarily desirable either as recent authors have pointed out. There have been several instances where trial judges have applied their own 'factors' to a particular discipline thinking they were appropriate for determining reliability, seemingly unaware that none of the criteria they proposed ultimately resulted in establishing 'reliability' or 'validity' in accordance with any scientific definition of these terms (Sales and Shuman, 2005).<sup>31</sup>

The difficulties of assessing evidence that straddles the traditional boundaries of science and non-science are yet to be resolved, with the higher judiciary reluctant to provide specific standards and procedures to guide trial judges in this regard, despite advice by some authors to do so (see DeCoux, 2007). Numerous commentaries on *Kumho* have noted that it specifically avoided setting the goal of scientific validity for non-scientific evidence (Sales and Shuman, 2005), and the Advisory Committee on the Federal Rules have also been reluctant to provide any more specific, codified guidance than that currently proposed in Rule 702, stating after the year 2000 review that '...[we will] make no attempt to set forth procedural requirements for exercising the trial court's gatekeeping function over expert testimony' (Advisory Committee, 2000).

Consequently, there are still several concerns over the inconsistent nature of judicial reliability enquiries, however, the Federal Judicial Center attempted to remedy the situation by offering more explicit guidance through the publication of the *Reference Manual on Scientific Evidence*. First published in 1994, this document was authored as an attempt to provide judges with 'quick access to information...that will be useful in dealing with disputes among experts' (Federal Judicial Center, 1994), and following the rulings in *Joiner* and *Kumho* an updated second edition was released in the year 2000. A published collection of opinion, online comments, and conference proceedings reflected the fact that this resource

---

<sup>31</sup> The authors cite *Antoine-Tubbs v Local 513 Air Trans. Div.* (1998) as one example where none of the seven criteria set by the trial judge actually established scientific validity, despite the court believing it had done so.

is generally well utilised by judges, with publication of third edition now in print (Committee on the Evaluation of the Reference Manual on Scientific Evidence, 2009).

The implications of the *Daubert* trilogy on resources and finances are yet to be studied in any detail, but it appears that funding for both judicial and non-judicial expertise is a critical factor in deciding how accurately the *Daubert* trilogy is applied in accordance with its original intent. Surveys have revealed that many judges who cite problems with applying *Daubert* in its current format do so due to time and monetary constraints, rather than an innate inability to conduct a reliability enquiry. A small subset apparently doubt their own ability to effectively evaluate reliability claims under the new requirements when faced with complex expert evidence, however this factor has not yet been demonstrated as influential concerning the number of motions filed or cases decided involving expert testimony.

*Daubert* was a federal case, and so as precedent was only binding in federal jurisdictions. Many states continued to use *Frye* interpretations of their individual evidence statutes until long after *Kumho Tire*, however, this conflict between State and Federal interpretation of expert evidence requirements appears to have eased, with the gradual adoption of state evidence rules conforming to those articulated in the Federal Rules of Evidence, which now embody a *Daubert*-style approach. Even in the minority of US states that still specifically affirm *Frye*, the use of general acceptance as the sole criteria for assessing expert evidence appears to be diminishing. In this regard, *Daubert* can actually be seen to have shifted the focus generally from *Frye* to the Federal Rules of Evidence, rather than onto *Daubert* itself, with most empirical studies suggesting that increased judicial scrutiny seen has been more in line with the Rules' general thrust towards a reliability focus, rather than following *Daubert*'s specific reliability criteria (Groscup et al., 2002).

What this exploration of the history of *Daubert* tells us is that the case itself cannot be considered in isolation, and held as a pinnacle of expectation for expert evidence in the United States or elsewhere. *Daubert* was simply one case among several that have attempted to define the expectations of expert evidence. While it offers useful insight and tentative suggestions for ensuring expert witnesses offer reasonable claims in court, it is ultimately the Federal Rules that are the primary authority for admission of expert testimony. Since the amendment in the year 2000, these mandate a reliability assessment for all expert testimony, but offer little in the way of suggestion as to how this is to be conducted. *Daubert* provides useful guidance in this regard, but it is by no means, by itself, a compulsory aspect of expert evidence assessment.

Numerous authors have suggested that *Daubert* is widely misapplied by courts (see Roisman, 2009), and if this is indeed the case, it becomes even more important for forensic practitioners themselves to undertake the responsibility of ensuring their theories and techniques are valid. A recent study involving interviews with judges and lawyers discussed the appearance of a phenomenon coined as the ‘*Daubertization*’ of experts, which occurs when an expert has testified so many times that they are no longer adequately scrutinised in court, forcing many attorneys to actively seek new experts for their clients (Waters and Hodge, 2005). Just how well courts are applying the intent of *Daubert* and the Federal Rules, specifically Rule 702, to forensic identification science evidence is an essential question that has so far remained largely unaddressed, and is the focus of the next chapter in this thesis.



## CHAPTER 3

# Forensic Identification Evidence since Daubert

---

Jonakait (1993) noted that the real question regarding the *Daubert* decision was not whether advocates and academics could interpret *Daubert* to justify the results they desired, but whether judges would do the same. He predicted that if *Daubert* was taken seriously, much of forensic science would be in trouble. Many legal academics envisioned *Daubert* and *Kumho* as the answer to their call for the use of ‘evidence-based’ forensic science. Most assumed that a large proportion of forensic evidence presented in criminal and civil courts across the country would now be subject to renewed judicial scrutiny.<sup>1</sup> Anecdotally, it appears that such examples have remained few when considering the many thousands of cases involving forensic identification evidence that have appeared before the courts. This chapter begins by exploring the practice of admission of forensic science evidence since the reliability mandates of FRE 702, *Daubert* and *Kumho*. It then presents the results of a large-scale retrospective study of challenges to the admissibility of forensic identification science since the *Daubert* decision.

The majority of empirical studies have indicated that judges are certainly taking a more active role in scrutinising expert evidence since the *Daubert* decision (Krafka et al., 2002)<sup>2</sup>. A 2001 RAND<sup>3</sup> report addressing trends in evidence admissibility in civil cases concluded that there was strong evidence that the *Daubert* opinion had changed the way judges assessed expert evidence, following an analysis of 399 federal district opinions from 1980 to 1999 (Dixon and Gill, 2001), although the authors themselves admitted that they were unable to determine whether outcomes had actually improved as a result.

---

<sup>1</sup> It is worth noting that such an examination of forensic science evidence has not been a wholly post-*Daubert* phenomenon. For example, *United States v Parks* (1991) represents a challenge under California’s *Kelly/Frye* principles (that requires firstly a demonstration of general acceptance of the technique in the relevant scientific community, followed by establishment of the reliability of the method) that resulted in the exclusion of fingerprint evidence due to a lack of evidence of any ‘science’ behind fingerprinting identification long before *Daubert* was extant.

<sup>2</sup> It is worth noting that virtually no empirical data on the application of *Daubert* was published before the year 2000, some 7 years after the Supreme Court’s ruling.

<sup>3</sup> The Research and Development Corporation (RAND Corporation) is a non-profit organization that is sponsored by private endowment in addition to the US Government and offers research and analysis to a variety of clients in order to ‘help improve policy and decision making through research and analysis’.

There was a significant increase in the number of cases where reliability was addressed after the *Daubert* opinion, but a decrease after July 1997. The proportion of evidence found unreliable also decreased in addition to the overall number of evidentiary reliability reviews after this time. Several authors who have commented on the RAND results theorise that this is due to parties responding to the new standards (Cecil, 2005; Dixon and Gill, 2001) however, also noted is that following a dramatic increase in use of the *Daubert* criteria in the period 1995-1997, a significant drop in the use of these five factors was seen in the period 1997-1999. The RAND analysis demonstrates that this was countered by an increase in the use of other criteria used to assess reliability, and by way of Dixon and Gill's explanation, demonstrates that the judiciary were perhaps becoming more competent in their assessment of evidence in civil cases through the use of other potentially important criteria, rather than rigidly following those proposed in *Daubert* (Dixon and Gill, 2001). An alternative theory — that of a slackening in the formerly rigorous assessments of scientific evidence — is not discussed by these authors, but is alluded to by others (see Saks and Faigman, 2005).

Some of the earliest data on judges' ability to actually perform a 'gatekeeping' role comes from Shuman and colleagues. This study was conducted from 1990 to 1991, and offers a contemporaneous pre-*Daubert* baseline. It failed to specifically address the question of how well judges can evaluate scientific evidence, but suggested that judges generally had greater confidence in their ability to scrutinise expert testimony, compared to that of laypersons (Shuman et al., 1993).<sup>4</sup> Kent (1999) later illustrated that while the majority of judges thought they could perform the role of gatekeeper adequately, more than 70% also reported limited or potentially out-dated education or experience with scientific methodology. Similarly, a 2001 survey that included questions regarding State judges' specific understanding of the *Daubert* criteria revealed less than 6% of judges conveyed an appropriate scientific definition of falsifiability, and only 4% demonstrated appropriate application of the concept of error rate (Gatowski et al., 2001). Another study concerning psychological evidence testimony also demonstrated that judges seemed overconfident in their self-assessments of how well they can identify invalid scientific methods (Kovera and McAuliff, 2000). Beecher-Monas (2007) argued nonetheless that judges *are* capable of providing the function of gatekeeper, providing they educate themselves about the evidence before them, and give valid reasons based on justifiable scientific and policy grounds for their decisions.

---

<sup>4</sup> While juries have been noted to have difficulties assessing expert testimony (Krauss and Sales, 2001), more recent studies have reached the conclusion that jurors appear not to differ greatly in their view of expert testimony when compared to the judiciary (see Heinzerling, 2006 pp.74-75), and that they generally perform better than they have previously been given credit for when assessing expert opinion (Diamond, 2008).

## The Admission of Forensic Identification Evidence

Moriarty and Saks (2005) have noted that ‘some forensic sciences have been with us so long, and judges have developed such faith in them, that they are admitted even if they do fail to meet minimal standards under *Daubert*.’ Elsewhere, Saks has commented that the judiciary engage in all forms of ‘judicial gymnastics’ in order to find such evidence admissible (Saks, 2009a). Saks has also characterised several ‘techniques’ by which the judiciary manage to shirk their gatekeeping duties (Saks, 2002), including; by refusing to hold an admissibility hearing; misapplying or misinterpreting *Kumho*; avoiding giving reasoning in a *Daubert*-style analysis; reversing the burden of persuasion onto the challenger; relying on general acceptance; relying on admission by other courts; over-emphasising the flexibility of the inquiry; bringing the standard down to meet the expertise, or; relegating the decision to one of weight rather than admissibility.

Risinger’s appendix appearing in the *Tulsa Law Review* considers 67 handwriting cases ruled on since *Kumho*, and he describes a good proportion of them finding handwriting admissible in such terms as ‘unsatisfactory’, ‘careless’, and ‘troubling’ (Risinger, 2007a). Risinger also describes the Sixth Circuit opinion in *United States v Jones* (1997), finding handwriting evidence admissible, ‘...one of the worst crafted and reasoned I have ever encountered in my forty plus years of reading opinions’ (Risinger, 2007b). La Morte (2003) has also analysed the ways in which judges admit evidence that, if properly analysed, should probably not have passed *Daubert* and *Kumho*’s requirements.

One such case characterising the ‘judicial gymnastics’ alluded to by Saks is that of *Llera Plaza I* and *II*. In *United States v Llera Plaza I* (2002), Judge Louis Pollak partially excluded fingerprint evidence on the grounds of its scientific unreliability following a *Daubert* inquiry, however this decision was ultimately reversed two months later, by the same judge, in *United States v Llera Plaza II* (2002), when he retracted his earlier ruling and admitted the testimony without restriction. Detailed analysis of these decisions has been undertaken by several legal scholars, who find much fault in Pollak J’s reasoning to reverse his initial decision to exclude fingerprint testimony in this case (Cole, 2004a; Faigman, 2009; Kaye, 2003b).<sup>5</sup>

---

<sup>5</sup> Including details such as; that much of the testimony offered in support of fingerprint analysis was based on scientifically flawed evidence; instances of deference to the British admission of fingerprints; conflating the individuality and individualisability of ten-fingerprint sets with that of single or partial prints; that fingerprint analysts are technicians rather than scientists does not alter the fact that there are no scientific studies (and hence there is no general acceptance in the scientific community) of attributions of identity based on ridgeology; and the flexibility of enquiry that *Kumho* states refers to the discretion to consider other indicia of validity besides those given in *Daubert*, and

Similarly, the appellate judge's ruling in *Deputy v Lehman Bros., Inc.*, (2003) reversed the trial judge's earlier exclusion of handwriting evidence. Among the seven reasons given by the trial judge for this particular exclusion were that the expert could not identify a set of scientific principles or standards that she applied in this case, and that she failed to adequately explain the inconsistencies in her reasoning process. The Seventh Circuit overruled the trial judges opinion, feeling that the trial judge had concentrated on the assessment of factors that were not applicable to handwriting analysis. While it did not subsequently rule handwriting admissible, it directed the District court to conduct a second analysis, on remand, '...in light of *Daubert* and this opinion'.

*United States v Mitchell* (2004) represents another case where multiple examples of Saks' 'judicial gymnastics' occur. In a rather unfortunate twist, this challenge resulted in the exclusion of witnesses attempting to testify *against* the use of fingerprints, rather than the exclusion of the fingerprint examiners evidence. The *Daubert* hearing, conducted before the original jury trial (*United States v Byron Mitchell*, 1999) was conducted over five days, running into nearly one thousand pages of transcript, and represents one of the most comprehensive attempts to analyse a discipline of forensic identification science under *Daubert* and Rule 702 to date.

Two months after the *Daubert* hearing concluded, the District Court denied the defendants motion to exclude the government's fingerprint evidence. The Court found that the governments latent fingerprint expert witnesses were capable of testifying in the trial proceedings, and subsequently, also agreed that the defence could call latent fingerprint expert witnesses to testify as the ability to make an identification. However, he also added; 'I am going to exclude evidence as to whether or not [latent fingerprint identification] is scientific, technical, or whatever. It has no relevance before the jury here. The question is whether or not an identification can be made by the examination of fingerprints – latent fingerprints.' Furthermore, it was ruled that 'This Court would take judicial notice that human friction ridges are unique and permanent throughout the area of friction ridge skin, including small friction ridge areas'.

In 2004, Byron Mitchell appealed to the Third Circuit challenging the trial judge's ruling in the 1999 *Daubert* hearing. Re-examining only the evidence presented in the *Daubert* hearing, the Circuit ruled that the government's fingerprint evidence clearly passed muster. Furthermore, in response to Mitchell's claims that the government was unable to

---

not, as Judge Pollak appears to have used, the discretion to ignore the fact that a lack of scientific research into a discipline should simply 'not count' towards arguments against judicial acceptance.

demonstrate that its evidence was correct, and failed to meet the standards required of ‘science’, the 3<sup>rd</sup> Circuit noted that *Daubert* requires ‘no more than that the Court satisfies itself that “good grounds” exist for the expert’s opinion’. Good grounds for admission plainly exist here.’ This reasoning rests largely on what Saks characterises as ‘lowering the standard to meet the expert’.

The court went on to clarify that, in their opinion, the existence of the *Daubert* factors acts as a guarantee that cross-examination and adversarial testing of the evidence will be possible, by the providing of raw material for confrontation of the expert via peer review and publication, forcing the expert to admit to the limitations of their methods under cross-examination. The court wrote: ‘*Daubert* did not require the party who proffered the expert testimony to carry the burden of proving to the judge that the expert’s assessment of the situation was correct – that was to be tested by the adversary process’.<sup>6</sup> Here, the court essentially sheds its gatekeeping responsibility, by stating that it is now up to the adversarial system, via examination and cross examination, to further filter any evidence that may not pass muster, in direct contradiction to *Daubert*’s mandate that the *trial judge* bears responsibility for ensuring that evidence is not only relevant, but reliable.<sup>7</sup>

The Third Circuit in *Mitchell* also agreed with the trial judge’s exclusion of any evidence attempting to make a distinction between ‘scientific knowledge’ and ‘technical or other knowledge’, finding that such evidence was unworkable and unnecessary as, per *Kumho*, Rule 702 applied to all categories of expert testimony. ‘That a particular discipline is or is not “scientific” tells a court little about whether conclusions from that discipline are admissible under Rule 702.’ This is an example of what Saks characterises as ‘turning *Kumho* on its head’. The Seventh Circuit engaged in similar reasoning in *United States v Havvard* (2001) when it stated that the “...standards of *Daubert* are not limited to scientific theory alone...therefore the idea that fingerprint comparison is not sufficiently scientific cannot be the basis for exclusion under *Daubert*’. In both these instances, the judges have used *Kumho*’s directive against its initial purpose. Saks has found that these rulings ‘irresolvably incoherent’, and wrote that ‘*Kumho* did not open new escape hatches, it closed old ones.’ In these examples, the judges have used *Kumho* to remove the questioned expertise from *Daubert*’s reach, representing the opposite of *Kumho*’s intention.

---

<sup>6</sup> Quoting from *Ruiz-Troche v Pepsi Cola Bottling Co* (1998), citations omitted.

<sup>7</sup> Consider: ‘Nor is the trial judge disabled from screening such evidence. To the contrary, under the Rules the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.’ (*Daubert v Merrell Dow Pharmaceuticals Inc.*, 1993)

However, taking advantage of a further loophole, the Third Circuit declined to agree that Mitchell's experts themselves were actually excluded by the District Court's bench ruling, or that the scope of their testimony was improperly limited. Essentially, they claimed that Mitchell's attorney's misinterpreted the District Court's ruling, as the court never specifically ruled that Dr Cole and Professor Starrs could not testify, it simply ruled that a specific *type of evidence* was not going to be allowed.<sup>8</sup> Although the defence counsel assumed that the former situation was the case, they failed to ask any particular question regarding the admissibility of these two witnesses, and thus the appellate court saw that no error was committed by the trial judge. They noted that 'to the extent that the record was ambiguous, the onus was on Mitchell's counsel to make a clear record, especially given the multiple, nuanced categories of testimony being discussed'.

In an example of 'reversal of the burden of persuasion', the defendant in *Havvard* (2001) also argued that the court improperly asked him to disprove the validity of fingerprint evidence at the *Daubert* hearing, rather than requiring the government to identify a scientific basis for fingerprint analysis, when it used the reasoning that he had not offered any evidence to undermine the reliability of the fingerprint evidence. Havvard's appeal on these grounds was dismissed by the Seventh Circuit, who responded curtly with; 'We think Havvard reads too much into the district court's observation that he had "offered no evidence in this case undermining the reliability of the methods in general"', and in only those few words dismissed this claim. Another instance can be found in *United States v Cromer* (2006), where the judge more blatantly stated; 'The Court concludes that Defendant [*sic*] has not put forth evidence that fingerprint identification is *not* scientifically reliable.'<sup>9</sup> McConnell J in *People v Partee* (2008) said that 'An individual, such as Partee, who seeks to challenge such a long-recognized scientific technique bears the burden of showing that the evidence is not scientifically reliable. Such a burden requires more than the vague claims and speculation raised here.' As Saks noted, the simplest reading of these opinions is that the 'courts have expected the opponent to prove that the proposed evidence did not meet [the admissibility requirements], rather than requiring the proponent prove that it did. [This] is erroneous application of the law' (Saks, 2002).

Several cases have used 'general acceptance via acceptance by the courts' as reasoning for the admissibility of forensic evidence. In *United States v Paul* (1999) the Eleventh Circuit ruled that the defendant's argument that handwriting analysis did not qualify as reliable

---

<sup>8</sup> *Viz.* 'I am going to exclude evidence as to whether or not [latent fingerprint identification] is scientific, technical, or whatever.'

<sup>9</sup> Emphasis added.

scientific evidence was without merit, as courts have long received handwriting analysis testimony as admissible.’ In *United States v Prime* (2002) the admission of handwriting testimony withstood *Daubert/Kumho* analysis by relying heavily on its historical acceptance by law enforcement and English and American courts, and in *State v Brewer* (2005), the court found that “[t]he testimony of the state’s expert with regard to firearms and ballistics is so well established [in the judicial system] that it does not require analysis under the state’s *Daubert*-type rule”.

Again, in *United States v Havvard* (2001) the Seventh circuit noted that ‘[t]he District court recognised that establishing the reliability of fingerprint analysis was made easier by its 100 years of successful use in criminal trials’ and ‘it properly considered the *Daubert* factors and concluded that fingerprinting techniques have been tested in the adversarial system’. Of note in these trials is that many of them cite cases prior to the adoption of *Kumho* as evidence for the admission of the evidence. Before *Kumho*, many admissibility decisions were based on the finding that handwriting testimony was not scientific, constituting ‘technical or other specialised knowledge’ under Federal Rule of Evidence 702. Much forensic testimony thus escaped entirely the rigorous scrutiny that *Daubert* demanded of all expert testimony, by virtue of its ‘non-scientific’ status, hence their use as examples of admission are from an era where different standards applied.

However, consider Gertner J, in the opinion in *United States v Green* (2005) who felt that the reliance on long-standing use of ballistics evidence in the courts runs the risk of ‘grandfathering in irrationality’, without proper *Daubert/Kumho* analysis. Gertner also realised that this reasoning ignores the mandate of *Daubert*, particularly when courts rely on pre-*Daubert* acceptance of a given scientific technique. Other courts have also found that ‘judicial’ general acceptance does not qualify as general acceptance under *Daubert*. The US District Court (E.D. Kentucky) in *United States v Sullivan* (2003) held that ‘the historical acceptance of fingerprint evidence in courts does not qualify as general acceptance for the purposes of *Daubert*...’ and that ‘[t]he reliability of ACE-V is not demonstrated by its use in prior court cases’. The district court in *United States v Saelee* (2001), where handwriting evidence was excluded entirely, also noted that ‘...the fact that [handwriting] evidence had been admitted in the past does not mean that it should be generally accepted now, after *Daubert* and *Kumho*.’

Some courts invoke a narrow interpretation of the ‘task at hand’ requirement in *Kumho* in order to find evidence admissible. Yet by doing so, they avoid any *Daubert*-style analysis of the science itself, and instead focus on the qualifications and experience of the expert in

question. As an example, the judge in *US v Prime* (2002) believed that ‘...the *Daubert* inquiry is not intended to ask the "larger question" regarding the reliability of a particular technique in general. Rather, the inquiry is case-specific.’ Within the confines of this case, the judge had ‘no trouble concluding that the premises of handwriting identification are sound’. By adopting a ‘case-specific approach’, judges then focus on qualifications and experience of the expert, thereby avoiding any of the ‘reliability’ questions posed in *Daubert* or *Kumho*. The judge in *United States v Williams* (2007) similarly focussed on the individual expert’s qualifications and experience in order to find her evidence admissible.<sup>10</sup> Yet the fact that an expert witness is qualified, or has sufficient experience, still fails to address the question of whether evidence tendered by that expert is scientifically valid.

Some courts have found reasoning to avoid analysis as suggested by the US Supreme Court. The majority opinion in *United States v Crisp* (2003) took the view that the US Supreme Court could not have wanted the non-science forensic sciences to be excluded, and therefore, *Daubert*’s criteria were not to be considered as touchstones of admission. The court in *United States v Santiago* (2002) ruled similarly: ‘[I]t is this Court’s view that the Supreme Court’s decisions in *Daubert* and *Kumho Tire* did not call this entire field of expert analysis into question’.

Some courts avoid giving reasons for their admissibility decisions entirely, and thus it is left to one’s imagination as to why the judge actually decided to admit the evidence. In *United States v O’Driscoll* (2003), Muir J gave no explanation at all in his finding that ballistics evidence was reliable. Citing *United States v Santiago* (2002), he merely stated ‘...[b]ased on [the expert’s] testimony we are satisfied that the field of ballistics [*sic*] is a proper subject for expert testimony and meets the requirements of Rule 702...we see no reason to exclude the ballistics evidence that was presented at O’Driscoll’s 1984 trial for kidnapping.’ Others have denied the need for an evidentiary hearing altogether. In *United States v Cooper* (2000), another case often cited by others attempting to address the admissibility of forensic ballistics evidence, the court noted that ‘there is nothing in *Kumho Tire* or *Daubert* that requires the Court to conduct a pre-trial evidentiary hearing if the expert’s testimony is based on well-established principles.’ Deeming it ‘not in the interests of justice’ to conduct a

---

<sup>10</sup> The opinion states: ‘[t]he government provided an exhaustive foundation for [the witness’s] expertise, including her service as a firearms examiner for approximately twelve years; her receipt of “hands-on training” from her section supervisor; attendance at seminars on firearms identification where firearms examiners from the United States and the international community gather to present papers on current topics within the field; publication of her writings in a peer-reviewed journal; her obvious experience with toolmark identification; her experience examining approximately 2,800 different types of firearms; and her prior expert testimony on between 20 and 30 occasions’ (*United States v Williams*, 2007 at 161).



potentially ‘lengthy [evidentiary] hearing’, when the ‘admissibility of well-established scientific principles can be determined at trial’, the defendants motion to challenge expert ballistic testimony was denied.

The court cited *United States v Nichols* (1999), a case regarding the analysis of explosive material, which ruled that ‘no pre-trial hearing was required because the challenged evidence does not involve any new scientific theory and the testing methodologies were neither new or novel’, despite the precedent of *Daubert* occurring some six years earlier. Similarly, the court in *United States v Reaux* (2001) denied the defendants motion to hold a pre-trial hearing on the motion to exclude fingerprint evidence, and found that it had ‘the authority to avoid unnecessary reliability proceedings where the reliability of the methodology is properly taken for granted’, despite *Kumho*’s warning that this discretion does not warrant the complete abandoning of the gatekeeping function.

### **Empirical Data on the Admission *versus* Exclusion of Forensic Identification Science Evidence**

While specific judicial decisions provide useful insight into individual judicial reasoning, there have been relatively few studies that have attempted to characterise patterns of admissibility or exclusion for any given discipline. A post-*Kumho* study by Fradella and colleagues (Fradella et al., 2003) examined a total of 83 criminal cases, yielding 98 relevant claims where a forensic scientific issue was challenged under *Daubert*, yet only twenty-five of these concerned the ‘forensic identification’ sciences<sup>11</sup>. Detailed numerical analysis was not undertaken, however, the authors concluded that the courts overwhelmingly accepted the identification evidence. The exception to this statement concerned the admission of handwriting analysis, with the authors having noted that the expert testimony in these cases was usually limited as to the similarities and differences between the known and unknown exemplars, preventing the expert concluding as to the existence of a ‘match’.

Risinger (2007a) recently undertook a more detailed analysis of handwriting cases post-*Kumho*. Cases where no reliability challenges were brought forward were excluded, as were those where the challenge was based solely on a lack of training or experience. 67 cases formed the set from which a qualitative analysis of the admissibility decisions was undertaken. McFarland similarly considered a qualitative analysis of 66 handwriting identification cases, selected from 300 federal and state cases that have considered this type of evidence since *Daubert* (McFarland, 2004). This somewhat selective example, published on the author’s website, claims to demonstrate ‘complete defeat for foes of QDE’

---

<sup>11</sup> N(25) = 8 fingerprint, 11 handwriting, 1 age estimation (visual), 5 unspecified

[Questioned Document Examination], and states that in cases where the trial judge has excluded the evidence, this was based on ‘an evaluation particular to a unique individual’, and often due to a qualification issue. She concludes her introduction by mentioning that ‘Cases from states on the *Frye* standard are included to show that all jurisdictions are in harmony as to the ultimate reliability and admissibility of expert handwriting evidence.’

Haber and Haber (2008) cite the ‘FBI [*sic*] Onin website’ as listing over 40 cases where fingerprint evidence has been challenged under *Daubert*. This data set was also referenced by Tom Ferriola in his article, appearing in the *Criminal Law Bulletin* (Ferriola, 2004) as evidence for the reliability of latent fingerprint examination. It should be noted firstly that this website is not affiliated with the FBI, rather is maintained by Ed German, a latent fingerprint examiner and former head of the US Army Criminal Intelligence Command, ‘as a private endeavour for both public service and personal interest’ (German, 2005). Furthermore, the last update for the section on legal challenges to fingerprint evidence was in September of 2005, and was provided by Paul Sarmousakis, an attorney, and Stephen Meagher, an FBI fingerprint specialist. None of these cases resulted in an exclusion of fingerprint evidence, and it would also be difficult to describe this sample as being free from bias.

Moenssens (Moenssens et al., 2007) noted 41 cases concerning challenges to forensic odontology evidence from 1993, with few examples of limitation. Bowers (in Faigman et al., 2007) has also referenced 19 similar cases. Peter Nordberg, a shareholder in a Philadelphia based law firm who specialises in complex federal civil litigation, authored a website that attempted to track federal appellate decisions on the admissibility of expert testimony since 1 Jan 2001 (Nordberg, 2006). The website is no longer current, as the most recent case listed was 15 Sep 2006, however, despite the fact that the sample is not an exhaustive representation of *Daubert* challenges to forensic science, the case summaries, with links to original opinions for many of them, provides useful insight into the workings of *Daubert* in the forensic context.

Of the 29 cases involving challenges to fingerprint testimony cited by Nordberg in this period, only one has resulted in exclusion of fingerprint testimony. Handwriting analysis testimony was admitted, having definitively been found ‘reliable’, in 14 out of 18 cases. Of the other four cases, no reliability analysis of handwriting expertise was considered in two of them. Handwriting testimony was excluded in the remaining two cases, although one of these exclusions was later reversed by the 7<sup>th</sup> circuit, on the basis that too much of the

District judge's analysis involved factors not appropriate to handwriting expertise under *Daubert*.<sup>12</sup>

Merlino's study (Merlino et al., 2007), found that twenty-five out of 37 proffers of forensic document examination (67.6%) were held to be admissible, and thirty four out of 39 proffers of fingerprint evidence (87.2%) were held to be admissible, from a total of 65 cases selected from the Lexis® database. The authors then proceeded to analyse the reasons given by judges for either the admission or the exclusion of the evidence. A significant difference was noted pre and post-*Kumho* in the number of mentions of reliability of the basis of the testimony, but not for the issues of relevance, probative value, qualifications, waste of time, application of the principle to the facts to the case, or general acceptance.

## The Exclusion of Forensic Identification Evidence

Information about the admission of evidence is useful for the forensic practitioner, however, the results are so varied, and the courts so inconsistent in the many ways they choose to admit it, that its practical value is often minimal. By analyzing the reasons that have been cited for *exclusion*, one can clearly envision what forms of evidence are failing to make it through the evidentiary filter, and why, as the courts slowly begin to awaken to the criticism around them. While complete exclusions of expert forensic evidence have been relatively rare, and most of these having been criticized as being the result of an overly-rigid application of *Daubert* (See *United States v Prime*, 2002), the judiciary are not completely united in their routine acceptance of forensic evidence.

Obtaining statistical data on admission and exclusion rates of forensic identification evidence post-*Daubert* is plagued by the familiar problems of selection bias and availability of unreported judgements, coupled with the sheer enormity of an inclusive search of the legal system for relevant cases. Studies that focus solely on appellate review decisions will inevitably suffer from selection bias, as not all district court cases end up being reviewed by a higher court. Studies that concentrate on reviews of published opinions, also suffer from selection bias, as Merritt cites figures indicating that close to 80% of all opinions remain unpublished (Merritt, 2001).<sup>13</sup> Furthermore, the policies governing which opinions were to be published differed substantially between circuits until the amendment to the Federal Rules of Appellate Procedure in 2007 (Mead, 2001), and several authors have commented on the

---

<sup>12</sup> See *Deputy v Lehman Bros., Inc.*, (2003)

<sup>13</sup> Certain legal and individual court databases have recently been publishing all Courts of Appeals decisions electronically, however there still appears to be wide variation in the extent to which this is carried out (Martin, 2007).

distorting effect this has on common law (Snowden, 2001). There is much data available characterising the ideology of reported versus unreported opinions, with most studies demonstrating that published opinions are not truly representative of all opinions given by the courts (Keele, 2007), this being particularly so for district courts (Olson, 1991). There are currently no studies available comparing non-published versus published *Daubert* opinions, hence it is impossible to define the extent to which non-published data may alter current perceptions of the application of the *Daubert* trilogy in courts.

## Methodology

In order to address some of these shortfalls, this author undertook to establish the relative frequency of exclusion versus admission of forensic identification evidence types following an admissibility challenge using the Daubert Tracker database. This represents an alternative to using a traditional legal databases such as Westlaw<sup>®</sup> and Lexis<sup>®</sup>, or by selectively sampling judicial cases. This database is accessible via the World Wide Web<sup>14</sup> on a fee-for-service basis, and is primarily designed to assist lawyers in research for a witness with specific expertise, in addition to providing historical details of any admissibility challenges, and outcomes, that the witness has faced. This database claims to be comprised of all reported and numerous unreported cases from both state and federal jurisdictions where an admissibility standard has been cited or mentioned in a decision (currently tracking cases related to over 165 authorities since 1993), and where a testifying expert's methodology or qualification has been challenged. It features Federal and State opinions, both reported and unreported. Now under licensing agreement with LexisNexis, the *Daubert* Tracker database provides an online repository of court decisions and documents relating to expert testimony and scientific evidence, in order to assist attorneys with case preparation (Daubert Tracker website, 04 Feb 2009). As yet, no legal scholars appear to have used this database for any published studies, and it is largely advertised as a resource for litigators to 'check their chances' on certain case types, as well as scrutinise any participating expert's credentials and court record.

The results of such a study therefore do not reflect 'the admissibility of forensic evidence types under *Daubert*', but represent the wider consideration of admissibility under a general 'reliability' requirement, applicable across many jurisdictions. There were several reasons for this approach. Firstly, a number of authors have reached the conclusion that challenges under the two major case-law examples, *Frye* and *Daubert*, essentially result in the same outcome (Bonifant, 2010; See Cheng, 2005). Secondly, it is recognised that most states in

---

<sup>14</sup> URL at [www.Dauberttracker.com](http://www.Dauberttracker.com)

the US have enacted similar evidence codes to the Federal Rules, which since the amendment in 2001 embody the ‘reliability’ ideals intended by the *Daubert* opinion (Bernstein and Jackson, 2003; Conley, 2005; See Cheng, 2005). Thirdly, the vast majority of judicial decisions recognise the fact that the admission of expert evidence is dependent on both a *relevance* and a *reliability* requirement, and it is now well-recognised that the question of whether an expert has tendered reliable evidence is not necessarily dependent on whether they have addressed issues specific to a *Daubert*-oriented inquiry. Fourthly, a more global analysis is likely to be of more use and interest to the forensic science practitioner who may be asked to give evidence in other States. Finally, the research encompasses only the period of *Daubert* and its progeny, thus accounting for its impact on other jurisdictions while not necessarily being restricted to those jurisdictions that have explicitly adopted the *Daubert* precedent.

A search of the Daubert Tracker database was conducted for forensic evidence disciplines from 1993 to 2008 via the drop-down list provided on the search page in four main areas; fingerprint analysis, firearms/toolmarks/ballistics (FATM) analysis, handwriting analysis,<sup>15</sup> and forensic odontology.<sup>16</sup> Where a case appeared more than once, usually because more than one expert was called to testify in the same forensic area, these were consolidated into one record to avoid double counting of decisions or cases.

## Results

A total of 551 cases from years 1993 to 2008 (inclusive) involving a challenge to expert witness testimony were retrieved. Any case involving a challenge to bullet lead analysis, recently withdrawn by the FBI as an incriminating forensic technique following several scathing reports regarding its inherent unreliability, was excluded leaving a total of 548 cases involving a challenge to forensic expert witness testimony in the four disciplines outlined above. In seven cases,<sup>17</sup> witnesses testifying as to the unreliability of the discipline were themselves excluded by the admissibility ruling. These were also subsequently excluded from the data analysis that follows. The remaining 541 challenges were analysed and categorised according to the final admissibility decision as either admitted, excluded, or limited (table 3.1 and 3.2). From a total of 541 challenges to forensic evidence in these disciplines, 67 challenges (12.4%) resulted in an outright exclusion, 14 (2.6%) resulted in

---

<sup>15</sup> The category of ‘handwriting analysis’ consisted of a combined search for both ‘Handwriting Analysis’ and ‘Forensic Document Examination’. Duplicate cases were excluded.

<sup>16</sup> The category considered as ‘forensic odontology’ consisted of a combined search for both ‘Dentistry – Forensic Odontology’ and ‘Bite-mark Expert’. Duplicate cases were excluded.

<sup>17</sup> Five of these were in fingerprint cases, and two were in handwriting cases.

admission with limitations on the scope of the testimony or were excluded only in part, and 467 challenges (85.0%) resulted in the evidence being admitted without restriction.

DISCIPLINE	N	% N	ADMITTED	% ADM	EXCLUDED or LIMITED	% EXC or LIM
FATM	207	38.3	170	82.1	37	17.9
FPA	176	32.5	164	93.2	12	6.8
FO	36	6.7	30	83.3	6	16.7
FDE/HW	122	22.6	96	78.7	26	21.3
TOTAL	541	100	460	85.0	81	15.0

Table 3.1 – Admissibility Challenges to Forensic Identification Science Evidence<sup>18</sup>

DISCIPLINE	N	EXCLUDED	% EXC	LIMITED	% LIM
FATM	207	34	16.4	3	1.4
FPA	176	9	5.1	3	1.7
FO	36	4	11.1	2	5.6
FDE/HW	122	20	16.4	6	4.9
TOTAL	541	67	12.2	14	2.6

Table 3.2 – Challenges either Excluded or Limited

In a further analysis using the grounded theory approach, the judicial opinions of all of the cases that were deemed to have been either excluded outright or limited ( $n_c=81$ ) were obtained and analysed in order to ascertain trends in judicial reasoning for the subsequent exclusion of such evidence. In order to extract as much detail as possible, open coding proceeded using a two-letter code based on the reasons given by the trial or appellate judge for the admissibility ruling (Table 3.3), resulting in 107 identifiable reasons given in the written judicial opinion for the exclusion or limitation of forensic evidence.

<sup>18</sup> FATM = Firearms and Toolmarks, FPA = Fingerprint Analysis, FO = Forensic Odontology, FDE/HW = Forensic Document Examination / Handwriting Analysis.

Code	Explanation	N Reasons
BP	Burden of proof not met by party (insufficient evidence tendered)	12
CK	Subject of testimony deemed a matter of common knowledge	2
CW	Testimony barred under confusion/waste of time discretion	3
EX	Expert's custom experiments not relevant or sufficiently reliable	6
IC	Inappropriate statements or conclusions made by expert	14
ME	Standard methods / techniques not demonstrably followed	6
NA	Testimony would not assist the trier of fact	2
NQ	Witness deemed not qualified to provide expert testimony	17
NR	Proposed testimony deemed not relevant to case at bar	8
OA	Witness attempted to testify outside area of expertise	6
PR	Procedural grounds for exclusion, such as late filing, etc	4
SP	Expert's testimony based on speculation	5
UK	Unable to establish reasons for exclusion from available record	2
ST	Underlying theory or technique not proved reliable	20

Table 3.3 – Open Coding

Two challenges resulted in no reasons being given for the exclusion of evidence at all, representing 2.5% of challenges considered, and these were excluded from any subsequent analysis of the data. This resulted in a total of 105 known reasons given in 79 independent challenges to forensic evidence in the identification sciences disciplines. A second, independent practitioner reviewed a random selection of these judicial opinions, and assigned codes from the selection generated above. This exercise demonstrated good concordance between two independent raters using the ‘open’ codes above, suggesting a high validity of the coding protocol.

The open coding, while providing some detailed information regarding the nature of judicial reasoning to exclude evidence, proved to be too granular for the level of analysis that was sought in latter stages of the study, and thus it was further refined via axial coding so that each case or challenge could be assigned to a broader description (table 3.4). This protocol resulted in the generation of six categories, summarised as; where the evidence was excluded due to a breach of procedural or technical rules, such as statutory time limits or rules of disclosure (P); when the witness was deemed not qualified to give expert opinion evidence (Q); when the evidence was excluded either under Rule 403 discretions, or as not relevant, or because the method/theory does not fit the circumstances of the case (R); when inappropriate statements were made, or inappropriate inferences were drawn by the expert witness, or witness attempted to testify outside area of expertise (W); when the expert failed to follow approved, recognised or accepted methods and techniques in their discipline (M) and; when

the underlying premise for the witnesses conclusions were not proven, due to failure to tender enough information about the theory or technique to allow analysis, a failure to meet or address *Daubert* or reliability thresholds, or through the conduct of experiments or use of 'experience' by the witness that was deemed to be scientifically unreliable (S).

While 'scientific reliability' is designated a category unto itself, the definition of reliability in an evidentiary sense means more than what was meant by the classification in the context of this coding – that of underlying scientific premise. Methodological shortfalls should also be considered in any discussion of reliability, as encompassed by Rule 702 of the Federal Rules of Evidence in that '...the testimony [must be] the product of reliable principles *and methods*'.<sup>19</sup> (Federal Rules of Evidence, 2008). The US Supreme Court in *Daubert* also noted that '...[T]he requirement that an expert's testimony pertain to "scientific knowledge" establishes a standard of evidentiary reliability', and that '...the word "knowledge" connotes more than subjective belief or unsupported speculation' (*Daubert v Merrell Dow Pharmaceuticals, Inc.*, 1993).<sup>20</sup> Reliability can therefore also be said to include consideration of whether the expert attempts to opine on an area that is outside the bounds of their particular field, the opinion is based on speculation, or the expert makes statements that remain unsupported by the bulk of their testimony. This notion is supported by other authors, agreeing that 'conjectures fail the reliability requirement for expert witness testimony' (Foster and Huber, 1999).

Hence, 'reliability', comes to mean far more than simply the underlying scientific premise on which the testimony rests. It includes consideration of such issues as witness conduct, including conjecture, speculation, and improper application of knowledge to areas outside of their designated expertise; methodological flaws in either their analysis or subsequent interpretation of results, and; the underlying scientific premise on which the expert's assumptions are based. This reasoning accounts for the final criteria for analysis, whether the evidence was excluded due to reliability concerns.

## Qualitative Analysis

In the data set analysed by this author, reliability issues represented 65.7% ( $n_R=69$ ) of all reasons given for exclusion of evidence given by forensic identification science witnesses. These reasons represented 45 admissibility challenges where a reliability concern was the

---

<sup>19</sup> Emphasis added.

<sup>20</sup> Original footnote omitted.



sole reason given for exclusion, and a further 6 challenges where both a reliability and another reason was cited.

Judicial Reasoning	Axial Code (Open Codes)	Explanation	No. Cases Citing*	% Cases Citing†
Procedural	<b>P</b> (PR)	Excluded due to a breach of procedural or technical rules, such as statutory time limits or rules of disclosure.	4	5.1%
Qualifications	<b>Q</b> (NR)	The witness was deemed not qualified to give expert opinion evidence.	16	20.3%
Relevancy	<b>R</b> (CK,CW, NR,NA)	Excluded as not relevant, including exclusions under Rule 403 (or similar) discretions.	15	19.0%
Witness Testimony	<b>W</b> (IC,SP,OA)	The expert made inappropriate statements, drew inappropriate inferences, or attempted to testify outside the area of his or her expertise.	17	23.8%
Methodology	<b>M</b> (ME)	The expert failed to follow approved, recognised or accepted methods and techniques in their analysis.	6	5.7%
Scientific Underpinning	<b>S</b> (EX, ST,BP)	The underlying premise for the witnesses conclusions were not proven, due to failure to tender enough information about the theory or technique to allow analysis, a failure to meet or address <i>Daubert</i> or other reliability thresholds, or through the conduct of experiments or use of 'experience' by the witness that was deemed to be scientifically unreliable.	27	36.2%
*Total exceeds N cases (79) as some cases were excluded due to reasons in more than one category.				
†Total exceeds 100% as some cases were excluded due to reasons in more than one category				

**Table 3.4 – Reasons for Exclusion of Evidence (Axial Coding)**

In total, reliability was cited as a reason for exclusion in 64.6% ( $n_c=51$ ) of admissibility challenges. The discipline representing the largest proportion of exclusions due to reliability was that of forensic odontology, with 100% ( $n_c=6$ ) of challenges being excluded on reliability grounds alone. The next highest was handwriting analysis, where 72% ( $n_c=18$ ) of exclusions cited reasons concerning the reliability of the testimony.

### ***Fingerprint Analysis***

164 (93.2%) challenges to fingerprint evidence resulted in unrestricted admission. A total of 12 challenges to fingerprint evidence resulted in an exclusion or limitation of the evidence. 5 (41.7%) of these challenges resulted in an exclusion due to grounds other than reliability; 2

of these challenges resulted in exclusion due to procedural errors concerning rules of disclosure; 2 were due to the witness not being adequately qualified to provide expert opinion; and one was due to rebuttal fingerprint testimony offered by the defendant being found to be irrelevant to the issue in question. 7 challenges (58.3%) resulted in an exclusion citing grounds of reliability. 2 of these resulted in only a partial exclusion, where the witness was allowed to testify, but only regarding the similarities and differences between known and unknown exemplars, and not that a particular latent print could be uniquely identified as the defendant. One of these cases was later reversed by the same judge after reconsideration (*United States v Llera Plaza I*, 2002), while the other involved footprint evidence rather than fingerprint evidence (*Hurrelbrink v State*, 2001). Of the remaining four cases, one was excluded on the grounds that the State failed entirely to rebut the defendant's admissibility challenge, and provided no evidence with which the judge could base a ruling otherwise.

#### ***Firearms, Toolmarks and Ballistics Evidence***

A total of 37 challenges (17.9%) to firearms, toolmarks and ballistics testimony resulted in either exclusion or limitation of the proffered evidence. The reasons for exclusion were unable to be ascertained in one of these challenges, and thus it was excluded from further analysis. Reliability concerns were mentioned in the reasons given for exclusion or limitation in 20 (52.8%) of challenges. While this appears to be lower than fingerprint evidence, in excluding cases that do not deal specifically with identification (where the expert was testifying regarding gun design, bullet trajectories, distance of firing), reliability played a role in 11 (64.7%) out of the remaining 17 successful challenges.

#### ***Forensic Document Examination***

Forensic document examination (including handwriting analysis) testimony was admitted without limitation in 96 (78.7%) challenges. A total of 26 challenges to handwriting analysis resulted in an exclusion or limitation on the proffered testimony. The reasoning behind one exclusion in this set was unable to be determined, and thus it was excluded from the analysis. Reliability reasons accounted for the exclusion in 18 challenges (72%). 15 of these challenges resulted in the judge making a negative finding specifically on the scientific validity of the technique used by the expert witness; yet 6 of those cases subsequently involved only a partial limitation of the testimony by preventing the witness making any statements as to whether the handwriting was specifically that of the defendant.

#### ***Forensic Odontology***

Forensic odontology evidence was admitted without restriction in 30 challenges (83.3%). Six challenges to the reliability of forensic odontology resulted in exclusion or limitation of the

evidence due to reliability reasons, accounting for 100% of the exclusions in this discipline. 5 of these challenges were in bite mark cases. The sixth regarded testimony given by a forensic odontologist who was also qualified in crime-scene evaluation and testifying thereto. It was decided on an appeal for ineffective counsel that the defendant's proffered (but uncalled) evidence in this last case regarding the positioning of a body would have been excluded under *Daubert*, even if it had been called by counsel (*Powers v State*, 2006). There was one instance of a challenge to the identification of a deceased individual from the dentition, however it was not successfully upheld.

### Reasons for Failing to Meet the Reliability Threshold

Through the axial and selective coding processes, it became apparent that there were clear trends in judicial reasoning for exclusion of forensic identification evidence. A narrative discussion of this various reasoning now follows.

**Sufficient Evidence.** The first of these reasons can be characterised as a failure to adduce sufficient evidence of reliability. Perhaps the most startling example of an exclusion happening due to a lack of evidence of reliability occurred in *Government of the Virgin Islands v Austin Jacobs* (2001), where the defendant moved to exclude the government's incriminating fingerprint identification evidence against him in a burglary case, on grounds that it did not meet the requirements of *Daubert* and Rule 702 of the Federal Rules of Evidence. Although the government had notice that the hearing would be to determine the admissibility of the fingerprint evidence under *Daubert*, it did not present any evidence or even produce the proffered expert witness for examination by the defendant or the court for the *Daubert* hearing. The territorial court subsequently excluded the testimony from appearing at trial, because it did not survive the threshold inquiry as required under *Daubert* and Rule 702. Following dismissal of the case against Jacobs, on appeal by the government, both the District Court<sup>21</sup> and the United States Court of Appeals for the Third Circuit upheld the territorial court's decision. The District court noted that 'the government failed to provide the trial judge with even the most basic information regarding the methods and scientific bases the expert used to arrive at her identification opinion. To say that the fingerprints were "compared" does not say anything about how they were compared'. The Third Circuit also noted that 'the government bore the burden of convincing the District Court that the methodology of its expert was reliable and could be appropriately applied to the facts at issue... Although the government could have satisfied its burden by providing

---

<sup>21</sup> At that time, the trial court in the Virgin Islands was known as the 'Territorial Court' (now known as the 'Superior Court'), and appeals were heard by the 'Appellate Division of the District Court'. Appeals from the District Court were subsequently heard by the U.S. Court of Appeal for the 3<sup>rd</sup> Circuit.

additional information, it did not do so. Therefore we conclude that the Territorial Court did not abuse its discretion by excluding the proposed expert.’ (*Government of the Virgin Islands v Austin Jacobs*, 2002).

**Statistics.** Another reason often cited in exclusions due to reliability was a failure to provide statistical information obtained via credible, scientific sources. In *People v Ballard* (2003), the appellate court found that the pronouncement of the expert that she was ‘99% certain’ that the defendant’s fingerprint was found in a stolen car ‘had no demonstrated basis in an established scientific discipline and rested solely upon Ms. Dyke’s personal opinion’. Failure of counsel to object to this statement constituted ineffective assistance of counsel, and the appellate court ordered a new trial.

In *United States v Rutherford* (2000) the defendant challenged the conclusion of the witness that he was the author of a signature or other writing on a document. The court wrote; ‘While the evidence adduced at the *Daubert/Kumho* hearing established that [the witness] meets the minimum requirements under Rule 702 to qualify as a non-scientific witness, there was no evidence adduced to support the nine-level scale of probabilities adopted by the American Board of Forensic Document Examiners (ABFDE) for conclusions as to handwriting identification. Accordingly, the Court shall preclude [the witness] from testifying to the degree of probability, confidence, or certainty underlying his proffered opinions’ (*United States v Rutherford*, 2000).

In other handwriting cases, the court has not denied the witness a chance to demonstrate his evidence in court, but limited the testimony to ‘...identifying and explaining the similarities and dissimilarities between the known exemplars and the questioned documents.’ Additionally, some courts have considered restricting the testimony of forensic document examiners as to their degree of certainty in determining the genuineness of a signature, in particular any reference to the nine-level ABFDE scale, such as *United States v Van Wyk* (2000), *United States v Hines* (1999), and *United States v Santillan* (1999).

In *Ege v Yukins* (2005), the evidentiary opinion describing the bitemark evidence came from a forensic odontologist who characterized the “match” of a mark on the victim’s cheek with the petitioner’s dentition in terms of overwhelming mathematical probability. The expert testified that in the Detroit Metropolitan Area, consisting of approximately three and a half million people, nobody else would match up. The court found that this evidence, particularly the testimony concerning the mathematical probability of an alternate random match, “lacked the proper foundation.” The judge noted that ‘forensic expert testimony regarding identification of the defendant based upon a statistical analysis requires a proper foundation.’

In *People v Wright* (1996), the same expert who testified in *Ege v Yukins* again came under fire. The odontologist in this case concluded that there was no one in the Detroit Metropolitan area of four million persons that would "even be close" to the unique pattern of defendant's bite mark, and that "there is probably no one in the world that would have this unique dentition." He based these conclusions on his professional training and statistics concerning bite marks from a 1984 article in the *Journal of Forensic Science*, which states if you have five unique points, that the chance of another individual making that same mark is 4.1 billion to one. The expert also stated that based on this article, if you had eight points of concordance, that no other person in the world could be considered a match. This evidence was admitted by the trial court, and the appellate court agreed that it was properly admitted, as such details went to weight and not admissibility. In a further appeal, the Supreme Court of Michigan disagreed with this finding, and directed the Court of Appeals to conduct a *Davis-Frye* hearing on both the methodology of bite mark comparison and the expert's use of statistical probabilities (*People v Wright*, 1998).

The Court of Appeals reconsidered its earlier decision, however, found that the bite mark testimony was still admissible, and that the statistical basis for the expert's conclusions was grounded in solid scientific method. Even though the Court of Appeals found the testimony reliable, the Supreme Court was clearly not satisfied with this result, as it ordered a new trial following the Court of Appeals' submission, instructing the trial court to undertake its own *Davis-Frye* hearing.

In *State v Fortin* (2007), the judge took the unusual step of requiring the production of a reliable database as an essential qualifier for the expert testimony of the forensic odontologist testifying as to the relatedness of two separate crimes<sup>22</sup>. This was deemed necessary in order to support his claim, and that of the pathologist, that 'from the thousands of cases they have reviewed in the course of their professional experience, they had never before seen such a combination of bite marks to the chin or breast on a sexual assault victim'.

**Methodology.** A failure to provide evidence that the method used for analysis in the particular case had been subject to any form of reliability testing also proved fatal to a number of challenges. Several forensic disciplines have attempted studies in order to prove the reliability of their technique, however, in several instances the fact that this has not directly related to the circumstances of the case has proved troublesome. Two fingerprint challenges have resulted in the finding that the ACE-V methodology used by latent

---

<sup>22</sup> So-called 'signature evidence'.

fingerprint examiners had not been subject to any reasonable test of scientific validity concerning *latent print* matching, compared with matching full, inked and rolled prints – *State of Maryland v Bryan Rose* (2007) and *Commonwealth v Patterson* (2005). In *Rose*, as the death penalty was sought, Souder J noted that ‘the court must be even more careful to determine whether the opinions which a party seeks to present at trial will be permitted.’ (*State v Rose*, 2007, Memorandum Decision) The defendant contested that the ACE-V methodology used by fingerprint examiners was not one that had been subject to scientific testing, thus the error rate was unknown, and the reliability of the method was therefore unproven<sup>23</sup>.

Souder J noted that although the ACE-V methodology appeared to be suited to testing, such tests had not yet been performed. She also noted that the principles underlying ACE-V, the uniqueness and permanence of fingerprints, still did not act as a substitute for testing of ACE-V itself, noting that no studies on the likelihood of partial prints taken from a crime scene would match only one set of fingerprints in the world existed. The judge also rejected the State’s argument that fingerprints enjoyed general acceptance outside of law enforcement, in areas such as DVI, security, biometrics, and immigration. She noted that the comparison of known exemplars was not the issue in this case, and that acceptance by these organisations did not demonstrate general acceptance of *latent* fingerprinting methodologies. In her conclusion, Souder noted that the ‘defendant demonstrated that there [were] no studies of ACE-V method to determine the reliability of the methodology.’ (*State of Maryland v Bryan Rose*, 2008).

In another example, a firearms examiner in *Sexton v State* (2002) declared that unfired bullets bore marks that matched those on fired cartridge cases, left by the magazine of a weapon involved in an aggravated assault<sup>24</sup>. He testified that if two cartridge cases showed the same magazine mark, then one could say with one hundred percent certainty that the two cartridges had been cycled through the same magazine. Under the Texas *Kelly* principles, consideration of acceptance by the scientific community, the existence of literature supporting the theory or technique, and a known error rate, weighed against admission of the

---

<sup>23</sup> In Maryland, the standard for admissibility of expert scientific evidence is set by two seminal cases, *United States v Frye* (1923) and *Reed v State* (1978). Under the *Frye-Reed* test, ‘...a party must establish that a technical or scientific method is reliable and generally accepted in the scientific community before the court will admit testimony based upon the application of the questioned technique.’ Furthermore, reliability must be demonstrated under the ‘gatekeeping’ obligation assigned to judges, which ‘applies not only to “scientific” testimony, but to all expert testimony’ (*State v Rose*, 2007, Memorandum Decision).

<sup>24</sup> Notably, no comparison with the magazine itself was actually carried out by the expert.

evidence in this case. The majority opinion noted that while the expert's '...record qualified him as a firearms identification expert, it [did] not support his capacity to identify cartridges on the basis of magazine marks only.' The court noted that 'the underlying theory of toolmark examination could be reliable in a given case, but under the *Kelly* criteria, the State failed to show that the technique applied in this case was valid... We granted review to determine whether the expert's testimony met the reliability requirement in *Kelly v State*. We conclude that it did not.' The judgement of the Court of Appeals was reversed.

Several judicial criticisms have been related specifically to the application of general analytical techniques to the particular circumstances of the case. In *United States v Fujii* (2000), the court found 'no need to weigh in on th[e] question...[of] whether handwriting analysis *per se* meets the *Daubert* standards, as its *application in this case poses more significant problems*'.<sup>25</sup> Among these problems were that this case involved handprinting, and the court was unconvinced that such a specialist field even existed, as handwriting comparison usually involved cursive writing. As the defendant was Japanese, it was offered that it would be difficult for an individual not familiar with the writing of a Japanese person, not writing in their native language, who was taught to write English in Japan, and who has strong cultural expectations to write in a uniform manner, to identify the subtle dissimilarities in the handwriting of individual writers; on this matter the government had no counter. Consequently, it was the specific facts of the case that led to the wholesale exclusion of the experts evidence.

Similarly, the district court in *United States v Saelee* (2001) noted that '[t]he Government has failed to prove that [the expert's] proposed testimony as to similarity between known and questioned documents would be the product of reliable principles and methods',<sup>26</sup> and subsequently ruled to exclude all testimony of hand-printing comparison evidence (c.f. handwriting comparison, which is considered to involve analysis of cursive letters) at trial. However, the trial judge also pointed out that '[this court] is not holding that handwriting analysis can never be a field of expertise under the Federal Rules of Evidence. It is merely holding that the Government has failed to meet its burden of establishing that the proffered testimony in this case is admissible...'

**Standards.** Experts who fail to adhere to recognised methods and protocols within the field have also found themselves excluded from giving evidence. In *Maryland v Rose* (2007) the judge included in her reasons for excluding the evidence that 'the defendant ... demonstrated

---

<sup>25</sup> Emphasis added.

<sup>26</sup> Citation omitted.

that the ACE-V method was neither perfect or perfectly applied to partial latent print identifications...'. In *Ramirez v State* (2001), the judge also noted that there was no written authority upholding the expert's method of analysis, and this was cited as a reason counting towards the exclusion of the toolmark expert's evidence. In *Ege v Yukins* (2005), the judge noted that the fact that the odontologist had used a non-dental forensic photo of the possible bite mark, in addition to using models that were made in excess of nine years after the original wound was inflicted to exclude other suspects, also went against admission of the evidence.

In *Eubanks v Hale* (1999), the document examiner admitted that he had not followed standard procedures in examining the challenged signatures by not using a microscope, however this was overturned on appeal, with the appellate court citing Rule 44(j) of the Alabama Rules of Civil Procedure, (Proof of Handwriting) '...Whenever the genuineness of the handwriting of any person may be involved, any admitted or proved handwriting of such person shall be competent evidence as a basis for comparison to prove or disprove such genuineness'. However, in another case, *Bourne v Town of Madison* (2007), the court found that the expert's methodology of enlargement was "inconsistent with the accepted methodology among forensic document examiners," which is to view the original document using a microscope. They concluded that the expert had not made a sufficient showing that his methodology for determining the authenticity of the signature was reliable, and the evidence was thus excluded in its entirety.

In *Keko v Hingle* (1999), forensic odontologist Dr Michael West testified that the suspect in a murder trial was the perpetrator after examining the body using his own controversial 'blue light' technique. He claimed that he had discovered tooth marks on the body, and after obtaining a warrant for dental impressions, proceeded to declare a 'match' between the defendant's dentition and the tooth marks that only he was able to see. In November of 1996, Judge Michael Kirby of the 25th Judicial District found that, under the newly enunciated *Daubert* standard, Dr. West's testimony should not have been presented to the jury in plaintiff's original murder trial as it was not found to be sufficiently reliable or accepted by practitioners in the field, and, granted the defendant's motion for a new trial.

**Explanation of Methodology.** Several experts have also been criticised for not being able to adequately explain their methodology. In *State v Swinton* (2004) the defendant challenged the admissibility of the forensic odontology evidence in two ways. First, photographs of a bite mark on the victim's body that were enhanced using a computer software program known as 'Lucis', and second, images of the defendant's teeth overlaid, or superimposed,



upon photographs of the bite mark that were made through the use of Adobe Photoshop. The first challenge did not survive, as it was determined that the 'Lucis' program was incapable of actually altering the dimensions of the image such that it might be susceptible to manipulation. The second challenge, however, was successful, as the expert could not articulate sufficiently how the visual effect of the defendant's translucent teeth superimposed over the bite mark was produced. Because he was not familiar with Adobe Photoshop, and was using the program for the first time for an odontological match, the expert secured the assistance of a Fairfield University chemistry professor, to scan these images and create the overlays by using the computer program to superimpose the defendant's dentition over the bite mark. This professor was not made available to testify at trial. The defendant argued that the reliability of what had come out of the computer was the issue, and that he could not test that reliability by questioning a person who merely had sat next to the machine. The defendant referred to the issue at hand as a 'black-box test' whereby the jury was being asked to trust the computer. Specifically, he could not testify as to whether; the computer processes that were used to create the overlays were accepted in the field of odontology as standard and competent; whether proper procedures were followed in connection with the input and output of information; whether Adobe Photoshop was reliable for this sort of forensic application, or; whether the equipment was programmed and operated correctly (*State v Swinton*, 2004). When asked how the computer actually superimposed the tracing of the biting edges of the defendant's teeth over the photograph, the odontologist was only able to manage '[the professor]... moved them together.' Importantly, the court noted that the Adobe Photoshop program was capable of actually altering photographs, and that 'a witness must be able to testify, adequately and truthfully, as to exactly what the jury is looking at.' The appellate court ruled that the Adobe overlays were improperly admitted, as the defendant should have had the opportunity to question someone who could testify accurately as to the reliability of the evidence and the processes used to generate them.

**Proof of Methodology.** Experts who fail to tender sufficient documentation of the identification method they have used in the case at bar have also found their evidence excluded. In *State of New Hampshire v Richard Langill* (2007),<sup>27</sup> Coffey J found that while ACE-V was a reliable method of analysing latent fingerprints, a failure to document this analysis, in combination with the possibility of a biased confirmation due to the lack of a 'blind' verification by the second examiner, resulted 'in an insufficient basis for the court to find that the principles were reliably applied to the facts of the case'. This led the trial court

---

<sup>27</sup> *Daubert* was explicitly adopted as precedent in New Hampshire in 2002 – See *Baker Valley Lumber Co. v Ingersoll-Rand Co.* (2002)

to exclude the fingerprint identification ‘until such time that this court is satisfied that the [New Hampshire Department of Safety Forensic Laboratory] has conducted a methodically reliable analysis on the latent print found on [the victim’s] bureau.’ (*State v Langill*, 2007). Gertner J noted that the lack of good records of the analysis in *United States v Green* (2005) ‘pointed against admission of the testimony’, even though it was ultimately allowed with limitation.

In *United States v Monteiro* (2006) the expert had failed to make any sketches, or take photographs of his comparison work, and his notes were scant. He had also failed to document the verification of his results by another qualified examiner. Therefore, Saris J required the government to produce documentation that proved the technique utilised in the analysis conformed to standards accepted in the forensic firearms analysis field and that the findings had been verified by an independent second examiner. Saris J granted the defendants motion to exclude ballistics evidence; ‘The government must ensure that its proffered firearms identification testimony comports with the established standards in the field for peer review and documentation.’ However, she also stated that she would permit the expert to testify if the government could demonstrate that it met the peer-review and documentation requirements.

In *Ramirez v State* (2001), a forensic examiner attempted to identify a knife that was allegedly used as a murder weapon in 1983 to the exclusion of all others. This case was decided under Florida’s *Frye* rule, requiring that ‘the basic underlying principles of scientific evidence have been tested and accepted by the relevant scientific community by looking to properties that traditionally inhere in scientific acceptance for the type of methodology or procedure under review – i.e. “indicia” or “hallmarks” of admissibility’.<sup>28</sup> The knife mark evidence in *Ramirez* was initially deemed admissible in the trial *Frye* hearing, however the appellate court disagreed with this finding. Among the reasons given for the Supreme Court’s decision was the fact that the expert had not taken any photographs, prepared notes or a written report delineating the basis for identification, and the Supreme Court of Florida subsequently reversed the original verdict and vacated the defendant’s sentence.

**Objective Standards.** The fact that a forensic comparison itself is a subjective one is not necessarily reason for exclusion when there exists objective standards by which to relate the findings. Lack of such objective standards, however, has proven fatal to admissibility in several cases. Notably, in *Ramirez v State* (2001), the judge wrote in his reasons for exclusion of the proffered testimony that the expert’s method, involving comparisons of

---

<sup>28</sup> See *Brim v State* (1997)

striation marks on cartilage, was not governed by objective scientific standards, and that this strengthened the case for exclusion of his evidence.

**Observer Bias.** The existence of observer bias has been noted as strengthening the case for exclusion in several challenges where evidence has been limited, including *United States v Green* (2005) and *State v Langill* (2007).

**Proficiency testing.** In a similar vein, several experts were excluded because of a failure to ensure that their so-called ‘proficiency tests’ replicated real-life conditions. Numerous disciplines undertake proficiency testing in an attempt to verify the ability of their practitioners, often as a substitute for demonstrating the reliability of the scientific theory itself. Recently, this practice has come under closer scrutiny by the judiciary. In *Maryland v Rose* (2007) it was also noted that proficiency tests were undertaken and passed, however, these were criticised by the trial judge as not being representative of real-world conditions. Therefore, they did not carry as much weight towards admitting the fingerprint evidence as they might have, although the evidence was eventually admitted, but with limitations.

**Certainty.** Instances where the expert has attempted to testify to absolute certainty have also been fatal to the admissibility challenge in a number of instances. The judge in *Maryland v Rose* (2007) found the expert’s testimony that the error rate associated with the ACE-V method for assessing fingerprints was ‘zero’ was not credible or persuasive, and formed part of her reasoning for excluding it. In *United States v Monteiro* (2006) the judge permitted the firearms expert to testify, with the caveat that ‘no examiner should testify to an absolute or arbitrary degree of statistical certainty, as the examiners opinion as to an identification is a subjective one’. She noted that the ‘lack of a *universal* standard for declaring a match is troubling, but not fatal under *Daubert/Kumho*, because a court may admit well-founded testimony based on specialised training and experience.’ She found that ‘a qualified examiner who has documented and had a second qualified examiner verify her results may testify based on those results that a cartridge case matches a particular firearm *to a reasonable degree of ballistic certainty*.’ Testimony that the toolmark identification technique used in *Ramirez v State* (2001d) ‘was infallible’, and ‘that it was impossible to make a false positive identification’ using this particular technique also went against admission of the evidence in this case, the judge noting that there was no data on error rate produced to verify this claim.

In another descriptor of exclusionary trends, judges are beginning to find that witnesses attempting to make an ‘absolute’ identification of the defendant, or ‘to the exclusion of all others’ are not relying on credible scientific reasoning. Souder J in *Maryland v Rose* (2007)

allowed the State fingerprint examiner to testify only on the proviso that none of the testimony addressed the ultimate question of whether any latent fingerprint was that of the defendant. Along similar lines, Gertner J in *United States v Green* (2005) allowed a firearms and toolmark examiner to testify as to his observations, however, Gertner did not allow him ‘to conclude that the match he found by dint of the specific methodology he used permits “the exclusion of all other guns” as the source of the shell casings.’ Gertner J also noted concerns regarding the reliability and testing of the expert’s methods<sup>29</sup> and subsequently found that they did not permit such a conclusion to be drawn. Gertner ultimately allowed the testimony, with restriction on the expert’s conclusions, stating as she did in *Hines*<sup>30</sup> ‘...that if the jury is able to see and understand what the expert saw, then the testimony may be admissible. If the jurors cannot see and understand the testimony, it amounts to nothing more than [...] the kind of *ipse dixit* which the court was concerned in *Joiner*.’

In *United States v Diaz*, (2007) the trial judge found that ‘the theory of firearm identification used by the SFPD Crime Lab is reliable under *Daubert*. The record, however, does not support the conclusion that identifications can be made to the exclusion of all other firearms in the world. Thus, the examiners who testify in this case may only testify that a match has been made to a “reasonable degree of certainty in the ballistics field”’ (*United States v Diaz*, 2007). The evidence was allowed but with limitations on the conclusions that the expert witness could draw.

Similarly, in *Wolf v Ramsey* (2003), the Court concluded that ‘while the expert can properly assist the trier of fact by pointing out marked differences and unusual similarities between Mrs. Ramsey’s writing and the Ransom Note, he has not demonstrated a methodology whereby he can draw a conclusion, to an absolute certainty, that a given writer wrote the note.’

**Relationship of Expertise to Case at Bar.** In the only complete exclusion to rest on such reasoning, the trial judge in *People v Wynne* (1999) excluded the testimony of the firearms expert not because he failed to be qualified as an expert, but because he ‘failed to lay a foundation to show that this expertise qualified him to make the relevant assessment from

---

<sup>29</sup> ‘The question of whether the expert’s technique or theory is scientifically reliable is a specific one: The issue is not whether the field in general uses a reliable methodology, but the reliability of the expert’s methodology *in the case at bar*...or what the Court has described as a question of “fit.”’ (*United States v Green*, 2005 at 119, citations omitted).

<sup>30</sup> *United States v Hines*, (1999)

the coroner's report'<sup>31</sup>. This is perhaps an extreme example, where the expert attempted to testify having not actually performed an examination of the material evidence.

Similarly, firearms identification experts have been excluded on the grounds that their testimony related to distance of firing (*State v Lopez*, 1996), gun design (*Barrett v International Armaments, Inc.*, 1999) or whether a gun had been dropped or thrown (*United States v Fisher*, 2002) does not relate to their core area of expertise.

**Experiments.** Experts who present their own hypotheses to the court but have failed to test their theories in a controlled, scientific manner have also been prevented from testifying. Several expert witnesses are tempted to conduct their own research, or 'experiments', in an attempt to use these results to later prove their hypothesis to the court. Such practices should be adopted with caution. In *Smith v State* (1997), the trial court excluded the testimony on grounds that the defence had failed to establish a proper foundation demonstrating that the experiments conducted by the expert witness were done so in a scientific manner, using comparable circumstances. The appellate court affirmed this decision. Similarly, in *Estate of Kenneth Griffin v Hickson* (2002), the plaintiffs relied on what were termed the 'Holmesburg experiments.' Based on the results of these experiments, the plaintiffs averred that the weapons used by agents Hickson and Martinez were capable of depositing the gun shot residue particles on the defendant. However, the judge noted that 'certain critical aspects of the Holmesburg experiments must be introduced via expert testimony if the tests are to be deemed admissible'. The plaintiffs were not prepared to present such evidence and therefore the evidence was not permitted.

## Discussion

Data concerning the ratio of admitted versus excluded evidence is difficult to accept without caveat. No single source routinely lists every challenge brought before the courts, in addition to the usual problems of sourcing unpublished opinions, or potential selection bias. Daubert Tracker was used in this instance partly because of the numbers of cases offered, and the fact that they were relatively easily filtered.

One point that became obvious from analysis of the judicial opinions where forensic identification evidence was excluded was the general lack of reference to the four *Daubert* criteria. Other authors too have noted that while judges may be excluding more evidence, they are doing so via legal doctrines available to them well before *Daubert* (Saks and

---

<sup>31</sup> Emphasis added.

Faigman, 2005). As the Supreme Court decision concerned Federal legislation, it was not a binding precedent on State and Territory courts, thus Dobbin and colleagues attempted to analyse the difference in judicial procedures used to assess scientific evidence between those States that have either explicitly affirmed the *Daubert* approach, or use a similar ‘reliability’ test to interpret their own Rules of Evidence<sup>32</sup> (referred to as ‘*Daubert* States’), and those that chose to retain their own standards (most of whom followed a ‘general acceptance’ test) in an attempt to realise the practical effect of the *Daubert* approach. They found no statistical difference between *Daubert* and non-*Daubert* States in their use of procedures to analyse scientific evidence (Dobbin et al., 2007), according with Cheng and Yoon’s state-based study, which lends weight to the argument that judicial scrutiny under either *Daubert* or *Frye* currently leads to essentially the same conclusion<sup>33</sup> (Cheng and Yoon, 2005). The effect of *Daubert* was also noted to have been ‘minimal’ in the Delaware Superior Court, with little differences in outcomes of *in limine* motions regarding expert evidence in the pre and post-*Daubert* eras in that state<sup>34</sup> (Waters and Hodge, 2005). Rice argues that judges will not have the knowledge or skills to assess such criteria as ‘error rate’, or to determine the use of suitable versus unsuitable methodologies, and hence will come to rely on the general acceptance of scientists as much in the future as they have in the past (Rice, 2000).

Groscup confirmed this trend in her analysis of 693 state and federal appellate decisions in criminal cases from 1988 to 1999, with results demonstrating that the *Daubert* factors were the *least* likely criteria to be considered in admissibility decisions (Groscup et al., 2002). Real<sup>35</sup> concluded his year 2000 essay by stating that *Daubert* essentially boils down to; the need to query the relevancy of the expert’s testimony; the expert’s qualifications, and; whether the evidence given will ultimately aid the judge or jury in reaching a verdict (Real, 2000).<sup>36</sup> Groscup’s analysis confirmed that this attitude prevails amongst the general judiciary, finding that the most commonly cited reasons for exclusion of evidence were that

---

<sup>32</sup> Which, in the majority of cases, mirror the Federal Rules language anyway.

<sup>33</sup> The two studies agree, despite the fact that Cheng and Yoon’s methodology has been described as ‘novel and subject to serious criticism’ (Vickers, 2005). Their study rests on the premise that the practical result of *Daubert* is reflected in the number of cases where a defendant chooses to remove their case to a Federal court. This assumption ignores other factors besides that of standard of scientific admissibility that influence this decision (a more detailed description of such factors is given in *Removal: What Every Litigator Should Know* (Leonard, 1999)). The alternative conclusion to their findings, that the *Daubert* variable has little or no effect on removal rate, is afforded little discussion in their conclusions.

<sup>34</sup> Although noting that Delaware had been using reliability criteria for assessment of expert evidence from as early as 1989, despite not formally adopting the *Daubert* trilogy until 1999. This potentially explains, perhaps at least partially, this apparent lack of impact.

<sup>35</sup> A District Judge for California at the time of publication of his article.

<sup>36</sup> Real’s conclusion is strange given the discussion he embarked on in the main text of his article concerning the application of a reliability standard, yet he makes no mention of reliability in his summation.

it was either ‘not relevant’, the expert was ‘not appropriately qualified’, or that the evidence would ‘not assist the trier of fact’, and that judges rarely excluded evidence on the grounds of ‘unreliability’ as suggested by *Daubert*.

General acceptance also continues its dominating role in admissibility decisions. Despite earlier scholarly interpretations that general acceptance is no longer a necessary or sufficient condition for admissibility (Faigman, 1995), Groscup found that it was the only one of the *Daubert* criteria that resulted in a positive predictive value for admissibility in criminal trials<sup>37</sup> (Groscup, 2004). A survey of 3,260 judicial opinions from 1993 to 2004 also found that general acceptance was the most referred to *Daubert* factor (Merlino et al., 2008).

An assessment of the accuracy of the percentage admission/exclusion rate reveals that the figures concerning admission rate are likely to be an underestimate of the true proportion of cases where evidence is admitted. A brief comparison was done with the Westlaw<sup>®</sup> database, using a search string for filtering fingerprint challenges<sup>38</sup> from 1993 to 2008. Of the 300 cases retrieved, only 153 of them specifically related to a challenge to the admissibility of fingerprint evidence. The other 147 cases usually cited a fingerprint case in order to prove an aspect of law related to another area, or the words in the search string were merely used in passing while the case itself did not involve any specific challenge to fingerprint evidence. Of the 153 cases deemed to be relevant to this study, 74 of them were not listed in Daubert Tracker. Fingerprints were deemed to be admissible in all of these cases, and in the majority of examples, this was ruled on without the conduct of an evidentiary hearing. Conversely, Daubert Tracker listed 90 cases that were not picked up by WestLaw’s search engine. If one was to combine these results, then the ratio for admission versus exclusion of fingerprint evidence alone becomes 95.2% admission and 4.8% exclusion, compared to the figures obtained by using only the Daubert Tracker database of 93.2% admission and 6.8% exclusion. These figures are somewhat artificial, but demonstrate that in general, the proportion of evidence found to be admissible is likely to be underestimated in studies relying on a single consolidated source of information.

### **Pre versus Post–Kumho decisions**

The number of exclusions due to reliability before the *Kumho* decision ( $n_c=20$ ), or after the *Kumho* decision ( $n_c=61$ ) was also analysed. 10 challenges (50%) resulted in an exclusion due to reliability reasons prior to *Kumho*, with the remaining 10 due to other reasons. Post-

---

<sup>37</sup> The other positive predictors of admissibility were relevance, qualifications, assistance to the trier of fact, and prejudicial impact.

<sup>38</sup> The search string used was *fingerprint /2 admiss! da(aft 1993 & bef 2009)* supplemented by *fingerprint /7 reliab! da(aft 1993 & bef 2009)*.

*Kumho*, 41 challenges (67%) resulted in an exclusion due to reliability reasons, and 20 challenges resulted in an exclusion due to other reasons. This difference in proportions of exclusions due to reliability between the reasons for exclusion pre-*Kumho* versus post-*Kumho* versus difference was not found to be significant using a Chi-Squared test ( $p=0.166$ )<sup>39</sup>. This may suggest that the notion of pre-*Kumho* courts relying on finding the majority of forensic evidence admissible due to its ‘non-scientific’ nature may have been unfounded. Alternatively, it could be considered that post-*Kumho* courts are not necessarily finding forensic scientific evidence unreliable, as predicted by many critics. Whether this is due to the inherent reliability of forensic identification evidence, or the poor application of reliability standards by the judiciary is still a matter for debate. Critics such as those referenced in this thesis would assert the latter, whereas undoubtedly, judges would assert the former.

### **Criminal v Civil Testimony**

Saks and Faigman cite numerous examples of what they characterise as ‘judicial casualness’ in criminal cases (Saks and Faigman, 2005). In this study, criminal defendant experts testimony was excluded outright in 26 out of 27 (96.3%) ‘successful’ challenges (defined by the evidence being either excluded or limited). By contrast, prosecution experts underwent 36 challenges resulting in the exclusion or limitation of their evidence. Only 24 of these resulted in complete exclusion (66.7%). Civil defendant expert witnesses were completely excluded in all of the successful challenges against them, with no instances of limitation, whereas plaintiff experts were completely excluded in 11 out of 12 successful challenges. This accords with Risinger’s analysis of federal and state court opinions pre and post-*Daubert*, which indicated that *Daubert* seems to have had a far larger effect on civil cases than criminal, claiming a 34-fold increase in civil opinions citing *Daubert*, compared to only a 5-fold increase in criminal cases (Risinger, 2000). He also noted that in cases where *Daubert* challenges were introduced into criminal cases, more than 90% were ruled in favour of the prosecutors [i.e. the government].

Risinger also expressed concern that there seemed to be a developing trend towards a stricter standard of quality control in civil but not criminal cases, at odds with legal precedent claiming that the standard of proof for criminal trials is more demanding than a civil proceedings (*Anderson v Liberty Lobby, Inc.*, 1986; see *United States v Taylor* 1972).<sup>40</sup>

<sup>39</sup> Using Fisher’s Exact Test the two-sided p-value was 0.1901, also considered non-significant.

<sup>40</sup> In accordance with the common law, in civil cases the plaintiff is required to bear evidence only that it was ‘more probable than not’ the defendant did what was claimed. In criminal cases, the plaintiff is required to offer evidence establishing that the defendant did what was claimed ‘beyond reasonable doubt’, which is considered to be a higher standard.



Waters and Hodge also noted that far more civil cases entered into *Daubert* challenges when compared with criminal cases in Delaware (Waters and Hodge, 2005). Berger has commented on the same observation (Berger, 2005), and claimed that application of this strict standard for civil trials is inappropriate; a subtly different conclusion to that offered by Risinger, in that Berger's criticism seems to be directed towards an inappropriate over-application of *Daubert* in civil cases, rather than an under-application in criminal cases. Other authors have also criticised the application of *Daubert* as being unreasonable in its requirements of civil plaintiffs (Melnick, 2005), claiming that its misapplication in civil trials has resulted in evidence that most experts would agree would form a basis for causation being rejected (Bernstein, 2007).

DeCoux offered two overlapping theories as to the one-sided nature apparent in prosecutorial success in evidentiary hearings associated with criminal cases (DeCoux, 2007), citing both the trial courts' unbounded flexibility in addressing reliability evaluations and the misapplication of the law by the judiciary as problematic. Neufeld noted, among other issues, that criminal defendants are usually indigent, appointed counsel by a financially constrained system, and are restricted to using expensive private crime laboratories in order to substantiate their case,<sup>41</sup> hence many have simply not been able to afford to mount a proper defence (Neufeld, 2005). Saks and Faigman accuse the judiciary of undertaking a "vague call to arms against junk science in civil cases, while keeping their hands off the government's proffers in criminal cases" (Saks and Faigman, 2005), and reference a bill that was before Congress in 1995 that essentially tried to amend Rule 702 so that reliability need not be scrutinised for criminal cases at all, requiring civil litigants to prove their scientific claims but not criminal prosecutors.<sup>42</sup> Harris noted in her recent empirical research (Harris, 2008,) that DNA and polygraph evidence is significantly more likely to gain acceptance in courts when supported by law enforcement, fuelling criticism of the potential under-application of *Daubert* in criminal cases where the interests of the government and politics come into play.

---

<sup>41</sup> And then only with the court's permission.

<sup>42</sup> The proposed amendment (which ultimately never passed Congress) read, in part:

(b) *Adequate basis for opinion. – Testimony...that is based on... scientific knowledge shall be inadmissible in evidence unless the court determines that such opinion*

*1) is scientifically valid and reliable;*

*2) has a valid scientific connection to the fact it is offered to prove; and*

*3) is sufficiently reliable so that the probative value of such evidence outweighs the dangers specified in rule 403.*

(c) ...

(d) *Scope. – Subdivision (b) does not apply to criminal proceedings.*

(Attorney Accountability Act, 1995)

## Conclusion

The cases described above represent a variety of reasons where judges have made the decision to exclude or limit evidence, finding that it does not satisfy the reliability aspect of its relevant statutes. Many of these decisions are not trans-case, that is, they do not look to rule on issues concerning reliability of entire disciplines applicable across more than the case at bar. While Souder J concluded her memorandum decision in *Maryland v Rose* by commenting that the proof presented by the State represented the ‘type of procedure that *Frye* intended to banish, that is a subjective, untested, unverifiable identification procedure that purports to be infallible’, she also noted that the decision to limit fingerprint evidence was not based on a finding that the entire discipline lacked reliability, but that the State, by ‘call[ing] one witness...[who] admitted he had no college degree, [whose] testimony revealed [he had] no scientific training...was unable to do so’ (*State v Rose*, 2007, Memorandum Decision).

Herein also lies the importance of specific training and experience for the forensic identification expert, to be able to avoid any of the cardinal faults that some other experts have brought to court. Forensic experts need to be assured of the soundness of their field by undertaking their own critical, unbiased analysis of the evidence suggesting the reliability, or otherwise, of their discipline. The National Academy of Sciences report released this year has noted that ‘[j]udicial review, by itself, will not cure the infirmities of the forensic science community’ (National Research Council, 2009). The judicial system has always required the help of experts, yet as this chapter demonstrates, the trend of unquestioned admission by judges may be turning in the wake of recent criticisms. However, the vast majority of forensic identification science evidence is still admitted without restriction, despite closer scrutiny by judges. What is also obvious from this data is that the reasons for their exclusions do not necessarily follow those specifically articulated as suggestions for assessment in the *Daubert* opinion, and so it is suggested that to base future research and practices on *Daubert*-style interpretations of what connotes admissible expert evidence is deeply flawed.

Nonetheless, in recognition of the fact that what *Daubert* has imparted is a notable change to the thrust of Federal Rules with the establishment of a reliability requirement, it would be folly to dismiss it as irrelevant entirely. So far, its main effect has been seen in the United States, and yet the United Kingdom has recently referenced *Daubert* and its progeny as being a potential direction for expert evidence reform in that country (see the House of Commons Science and Technology Committee, 2005; and Woolf, 2009). A further key

aspect to consider is whether *Daubert* has had, or is likely to have, any effect on the admission of expert evidence or its relevant legislation in Australia, and it is the next chapter that explores this notion.

## CHAPTER 4

# Australian Evidence Law and the Admissibility of Scientific Expert Evidence

---

While it can be argued that *Daubert* has had relatively little practical effect on the admission of mainstream forensic identification evidence, it has undoubtedly generated intense discussion amongst the legal and forensic science fraternities. The *Daubert* decision, and the associated consequences, have been watched closely by judicial systems outside the US, particularly that of Australia and the United Kingdom. A key question remains as to how much influence *Daubert* will have on our own legal doctrine regarding the admission of expert testimony. There is good reason to suspect that introduction of such a mandate in Australia is likely to generate a similar amount of discourse from legal academics and forensic scientists in this country, were it to occur. This chapter explores the question of whether or not *Daubert*-like interpretations of the admission of expert testimony are prevalent in Australia, whether they are likely to continue, and where trends in admissibility of expert testimony in this country are headed.

## Australian Evidence Law

A rule common to all legislation in Australia known as the *opinion rule* expressly forbids witnesses to give their opinion regarding matters that are before a court.<sup>1</sup> Despite this rule, principles derived through the common law<sup>2</sup> provide exceptions to the opinion rule that grant a basis for expert witnesses to give opinion evidence. In those states and territories that are yet to adopt the Uniform Evidence Legislation (Queensland, Northern Territory, Western Australia and South Australia) it is under these common law principles that expert witnesses are permitted to testify their opinion in courts. These criteria (often referred to as ‘rules’, although they are not formally legislated as such) are known as ‘exclusionary’, because failure to meet them generally results in the evidence being excluded from the trial

---

<sup>1</sup> The rule is expressed specifically in the various *Evidence Acts* as: ‘Evidence of an opinion is not admissible to prove the existence of a fact about the existence of which the opinion was expressed’. (Section 76, *Evidence Act* (Cth) )

<sup>2</sup> That is, law derived through case law and precedent, rather than legislation

proceedings. Freckelton and Selby (Freckelton and Selby, 2009) have summarised these ‘common law rules of expert evidence’ as follows:

1. The *expertise rule* – experts must have demonstrable knowledge outside the knowledge of the ordinary trier of fact. For example, an expert is not merely so because they have an opinion on a subject, they must prove that they have had experience, training or a special skill in the relevant area in order for them to qualify as such.
2. The *area of expertise rule* – experts must testify in an area that is relevant to their area of expertise.
3. The *common knowledge rule* – the expert cannot testify if the subject matter they are giving evidence in falls within the scope of ‘common knowledge’ that the ordinary trier of fact could be expected to know.
4. The *basis rule* – the expert cannot give evidence based on hearsay, or the work of others, unless the basis of that work is also admitted in evidence.<sup>3</sup>
5. The *ultimate issue rule* – an expert witness cannot give evidence about a matter that is for the court to decide. For example, an expert cannot state their opinion regarding the defendant’s guilt or innocence, as this is the function of the trier of fact.

While the basis for admission of expert evidence in those States and Territories that operate under these common law principles is derived from case law and precedent, in Federal cases the *Evidence Act* 1995 (Commonwealth) specifically legislates an exception to the opinion rule, and thereby grants authority for an expert witness to tender their opinion in court through Section 79 of the Act, which states:

If a person has specialised knowledge based on training, study or experience, the opinion rule does not apply to evidence of an opinion of that person that is solely based on that knowledge.

This is the primary pathway by which experts have authority to express their opinion in all federal courts in Australia and the Australian Capital Territory. Four other Australian States or Territories have mirrored this legislation for use in their respective state courts,<sup>4</sup> and thus as a collective, this body of laws is referred to as the Uniform Evidence Legislation (UEL). An important aspect of the Uniform Evidence Legislation to be noted is that it represents a move away from several of the cardinal common law tenets regarding expert testimony.

Firstly, the UEL does not explicitly incorporate an area of expertise rule. The Australian legislation does not use the terminology ‘expert’, thus any evidence that could be perceived

<sup>3</sup> The basis for this exception was because of the legislatively enacted *hearsay rule*, which forbids evidence not based on first-hand knowledge or experience of the facts in issue

<sup>4</sup> As of 2009, the following states have adopted the language of the Commonwealth Evidence Act in their own legislation: New South Wales (1995), Tasmania (2001), Norfolk Island (2004) and Victoria (2008).

as not being ‘ordinary knowledge’ is potentially admissible.<sup>5</sup> Expert witnesses do not necessarily have to claim a specific area of expertise, they merely need to demonstrate that they have knowledge or experience of some matter ‘outside the experience and knowledge of the judge and jury’ (*Murphy v The Queen*, 1989).

Secondly, the ‘ultimate issue rule’ and the ‘common knowledge rule’ are also rendered obsolete under the UEL. Section 80 states:

Evidence of an opinion is not inadmissible only because it is about

- a) a fact in issue or an ultimate issue
- b) a matter of common knowledge

However, despite the wording of s80(a), most Australian courts still generally find that experts drawing blatant conclusions about the ultimate issue<sup>6</sup> usurp the authority of the court, and thus are not permitted (Freckelton and Selby, 2009). Clarification of this apparent inconsistency was made by Austin J, who stated: ‘If an expert chooses to address an ultimate issue which is a mixed use of fact and law, and does not have expertise in law and assessment of evidence, there is a probability that opinion will be regarded as outside the scope of the experts specialised knowledge [and thus excluded]’ (*Dean-Willcocks v Commonwealth Bank of Australia*, 2003).

The basis rule also appears to be somewhat liberally interpreted in Australian courts, and this was the case in most jurisdictions even before the adoption of the UEL. In a strict application of this rule, an expert could not rely on the data or opinions of others when giving his testimony<sup>7</sup>, however, Section 60 of the UEL now allows for a flexible interpretation:

The hearsay rule does not apply to evidence of a previous representation that is admitted because it is relevant for a purpose other than proof of the fact intended to asserted by the representation.

Section 144(b) also provides:

Proof is not required about knowledge that is not reasonably open to question and is:

- a) ...
- b) capable of verification by reference to a document the authority of which cannot be reasonably questioned.

<sup>5</sup> See later on in this chapter for a discussion on the implication of this precedent in criminal trials.

<sup>6</sup> The ‘ultimate issue’ in a criminal trial, for example, would be the defendant’s guilt or innocence

<sup>7</sup> This was often difficult to achieve in practice. Strictly speaking, under the basis rule a medical expert witnesses could not, in court, refer to case notes and histories that he or she had not taken directly from the patient.

Thus there remains a need to disclose the *source* of such evidence as a basis for the opinion, however the source now need not be separately proven under the rules of evidence (Freckelton and Selby, 2009). For example, if the witness refers to well-established scientific principles, such as the histological diagnosis of cancer, the myriad of textbooks and research publications need not be investigated or admitted into evidence as these are sources that cannot be reasonably questioned.

Aside from situations concerning irrelevant testimony or lack of ‘expertise’, Australian courts can generally refuse to admit (or, subsequently exclude) expert scientific evidence only under what is known as their ‘general discretionary authority’, provided for by ss135 and 136 of the UEL<sup>8</sup>. Section 135 provides:

The court may refuse to admit evidence on the basis that its probative value is outweighed by the danger that the evidence might:

- a) be unfairly prejudicial to a party
- b) be misleading or confusing
- c) cause or result in undue waste of time

This discretion is usually used in the context of excluding expert testimony when the expert has attempted to offer an opinion that oversteps the boundaries of the subject on which he or she is technically permitted to give an opinion on.<sup>9</sup> One author concluded that as a result of the discretionary authority bestowed upon judges, explicit expression of the common law basis rule is unnecessary in Australian evidence legislation; where there is absence of identification of those facts which the evidence is said to rely upon, its probative value is likely to be low, and thus “...may be so low as to have no weight at all, and thus may not be admitted on discretionary grounds” (Nell, 2006).

The Australian Law Reform Commission (ALRC) supported this view, claiming that the reluctance to explicitly include the basis rule in Australian evidence legislation is due to the fact its principles are already embodied in two other orthodox concepts. The first of these is the notion that the lower the relationship between the facts proved and assumed, the less weight the opinion will be afforded, and the second, affirming Nell’s conclusion above, is that where evidence has so little correlation to the facts of the case, it automatically has so

---

<sup>8</sup> Section 135 uses the words ‘may *exclude* evidence’, whereas s136 uses the words ‘may *limit* evidence’. Note that this essentially means that the judge can still choose to admit evidence that may form part of wholly inadmissible testimony. It also does not *require*, it merely suggests that it *may* be excluded or limited, thus essentially giving the judge an option.

<sup>9</sup> In other words, because the expert has failed to base their opinion wholly or substantially on their specialised knowledge, it is therefore likely to be misleading, or cause a waste of time, and is thus excluded. For some relatively recent examples, see *Tante v Roth* (), *NRMA Limited v Morgan & Ors* (), *Makita (Australia) v Sprowles* (2001), *R. v Tang* (2006)

little probative value that it becomes inadmissible (Australian Law Reform Commission, 2005b).

Despite these iterations that s135 is the tool by which unreliable evidence can be dismissed, exclusion of scientific evidence in Australian courts using this line of reasoning is often difficult, and it has been rarely applied in cases where the reliability of evidence is a concern. Hunt J admitted: ‘...I [could] not [...] imagine a case where relevant scientific evidence would be excluded in the exercise of discretion because of a tendency to mislead, prejudice or confuse a jury’ (*R v Karger*, 2001).

While the treatment of expert evidence has never been vastly different between courts in different States and Territories in Australia, it is generally expected that the Uniform Evidence legislation, like the Federal Rules of Evidence in the USA, will gradually be adopted in one form or another by the remaining States, leading to legislative commonality in approaches to all forms of evidence in Australian courts (and not just that of experts). Former High Court Justice Michael Kirby noted:

*The more states that join the Uniform Evidence Act, the more it becomes a self-fulfilling prophecy<sup>10</sup> – the more likely it is that the remaining States will adopt the Uniform Evidence Act. This is because whatever their reservations may be, the remaining states will prefer to be part of the national scheme. The training of lawyers in those states has to include the Evidence Act because of federal jurisdiction, which applies in all jurisdictions in Australia, including the non-Uniform Evidence Act jurisdictions. I would think that within a decade, most, if not all of the other jurisdictions will have fallen into line.* (Kirby, 2010)

The continued adoption of the Uniform Evidence Act by the remaining States suggests that, if anything, the admission of expert testimony will shift towards a more liberal approach, rather than a more prescriptive one, in keeping with the UEL’s current basis.

## The Admission of Scientific Evidence in Australia

Case law supporting the interpretation of evidence legislation as it relates to scientific evidence is far more varied than that in the United States, with Australia so far lacking a seminal precedent such as *Daubert* to guide interpretation of the UEL in this regard. One of the earliest Australian cases to discuss admissibility of scientific evidence is *R v Parker* (1912), where the judge ruled that evidence of such a nature was admissible when it was

---

<sup>10</sup> Australian Law Reform Commission *Evidence (Interim)*, ALRC 26, 1985; *Ibid*, *Evidence* (ALRC 38, 1987); *Ibid*, *Uniform Evidence Law* (ALRC 102, 2005)



found to be ‘generally recognised by scientific men’. *Frye* subsequently used similar language in the well-known American case in 1923, and while Australia has not formally adopted the *Frye* test, it has been cited in a number of Australian cases since then, appearing to have been ‘inadvertently’ introduced into Australian law (Bernstein, 1996). Officially, the adoption of *Frye* has been rejected by several Australian legal sources,<sup>11</sup> with reasons cited similar to those that caused its demise in America, however, a similar language has remained dominant in Australian judicial opinions regarding the admissibility of scientific evidence (Freckelton and Selby, 2009), suggesting at least some continued deference to the US precedent.

Scientific evidence has not been considered a special category of expert evidence in Australia since *Hocking v Bell* (1945). Unlike the US Federal Rules, the word ‘scientific’ is not used in the Australian legislation;<sup>12</sup> the UEL simply uses the term ‘specialised knowledge’. A framework for approaching admissibility of evidence under the UEL is contained in the Introductory Note to Chapter 3 of the *Evidence Act*, and insofar as it relates to expert evidence, this framework suggests that assessment of its admissibility is based on a three stage process – firstly by assessing whether the evidence in question is *relevant* to the case at bar; secondly, consideration of whether the evidence falls under the *specialised knowledge* exception to the opinion rule; and thirdly, by consideration of whether any aspect of the evidence would be *subject to exclusion* under the general discretionary authority of the trial judge (Australian Law Reform Commission, 2005b).

It is the phrase ‘specialised knowledge’ that has generated controversy in interpretation of the *Evidence Act*,<sup>13</sup> and whether this criteria is met appears to be the fundamental question considered when assessing the validity of expert testimony in Australia. A seminal case involving assessment of whether expert evidence constituted ‘specialised knowledge’ was *R v Bonython* (1984), where it was noted that the expert must give an opinion that is based on a ‘reliable body of knowledge’ in order for it to meet this principle. However, as Freckelton points out:

*The Bonython approach seems to look to the field of knowledge, and whether it is regarded as reliable by those working in the area; it is analogous to the Frye test for sure, and yet the*

---

<sup>11</sup> Einstein J, for one, has stated that the adoption of the UEL constitutes a direct rejection of the *Frye* test, in that it does not call for an assessment of reliability or general acceptance (Einstein, 2000). See also *Idoport Pty Ltd v National Australia Bank Limited* (1999); the Australian Law Reform Commission Report (Australian Law Reform Commission, 2005b), and the ALRC Discussion Paper 23 - Evidence (Australian Law Reform Commission, 1985) both of which state that *Frye* has not been, and should not be adopted in Australia.

<sup>12</sup> Compare US Federal Rule 702.

<sup>13</sup> And indeed, the case law prior to its introduction.

*decision of Johnson and plenty of other cases is that just because someone has an unusual approach to things does not rule them out. I think that these are potentially compatible [approaches] in that as long as someone isn't standing entirely alone with bizarre views, then the question is one of what weight is to be given to the evidence.* (Freckelton, 2009)

In *HG v The Queen* (1999), Gaudron J stated that despite the introduction of the UEL, '[...]here is no reason to think that the expression "specialised knowledge" gives rise to a test which is any respect narrower or more restrictive than the position at common law'. Elaborating further, the opinion clarifies that establishing whether the expert has 'specialised knowledge' requires consideration of whether that knowledge is sufficiently organised, or sufficiently recognised, to be accepted as a reliable body of knowledge or experience, thus reaffirming the direction taken in *R v Bonython* (Einstein, 2000), yet using an unmistakeable *Frye*-style language.

Yet while *HG* and *Bonython* both use the terms 'reliable' and 'general acceptance', their semantic is not equivalent to either the *Frye* or *Daubert* cases. 'General acceptance' in the Australian context has often meant general acceptance *by the court*, rather than any particular scientific or academic body.<sup>14</sup> In *R v Lewis* (1987), Muirhead AJ wrote that even though there appeared to be 'no established universal view as to the reliability of forensic odontological evidence', this did not, in itself, render the evidence inadmissible. The Victorian Supreme Court has also held that providing *the judge* is satisfied that a field of expertise exists, it is no barrier to the admission of expert evidence that an expert within that field does not hold the views of the majority (see *R v Johnson*, 1994).

Concerning 'reliability', Australian courts tend to focus on establishing the existence of a 'reliable *body of knowledge*', rather than exploring any construct that involves assessment of a particular theory or technique akin to the *Daubert* enquiry. Use of the term 'reliable' in the Australian context has often echoed the US concept of 'general acceptance' more than anything else, despite blatant denials of 'general acceptance' as a cornerstone of expert evidence admissibility. As an example, consider *Casley-Smith v F.S. Evans and Sons Pty. Ltd.* (No.1) (1988). In this case, the opinion directs courts to 'directly assess reliability and not focus on the general acceptance of the theory or technique...'. Alas, it then continued with, '...by establishing whether members of that intellectual community would consider the area reliable', thereby ultimately still only referring to a consideration of 'general acceptance' within a relevant field.

---

<sup>14</sup> This began with the *Bonython* opinion, if not earlier. The phrase '....sufficiently organized or recognized to be accepted as a reliable body of knowledge or experience' does not specify which community that it is to be recognized or accepted by.

Australian judicial opinions that have addressed evidence legislation since *HG v The Queen* have specifically avoided direct discourse concerning interpretations of the term *reliable*. This is despite widespread academic recognition of its antithesis – *unreliability* – as a significant issue regarding scientific evidence in Australia since the early 1990's (Bourke, 1993a). The general trend in Australian courts has been to admit evidence, even when its basis is dubious, citing the precedent that reliability is a matter of weight, rather than admissibility (Freckelton and Selby, 2009). The leading case in this regard is *Commissioner for Government Transport v Adamcik* (1961). In this somewhat controversial case, a single expert witness testified for the plaintiff that the plaintiff's leukaemia could have been caused by his physical injury.<sup>15</sup> The jury subsequently found for the plaintiff, despite substantial evidence to the contrary being presented by the defence, and in the subsequent failed appeal to the High Court, Windeyer J held that providing the jury relied on admissible evidence in order to reach a conclusion, then that conclusion shall stand, even if the judge may subsequently disagree with the verdict. In *Adamcik*, the expert's evidence was allowed on the basis that he was a qualified medical practitioner, and thus was entitled to give an opinion on medical matters as per the expertise rule under common law. Assessment of the reliability of his methodology was for the jury to decide, and they could assign it the weight they felt appropriate.

But as Michael Kirby has said;

*That's a different matter to what is now normally the case in civil trials. Now such evidence would have to come before a trial judge. And the trial judge would have to give reasons. The trial judge's reasons could be scrutinised by an appellate court. The appellate court is not limited, in most jurisdictions of Australia, to reviewing the law. It is also generally obliged to review the factual determinations at trial. So that case might not be repeated today in the circumstances of civil trials conducted before judges. In jury trials, there was a very strong principle that if there was any evidence before the jury that was properly before them, it was up to the jury to decide which evidence they accepted or not. Appellate courts would very rarely interfere. They would only generally interfere if, as it was put, the result was against the evidence, the weight of the evidence, or the evidence overwhelmingly preponderated to the opposite conclusion. That was very rarely done in the days of jury trial (Kirby, 2010).*

Yet despite the relative age of this decision, and noting the comments made by the former High Court justice, matters such as reliability still tend to go to weight, rather than overall

---

<sup>15</sup> The plaintiff's expert also admitted that his opinion was based on his own 'theory' of a link between emotional distress and illness, and that he had actually only treated one other case of leukaemia in his practising career.

admissibility of expert evidence. In *HG v The Queen* (1999), Gleeson CJ ruled that although the expert witness ‘failed to [explain]... the precise factual matters upon which his conclusions with respect to the complainant were based’, the evidence was not deemed inadmissible due to this failing.<sup>16</sup> Gaudron J also commented in the same case that ‘...these are matters that might persuade a judge, in the case of trial by judge alone, or a jury not to accept [the expert witness’s] evidence [however,] they are not matters bearing on its admissibility as opinion evidence’.<sup>17</sup>

Justice Heydon recently argued that experts still have a duty to expose their reasoning processes to the court, so that a determination of the reliability of their methodology can be made. In *Makita (Australia) v Sprowles* (2001), he stated that ‘...the prime duty of experts [in giving evidence] is to furnish the trier of fact with criteria enabling evaluation of the validity of the experts conclusions...the trier of fact must arrive at an independent assessment of the opinions and their value...’. He elaborated further, stating that ‘the opinion of an expert requires demonstration or examination of the scientific or other intellectual basis of the conclusions reached... If all these matters are not made explicit, it is not possible to be sure whether the opinion is based wholly or substantially on the expert’s specialised knowledge. If one cannot be sure of that, the evidence is not admissible’ (Heydon, 2000). Justice Heydon maintains that s79 of the UEL *implies* a reliability requirement,<sup>18</sup> and advocates exclusion of evidence on the grounds of a lack of reliable methodology on this basis.

Heydon J has been one of few judicial advocates for a ‘reliability’ assessment of expert testimony. In *Makita*, Heydon proposed a set of seven criteria for establishing the reliability of expert testimony, namely:

1. It must be agreed or demonstrated that there is a field of ‘specialised knowledge’;

---

<sup>16</sup> ‘...[It is] not thereby require[d], as a condition of admissibility, that the assumed facts on which the opinion is based are established by the evidence. If at the end of the evidence they are not established, the weight to be accorded the opinion will be reduced, perhaps to nil. But that is not a matter of admissibility’. (*HG v The Queen*, 1999). It is acknowledged that this quote is technically referring to a failure to establish a factual basis for the opinion (i.e. the common law basis rule), which may not strictly speaking be the same as establishing reliability, but the two concepts are inextricably linked.

<sup>17</sup> Compare the American approach as articulated in *Daubert*, where Blackmun J stated ‘...the Rules of Evidence – especially Rule 702 – do assign to the trial judge the task of ensuring that an expert’s testimony...rests on a reliable foundation...’ (1993 at 37).

<sup>18</sup> A similar conclusion was reached in *Ocean Marine Mutual Insurance Association (Europe) OV v Jetopay Pty Ltd* (2000), where the Court stated more simply that it required ‘exposure of the reasoning process...[so as to demonstrate]...that the opinion is based on ... specialised knowledge’.

2. There must be an identified aspect of that field in which the witness demonstrates that by reason of specified training, study or experience, the witness has become an expert
3. The opinion proffered must be ‘wholly or substantially based on the witness’ expert knowledge’;
4. So far as the opinion is based on facts ‘observed’ by the expert, they must be identified and admissibly proved by the expert;
5. So far as the opinion is based on ‘assumed’ or ‘accepted’ facts, they must be identified and proved in some other way;
6. It must be established that the facts on which the opinion is based form a proper foundation for it; and
7. The opinion of an expert requires demonstration or examination of the scientific or other intellectual basis of the conclusions reached: that is, the expert’s evidence must explain how the field of ‘specialised knowledge’ in which the witness is expert by reason of ‘training, study or experience’, and on which the opinion is ‘wholly or substantially based’, applies to the facts assumed or observed so as to produce the opinion propounded.

While the first three points re-iterate well established Australian principles, the remaining four bear a remarkable similarity to the validity thrust embodied in the *Daubert* trilogy, integrating what is essentially an assessment of the reliability of expert’s underlying methodology into the admissibility enquiry.

Heydon’s approach, however, has been controversial, as evidence scholar Ian Freckelton has commented;

*Well that’s controversial too, I mean whether he was meaning to do that, or whether he was simply talking in terms of what’s optimal practice. There was a controversy in a number of cases after Makita about what status should be accorded to what he was saying, and most of the leading decisions now suggest that he wasn’t trying to lay down rigid criteria for admissibility. (Freckelton, 2009)*

Later cases certainly appear to have paid minimal regard to Heydon’s recommendation. The Federal Court chose to ignore the *Makita* ruling in *Cadbury Schweppes Pty. Ltd. v Darrell Lea Chocolate Shops Pty., Ltd.* (2006). Heerey J stated; ‘...I accept the submission of senior counsel that this aspect of *Makita* has not been followed in the Federal Court...[Consequently,] [t]he lack of proof of a substantial part of the factual basis of [the expert’s] opinion does not of itself, render [it] inadmissible under s79. Such lack of proof merely goes to the weight which may be given to the opinion.’ The Chief Justice of the NSW

Supreme Court in *R v Tang* (2006) also recently commented that ‘...the focus of attention must be on the meaning of the statutory phrase “specialised knowledge”, rather than by invoking extraneous ideas such as “reliability”....’.<sup>19</sup>

The Australian Law Reform Commission (ALRC) comment that while it is possible to construe Heydon’s comments in *Makita* as ‘rules for admissibility’, they prefer to read them as a statement ‘underscoring the relationship between relevance and weight’ (Australian Law Reform Commission, 2005b). Ian Freckelton again commented:

*The approach of the ALRC reports was to get away from rigid criteria for admissibility across the board in the Evidence Act, to do with hearsay and also to do with opinion, allowing more evidence, but to do it in a flexible way, with most of the work being done by the discretions, rather than formulating rules. Hence we deliberately did not include an area of expertise rule, we deliberately didn’t include a basis rule, and we recommended against, or we abolished, the common knowledge and ultimate issue rules – so to get away from that kind of rigidity that would follow from a formulaic application of Makita.’ (Freckelton, 2009)*

The *Makita* ‘reliability’ criteria have thus had little impact on the admission of evidence in Australian courts so far, with judges continuing to hold that reliability continues to go to weight, rather than admissibility of an expert opinion (Freckelton and Selby, 2009).<sup>20</sup>

## Commentary on Expert Scientific Evidence in Australia

Empirical studies on the admissibility of scientific expert evidence in Australia are relatively few in number. An early study attempted analysis of legal and scientific opinion of the extent of a perceived problem with unreliable scientific evidence via analysis of five case studies and informal interviews conducted with an ad hoc selection of public and private solicitors, defence and Crown barristers, Supreme and County Court Victorian judges, and forensic scientists from both independent laboratories and police-associated services (Bourke, 1993a). The lack of a formalised interview rubric makes detailed comparison with other data difficult, as this study relies on what are essentially anecdotal claims. However, the author

<sup>19</sup> The evidence given by the forensic anatomist in *Tang* was ultimately excluded on the basis that a substantial portion of her opinion was not substantially or wholly based on a body of specialised knowledge, as required by s79. Spigelman CJ said, of the body mapping and posture comparison aspect of her evidence: ‘That evidence barely, if at all, rose above a subjective belief and it did not, in my opinion, manifest anything of a “specialised” character. It was not, in my opinion, shown to be “specialised knowledge” within the meaning of s79.’

<sup>20</sup> See *Sydneywide Distributors Pty. Ltd. v Red Bull Australia Pty. Ltd.* (2002), *Cadbury Schweppes Pty. Ltd. v Darrell Lea Chocolate Shops Pty. Ltd.* (2002), *Jango v Northern Territory* (2004) .

stressed that while no conclusions could be drawn regarding the percentage of criminal trials that potentially contain unreliable evidence, there was indisputable qualitative evidence to suggest that the problem of unreliable scientific evidence was wide-ranging and generally not recognised by the legal profession. Bourke subsequently recommended several solutions encompassing both the legal and scientific communities, including a reformation of the Rules of Evidence to exclude scientific testimony that is deemed unreliable (Bourke, 1993b).

The Australasian Institute of Judicial Administration Report *Australian Judicial Perspectives on Expert Evidence: An Empirical Study*<sup>21</sup> (Freckelton et al., 1999a) published the results of a more formalised survey of Australian judges some five years after Bourke's analysis. The authors claim that the most significant concern amongst judges who have trial experience has been that of bias or partiality among expert witnesses, particularly a lack of independence in forensically commissioned medical practitioners and accountants. In the wake of this result, and in conjunction with a similarly reported observation in an enquiry into the English judicial system<sup>22</sup>, the authors proposed an 'Expert Witness Code of Conduct', which has now been adopted in slightly modified form by the Federal courts.<sup>23</sup> The requirement for expert witnesses to make an additional oath or 'statement' highlighting their over-riding duty to the court has been met with considerable controversy.

Edmond (2003) has claimed that experts have always testified under oath, and that it does not follow that clarifying further their duty to the court will produce more objective evidence. He commented that there exists very little empirical evidence to suggest that there is actually a problem with bias in Australian courts, ultimately arguing that the AIJA Report is representative of judges' *opinions*, not facts. Edmond is highly critical of the overall AIJA survey design and subsequent analysis, citing inappropriate questions, inadequate explanation of questions or terminology so as to deny meaningful comparison of results, a general reification of judicial perspectives, and the inappropriate extrapolation of judge's views to jurors (Edmond, 2005). Wheate's more recent survey of forensic scientist's and lay juror's attitudes demonstrated that their views on bias differ substantially to that of the judiciary as proposed by the authors of AIJA survey report (Wheate, 2007).

Tellingly, Street (2005) notes that despite the judiciaries concerns, the existence of bias in an expert witness is yet to prove fatal to the admissibility of scientific evidence in Australian

<sup>21</sup> Hereafter referred to as the AIJA Judges Survey

<sup>22</sup> Known in judicial circles as *The Woolf Report* (See Woolf, 1996)

<sup>23</sup> See *Schedule K – Expert Witness Code of Conduct* of the NSW Supreme Court Rules as an example.

courts, with bias so far also having only been considered a matter of weight.<sup>24</sup> While the AIJA Judges Survey concluded that the majority of judges have experienced bias, relatively little discussion is afforded on what percentage of judges believe that bias is significant, or whether that bias can be, or is, sufficiently dealt with at the time of the trial. A historical ‘lack of use’ of ss135, 136 or 137 to exclude or limit evidence that is perceived as partisan may indicate that the problem is not as significant as the authors suggest. Bell claimed that these recent amendments to Australian procedure have essentially ignored the larger issue of reliability, and as such ‘evaded the *raison d’être* of expert testimony’ (Bell, 2006). He noted that ‘the current emphasis on swearing allegiance to the court puts the essence of evidence secondary to legal routines, which are irrelevant to the logic of science’.

Expert evidence procedure has also taken a turn following the Woolf Report (Woolf, 1996) and its subsequent induced reform in English expert evidence law. Australia has legitimised a technique in civil trials that has become known as ‘hot-tubbing’, or ‘concurrent expert evidence’. Rather than subjecting numerous, competing expert witnesses to judicial examination and cross examination, leaving the trier of fact to assign weight accordingly, ‘hot-tubbing’ is essentially the practice of a pre-trial conference, directed by the trial judge, at which all of the expert witnesses gather and attempt to provide a consolidated statement on matters that they agree, or disagree if no compromise can be reached. Proponents of this concept argue that this in turn leads to better expert evidence as there is less distortion of expert opinion by inept leading from counsel, a better critical discussion of concepts among scientific peers, an enhanced capacity of a judge to ascertain which opinions to accept as he or she can view the interaction between the experts, and is quicker than traditional examination and cross-examination processes (McClellan, 2007).

Edmond again criticised these reforms in the Australian context, believing that the contention that experts will moderate their own views, or abandon ‘junk science’ in the presence of their peers is a romantic view of expert controversy that is unlikely to play out in the real world (Edmond, 2003). He has cited Saks’s article (Saks, 1995) and other literature from as early as the 1960’s, demonstrating that similar approaches have had very mixed results. Edmond’s own personal research<sup>25</sup> demonstrates that lawyers generally dislike the concept, and even though experts were generally favourable, several doubted their ability to reduce the amount of disagreement either in or out of court. While Edmond cited ‘hot-tubbing’ as a potentially useful tool, he believes it has limited potential amid the lack of evidence suggesting it has actually increased access to justice (Edmond, 2009).

<sup>24</sup> As ruled in *FGT Custodians v Fagenblat* (2003) VSCA 33.

<sup>25</sup> Conducted via telephone conversations with various lawyers and experts.



The former High Court justice, Michael Kirby, also expressed reservations about this approach:

*Some doctors are extremely articulate. Some are cautious, less experienced, nervous, lacking in self-confidence, but possibly right. The 'hot-tub' process puts a great premium on the capacity of the medical witness, or other witness, to become a kind of advocate, and to speak with energy and persuasion, sometimes perhaps to brow-beat somebody who may not have quite the same skill in language, or even perhaps the same first language as the other witness. So I have reservations about these procedures.* (Kirby, 2010)

At this time, hot-tubbing is limited to civil cases in federal courts (McInnis, 1995), with its subsequent failure to be adopted in other Australian courts potentially due its cumbersome process, its expense, and the perception that the procedure is too adversarial (Davies, 2004). It appears unlikely that this approach would be used in criminal trials, as Edmond explains:

*The other thing for the adversarial trial is how concurrent evidence would conform, where would it go in the trial, because the accused is not obliged to do anything until its heard the case of the State, and so they would have to put their own expert in as part of the prosecutions case, and there are tactical advantages in where it goes, and so forth. These things become much more messy. [Concurrent evidence] is much more free-form, and in a criminal trial, testimony is quite strictly managed, because if an expert says something out of line, and the same with other witnesses, it can abort trials.* (Edmond, 2009)

Judicial attitude towards reliability was also explored in the AIJA Judges survey. The majority of judges responded that reliability should not be a precondition to admissibility of expert evidence (Freckelton et al., 1999b). Very few thought that 'falsifiability' should be a key determinate of reliability, however, the implication of this is questionable, given the poor understanding of these terms by US judges, and noting that the authors of this survey made no attempt to define (or elicit a definition from the participants) of these phrases.

While several US authors justify the *Daubert* decision as necessarily 'shielding' juries from evidence that they may not comprehend, avoiding the risk that they will use it inappropriately in their determinations, Freckelton notes that emerging judicial rhetoric in Australia supports the notion that juries have no requirement for such shielding, with evidence suggesting that Australian jurors *do* have the capacity to deal with scientific evidence, providing appropriate judicial guidance is given (Freckelton, 1997). Wheate noted that due to the significant differences in juror selection processes in Australian and the US, application of American juror studies to Australian cases are limited (Wheate, 2006). Further

evidence by Wheate supports the notion that Australian juries appear to understand evidence enough to reach a satisfactory verdict (Wheate, 2007). The judiciary support this attitude, with a High Court ruling relatively recently in *Velevski v The Queen* (2002) that juries *are* able to competently assess conflicting evidence.

A companion report to the AIJA Judges Survey was released by Freckelton and colleagues in 2000, which analysed the results of a similar survey, this time for Australian magistrates. The AIJA Magistrates Survey highlighted that the factors bearing most heavily on the acceptance of expert evidence were a clear explanation, demonstration of appropriate practical experience, and impartiality when presenting evidence (Freckelton et al., 2000). It also concluded that Australian courts are *not* concerned with experts usurping the judicial function, juries (or judges) being incompetent to resolve disputing expert claims, or with witnesses not having proper experience or expertise (See Kirby, 2003).

The Australian Law Reform Commission (ALRC) is a permanent, independent statutory corporation that conducts inquiries into areas of law reform as requested by the Attorney-General of Australia. Its recommendations do not automatically become law, however it has a powerful influence on the direction of Australian law reform, with the vast majority of its recommendations being partially or wholly implemented since its inception in 1975. The most recent ALRC report on evidence (Australian Law Reform Commission, 2005b) continues to focus on the potential for the judiciary to adopt practices that help them ensure expert evidence is presented in a way that assists them in assessing whether the evidence complies with s79 of the Acts – specifically, whether the evidence is related to the specialised knowledge of the expert, and whether the expert is appropriately non-partisan to the benefit of the court, rather than the parties. The ALRC report emphasised that the majority of concerns raised by the Australian legal community in relation to expert opinion evidence are procedural in nature, including those related to costs, delay and bias, and reveals that there is little support for the notion that ‘reliability’ concerns akin to those expressed in the US are an issue that should be addressed through adoption of more restrictive evidence rules in this country.

## **The *Daubert* Trilogy in Australian Law**

Several authors have called for the establishment of criteria for the assessment of expert scientific evidence in Australia, akin to the *Daubert* –induced reforms to Rule 702 of the US Federal Rules of Evidence. Since the publication of the *Daubert* case in the US, Wilson predicted that judges in Australia would become increasingly reluctant to admit evidence of

a novel scientific technique unless they could be convinced of its reliability, comprehensibility and relevance (Wilson, 1994). Odgers and Richardson (1995) also claimed that there seemed to be a trend towards evaluating ‘reliability’ in Australian courts, which the *Daubert* decision would only serve to reinforce. Kinley and Rose also saw the potential for *Daubert* to make its way into Australian courts, albeit by implementing a qualitative shift in the extent of the evidentiary enquiry<sup>26</sup> rather than changing its ideological focus (Kinley and Rose, 1999). Burbidge predicted in 2001 that Australia would adopt a more rigorous approach to the admissibility of expert evidence in the wake of *Daubert* and *Kumho* (Burbidge, 2001). Freckelton and Selby have also been advocates of exercising the *Daubert* approach in Australia, claiming that it provides a ‘sophisticated means of distinguishing between evidence that is not yet capable of being effectively evaluated by courts from that which has been tested within the medium of peer-review and debate’. They too have predicted that the *Daubert* decision will exercise a profound ongoing influence on the development to Australian expert evidence law (Freckelton and Selby, 2009).

So far, very little of this prophetic commentary appears to have played out. Primarily, there are fundamental differences in the overall approach to trials and litigation between the United States and Australia that render adoption of any of their legislation (or case law) as precedent unlikely. Remembering that *Daubert* was introduced in order to reform the presentation of expert evidence to juries, it is important to note that the civil jury trial has been largely abolished in Australia, and is mainly used as a format in defamation cases where the expression of community values is considered significant (McClellan, 2007). Even in criminal matters, most often a judge sits alone as the trier of fact in this country.<sup>27</sup> While section 80 of the Australian Constitution appears to confer a right to jury trial,<sup>28</sup> the use of the words ‘on indictment’ in the original text have subsequently led to interpretations whereby only those who have committed an indictable offence are so entitled. To further complicate the issue, the prosecution also may play a role in deciding the format of the trial in some jurisdictions (Fricke, 1997). By comparison, a US citizen’s right to jury trial is fundamentally guaranteed by Article Three of the US Constitution.<sup>29</sup> This right is extended

---

<sup>26</sup> They viewed *Daubert* as a call for judges to ‘simply ask more questions’.

<sup>27</sup> The United States figures too, are comparatively small, with only 1 or 2 percent of civil cases going before a jury, and less than 10 percent of criminal cases, however, 98 percent of the world’s civil jury trials and 90 percent of the world’s criminal jury trials are conducted in America (Starr and McCormick, 2000).

<sup>28</sup> Section 80 reads: ‘The trial on indictment of any offence against any law of the Commonwealth shall be by jury, and every such trial shall be held in the State where the offence was committed, and if the offence was not committed within any State the trial shall be held at such place or places as the Parliament prescribes.’

<sup>29</sup> Article Three states: ‘The Trial of all Crimes...shall be by Jury; and such Trial shall be held in the State where the said Crimes shall have been committed.’

to the individual states by the 14<sup>th</sup> Amendment, and is protected in civil cases where the value of the sum in question is greater than twenty dollars (Cornell University Law School, 2009). Jury trials are subsequently far more common in those types of cases likely to involve scientific expert witness testimony in the United States.

Michael Kirby offered the following explanation for this shift in trial format in Australia:

*First of all, the concern in the 1960's and 70's that jury trial lead to lengthy proceedings, because things had to be explained, which it was said that judges would already understand. I don't think that was ever a very good argument, because jury trial actually lead to the compacting of evidentiary material in order that it would be understood by a lay person. This therefore led to a simplification of the case. Secondly, there was the objection that juries were sometimes unduly generous to plaintiffs, because of a feeling of sympathy towards plaintiffs, in particular, in defamation cases they were too generous, and against the newspapers. This lead to a lot of campaigning by the news media to get rid of them. Thirdly, there is an objection on the part of some people to a jury trial that it involves a determination of matters by lay people who don't give reasons, and reason is said to be a feature of a rational legal system (Kirby, 2010)*

The impetus for the *Daubert* decision also supposedly arose from the manifestation of so-called 'junk science' in American courts. The United States has a considerably different expert witness culture to that of Australia, involving the use of contingency fees and an availability of experts regarded as 'for hire', along with the absence of any rule that costs [generally] follow the event in civil trials (Einstein, 2000). Additionally, the 'loser pays' philosophy that reputedly acts as a disincentive to litigate in this country is not well established in the United States, and Australian courts also tend to award more modest damages compared to similar US cases. As a result, there tend to be fewer speculative or 'fringe' cases heard in Australia, which in turn limits the incidence of spurious scientific claims being made in court (Freckelton, 1997). Kirby has also commented;

*I think that ['junk science'] may be a problem in the United States. In Australia I don't think that has been a big problem. There was a case involving autonomic dyspraxia. That was said to be 'junk science'. I don't know if subsequent advancements and knowledge in that area revealed that it was junk science. You do occasionally get that. But normally in our courts it's a genuine dispute between experts. Judges nowadays will have to resolve these matters, and judges normally have a lot of common sense. They also have a lot of experience. If there is a whiff of 'junk science' being put up to them, they won't like it. I don't know for certain, because its been a long time since I've had anything to do with the conduct of trials, but I*

*don't believe that that is a major problem, or has been a major problem, in Australia*<sup>30</sup>. [...] *I think we've got too many filters for that in the experienced people we appoint to the judiciary.* (Kirby, 2010)

Ian Freckelton agrees:

*We don't have the toxic tort litigation that the US was experiencing - it was mostly in relation to toxic torts that the concerns were arising, and there are other significant differences – they have contingency fees, they don't have the loser paying the winner's costs, which is a very significant disincentive to bring spurious allegations, so there are a number of aspects of the legal culture and environment which are different in the United States. The problems here are not as pronounced as they were at that stage in the States.* (Freckelton, 2009)

Australian judges are almost always appointed from the ranks of senior barristers or solicitors. While Federal judges in the United States are nominated by the President and then undergo a rigorous confirmation process via a Senate committee, most States in the USA allow for the election of trial judges in their respective court systems. There are often no pre-requisites set for candidacy in this regard, and the danger remains that these elections rest on popularity rather than competence. This may or may not influence the competence and motivation of these judges in making rulings regarding scientific evidence in courts, particularly in high profile cases. Most legal scholars denounce this system, but recognise that it is considered a fundamental right of the people. *'I think it's a terrible system. But a majority of our states do it that way. We think that's democracy... I think it's a terrible way to pick our judges.'* (Saks, 2009a) A senior US federal judge agreed with this sentiment, but commented that a change in this system would be very unlikely in the future, noting that to remove a voting process from the general public would likely be seen as an attack on their fundamental rights as citizens.

It was specifically noted in the Australian Law Reform Commission's discussion paper on the uniform evidence acts (Australian Law Reform Commission, 2005a), that *no* comments as to *Frye*, *Daubert*, or any related tests of acceptance of the fields of special knowledge were raised by the stakeholders, suggesting that the Australian legal fraternity do not see either case as relevant to interpreting our own statutes. Several judicial opinions have quoted *Daubert's* definition of *knowledge* in attempts to aid the interpretation of s79 and

---

<sup>30</sup> Yet compare *Clark v Ryan* (1960)

‘specialised knowledge’ in our own UEL,<sup>31</sup> however, the Australian judiciary appear resistant to any notion that the *Daubert* approach is applicable to expert evidence in this country. In *HG v The Queen* (1999), Gleeson CJ specifically indicated that ‘there was no need to determine the relevance of *Daubert* for Australia’ at that time. Similarly, Barr J also avoided any application of the *Daubert* principles in *R. v Gallagher* (2001) stating that ‘...[t]he possibility might have been raised of applying the rule about reliability found in *Daubert v Merrell Dow Pharmaceuticals* (1993) 509 US 579. However, I found it unnecessary to consider questions of that kind...’.

Sperling J<sup>32</sup> noted that the *Daubert* approach had significant limitations, including its narrow focus, the resulting exclusion of theories that have not been widely accepted or validated, and its cost in the form of pre-trial hearings. He recommended that one should be ‘slow to introduce the approach here’, and expressed his view that its application was not warranted in the current Australian legal climate (Sperling, 2000). Kirby J also warned that ‘before adopting [*Daubert*-style] criteria in Australia, careful consideration should be given of the discussion that has ensued following their promulgation in the United States’ (Kirby, 2003). Spigelman CJ made the Supreme Court of NSW’s view abundantly clear in *R v Tang* (2006), where he stated; ‘...I do not mean to suggest that *Daubert* and its progeny in the United States has anything useful to say about s79 of the [Australian] Evidence Act[s]. Rule 702 of the Federal Rules of Evidence (2004), which fell to be interpreted in *Daubert*, is in quite different terms to s79.’

Much of the case law cited by authors in support of a developing trend towards a *Daubert*-style approach in the Australian context appears to have a tenuous link, and it is yet to be explicitly held in any opinion that a *Daubert*-style analysis holds a significant place in Australian case law. Wood (2003) cites *R v Gilmore* (1977) as evidence of a potential *Frye-Daubert* ‘hybrid’ approach in Australia. In this case, the Supreme Court of NSW ruled that evidence of spectrographic voice analysis was found to be improperly rejected by the trial judge as it ‘had reached the standard of scientific acceptance and reliability necessary for its admissibility into evidence [in the United States], and thus, [that decision] should, in my view, be matched in this State’.<sup>33</sup> A careful reading would find that this decision rested largely on the premise of general acceptance, with little more than a passing reference to

<sup>31</sup> ‘...The meaning of “knowledge” in s79 is, in my opinion, the same as that identified in the reasons of the majority judgment in *Daubert v Merrell Dow Pharmaceuticals Inc.* [1993] USSC 114; 509 US 579 (1993) at 590...’ (*R v Tang*, 2006 at 137). Gleeson CJ, had previously described the suggestion that the s79 definition of ‘specialised knowledge’ should follow that given in *Daubert* as ‘highly controversial’ (see *HG v The Queen* (Transcript), 1998)

<sup>32</sup> Then a Justice of the Supreme Court of New South Wales, now retired.

<sup>33</sup> Street CJ, (*R v Gilmore*, 1977 at 941)

‘scientific reliability’ in the *Daubert* context. Other authors also cite *Osland v The Queen* (1998) as evidence of a *Daubert* approach in Australian courts (Moles 2009; Wood, 2003). Yet the reference to *Daubert* in this opinion is only a footnote citing the fact that Battered Woman Syndrome has failed to meet a *Daubert* challenge according to the analysis of various US legal scholars.<sup>34</sup> Callanan J’s opinion does not declare that this testimony should have subsequently been excluded because of these controversies, and does not otherwise discuss the application or endorsement of a *Daubert*-style approach to evaluating scientific evidence in Australia. Similarly with *State Rail Authority of New South Wales v Earthline Constructions Pty. Ltd. (In Liq)* (1999), cited as approaching the ‘spirit’ of *Daubert* by another author (Bell, 2000), there is nothing in the text of the opinion that discusses reliability approaches to evidence, or the *Daubert* decision. The case itself refers to non-expert witness testimony and the role of demeanour in assessing witness credibility, and cannot realistically be construed as being relevant to any claims of support for the *Daubert* approach to assessment of scientific evidence. Recent cases such as *Velevski v The Queen* (2002) demonstrate that support of a *Daubert*-like ‘gatekeeper’ role for judges is virtually non-existent on the judicial circuit.

Gary Edmond has been a vociferous opponent of *Daubert* in Australia. He has recommended that judges should not attempt any reconciliation with *Daubert* in Australian law, given that the decision in the US arose as a result of perceived problems infecting the civil justice system such as ‘junk science’ – a concept of questionable relevance to Australia’s legal and scientific milieu (Edmond, 2000a). He has published numerous articles criticising *Daubert*’s approach, which he feels is based on a romanticised, out-moded and overly-idealistic model of scientific method that bears no resemblance to the way science works in the modern world.<sup>35</sup> That is not to say, however, that Edmond does not believe in the concept of a ‘reliability’ enquiry. In fact, he has stated that he is very much in favour of a reliability approach to assessing the validity of expert evidence, but he does not agree with the *Daubert* model (Edmond, 2008a). He has accused the Australian judiciary of having a ‘general disinterest’ in reliability, which expresses an over-confidence in the traditional means of testing evidentiary reliability such as cross-examination, rebuttal, jury directions and warnings, amid very limited empirical evidence that these techniques have any significant role in highlighting potentially unreliable evidence to the trier of fact. A reliability approach would, in Edmond’s view, have numerous advantages, including less erroneous convictions,

<sup>34</sup> At the time this opinion was written, BWS had not actually been excluded in any actual adversarial proceedings. Since *Osland*, BWS has been ruled as both admissible in the U.S. by the 7<sup>th</sup> Circuit in *United States v Young* (2002), and inadmissible by the 5<sup>th</sup> Circuit in *United States v Dixon* (2005).

<sup>35</sup> See (Edmond and Mercer, 1997; Edmond, 2000a, 2003, 2007; Edmund and Mercer, 2004)

preventing subversion of the courts by ‘cliques’ of experts, saving time and money via exclusion of unreliable evidence, and embodying a flexible legal standard that circumvents the problem of defining abstract concepts like ‘science’ (Edmond, 2008b).

In providing an example of why Edmond believes reliability approaches to assessment of expert evidence are superior, he criticised the Australian criminal justice system for its continued display of bias in its existing approach to expert evidence admissibility, similar to that exposed by Saks, Faigman and others in the United States criminal justice system. The issue in Australian courts has been that of ‘ad-hoc’ expertise. This provision has been recognised since *Butera v DPP* (1987) and continues under the UEL, where it is noted that s79 is sufficiently broad to allow experts who, purely by their repeated exposure to certain facts or evidence, became ‘qualified’ beyond that of a layperson. Therefore, they become ‘experts’ only for the purpose of the particular case – i.e. ‘ad-hoc’ experts. Edmond noted that this provision has only been recognised for the prosecution in criminal cases, and by allowing this testimony, judges seem to forget that s137 of the UEL<sup>36</sup> expressly forbids the admittance of evidence that is likely to be prejudicial to the accused in criminal cases (Edmond, 2008a). He suggests that s79 of the UEL should be amended in order to include a ‘reliability’ requirement, explaining that in doing so, all the judiciary would be asking for was ‘evidence-based’ science to be presented in court, in line with any reputable scientific disciplines’ standards.

Freckelton sees that it is possible, although remote at this time, that a reliability threshold may make its way into Australian law:

*It is possible. What we are seeing in many of the common law decisions is reference to reliability, and of course there is no enunciation as yet of what reliability is being taken to mean. Some of those decision were pre-Daubert, and so they were more vague as a result of that, but to my knowledge Daubert has not been applied in Australia at any stage to give meaning to reliability, so the scenario in which I could see that entering is if Daubert comes in via specialised knowledge, or to give meaning to probative value in the expert context, in which case it may be constructive to identify criteria or indicia to define it. (Freckelton, 2009)*

---

<sup>36</sup> Section 137 of the UEL reads: ‘In a criminal proceeding, the court *must* refuse to admit evidence adduced by the prosecutor if its probative value is outweighed by the danger of unfair prejudice to the defendant.’ (Emphasis added).



Yet evidence from case law suggests the judiciary are reluctant to do any such thing at the current time. Edmond, however, is of the view that a requirement to demonstrate reliability should already be ‘read in’ as part of the current semantic of ss55 and 56:

*I think that they should just read ‘reliability’ into ‘knowledge’ I don’t think that realistically, or any philosophers in our tradition could say that you could have knowledge that wasn’t reliable in some way. I think that’s true of standard dictionary definitions as well. And specialised knowledge as well. The fact that the sections 55 and 56, which deal with relevance in the Uniform Legislation say that the evidence has to be able to ‘rationally effect’ ...so you know its not knowledge of astrology, it has to be something else, a rational thing, specialised knowledge in that way has to be a reliable form of evidence. (Edmond, 2009)*

Despite the ALRC Report noting that a significant number of judicial officers and legal practitioners expressed similar concerns about the lenient approaches to admission of expert evidence, in addition to concerns regarding the quality of such evidence (Australian Law Reform Commission, 2005b), the Australian judiciary appears loathe to amend the current admissibility standard, and are particularly adverse to suggestions of establishing ‘criteria’ for use in assessing expert evidence. The Law Council of Australia have indicated that they favour an approach that focuses on the weight to be given to expert testimony, rather than adopting further constraints on admissibility. Ian Freckelton has admitted that while *Daubert* is a potentially useful tool in deciding the probative value of expert evidence, it is not necessary to further amend the current evidence legislation concerning admissibility, as he believes the discretionary provisions in the existing Act are flexible enough to allow exclusion of unreliable testimony in this spirit. He has argued for criteria in order to assist in the determination of what constitutes ‘unfair prejudice’, and ‘probative value’, however the ALRC dismissed this as unnecessary in their most recent report. They note that determining the admissibility of expert evidence by reference to the scientific validity as suggested by *Daubert* is not always practical or appropriate, especially where these fields inevitably rely on some degree of subjective interpretation (Australian Law Reform Commission, 2005a).

Consequently, the ALRC have recommended less formal options for dealing with concerns over expert witness reliability, such as the training of experts and lawyers, fearing that the adoption of more formal admissibility rules would lead to the courts concentrating on technical formal compliance without regard to their true intent. The ALRC report stops well short of any suggestions that Heydon’s *Makita* criteria should be routinely used by judges in assessing expert evidence, relaying concerns that such an approach may interrupt the smooth

running of trials by requiring such meticulous consideration of expert evidence by the trial judge. However, they have endorsed use of the *Makita* criteria not by judges, but as a preparation tool for expert witnesses. The report advised that if this alone is complied with ‘not only will any admissibility problems be avoided, the expert testimony is likely also to be compelling...’.

## Conclusion

The approach to the admission of expert evidence in Australia is decidedly more liberal to that of the United States. This approach has been guided by a combination of circumstances, case law, legislation and history that differs markedly from the US experience with expert scientific evidence. Australia appears to have different concerns regarding expert witness testimony, and the Australian legislative community, in conjunction with the judiciary, have emphasised approaches to remedying these potential problems that concern alteration of lawyer and expert witness behaviour, rather than attempting any form of statutory or judicial compensation via the introduction of stricter evidence law reforms.

Despite attempts by some members of the legal community to introduce *Daubert*-like reforms into Australian evidence law, there has been considerable resistance from the judiciary and Australian evidence scholars alike. This is undoubtedly due to the perceived ‘failure’ of the *Daubert* reforms in the US, as well as the perception that such reforms are unnecessary in this country. The adoption of the UEL is likely to stifle future attempts at introduction of *Daubert*-like (or *Makita*-like) criteria for admissibility of expert testimony in Australia, particularly as more States take up the mantle of uniformity. The liberal admission of expert evidence in Australia as endorsed by the Uniform Acts is unlikely to change in the near future:

*Once you have a Uniform Evidence Act, which is an attempt to state the whole law on a particular subject, then there is a natural disinclination to go back to the old principles of the common law. You have to really show self-restraint and resist going back into common law approaches. You have to concentrate on interpreting the statute. Once parliament has spoken, what parliament says is the law, and has a higher legitimacy than what judges have said in the past. Therefore, generally speaking, the High Court of Australia has insisted on examining what the parliamentary provision is, and what the words mean. That was illustrated in Weiss v The Queen<sup>37</sup> ... I believe it will be the same with the Uniform Evidence*

---

<sup>37</sup> See *Weiss v The Queen* (2005)

*Act. There will be a lot of resistance to trying to elaborate and to gloss with judicial dicta the provisions of the Evidence Act.* (Kirby, 2010)

While there is general agreement that ‘junk science’ has not been a significant problem in Australia, the question remains as to what problems forensic science faces in this country. These, on the face of things, do not appear to be vastly different to those faced by forensic science in the United States or elsewhere. The legal fraternity in Australia, perhaps rightly, are loathe to introduce reforms within their own sphere of influence in order to address criticisms of forensic science expressed by others. *Daubert*, as a corrective mechanism for some of these concerns, appears to have little chance of becoming extant in this country. Courts in Australia have been resistant to the notion that criticism of forensic science should subsequently lead to its exclusion, as evidenced by the historically liberal attitude to admission of expert testimony. Courts cannot, and should not, be relied upon to instigate a proper assessment, complete with punitive measures, of the reliability of any given discipline. Consequently, it is left to those who practise within the field to meet these criticisms head on. This requires a careful understanding of the criticism, as well as a detailed investigation of the discipline’s own standing and practices. It is an investigation of this aspect of the reliability of forensic odontology, in particular bitemark analysis, that we now turn.

## CHAPTER 5

# Uniqueness and Individualisation in Forensic Science

---

Of the fundamental theories said to underpin identification of individuals from not only bitemarks, but also fingerprints; frontal sinuses; bullets from the guns which they were fired; and so on, one of the most consistently adhered to is that of the uniqueness of the particular feature in question. Long cited as one of the cornerstones of forensic identification, the concept of ‘uniqueness’ has received renewed scrutiny thanks to several authors, who have argued that there is no scientific basis for such a claim when considering human physical characteristics (Cole, 2009; Kaye, 2009; Saks and Koehler, 2008; Saks, 2009b). Given that the fundamental tenet of bitemark analysis theory is itself supposedly that the dentition is unique, this chapter explores the origins of the uniqueness assumption in addition to the evidence for it, in an attempt to establish its role in forming an evidence base for forensic conclusions of identity.

## The Uniqueness Fallacy

There appears to be widespread assumption among forensic practitioners that any given forensic trait can be considered unique to an individual. A web-base survey of 72 forensic odontologists found that 91% of respondents believed that the human dentition was unique (Pretty, 2003), and 78% of those believed that this uniqueness could be represented on human skin during the biting process. An Australian forensic odontologist made the statement in a radio interview in the year 2000 that ‘... there's plenty of evidence and research going back to the early '60s which suggests a human bitemark is totally unique to the individual’ (See *Law Report*, 2000), suggesting that opinions in this country may not be dissimilar.

Forensic dental experts, as well as fingerprint experts, firearms and toolmark examiners, and anthropologists have continually invoked this principle as support for their conclusions of identity. In one of the seminal bitemark cases in Australia, *R v Carroll* (1985), the uniqueness principle was cited as support for the identification of the perpetrator via a

bitemark left on the murder victim's thigh. The following exchange occurred during the examination of the chief prosecution witness, Dr Romaniuk.

[Counsel] Are persons' dentitions unique? — [Witness] Indeed, yes.

In what sense? Could you elaborate on that? — There is no one individual – every individual is a unique person, and the dentition is a unique feature of that person; just as all our facial features are different [...] and so the dentition is a unique feature. (Trial transcript, p405)

And later:

[Counsel] Is a unique pattern created? — Yes, it is very unique indeed because of the shape of the teeth – indeed the relationship of teeth to teeth; the presence of defects in the anterior teeth; the nature of the relationship of the lower teeth to one another, and also the whole picture when one takes the dentition as a whole – it is a unique dentition. (Trial transcript, p411)

In the earlier committal hearing for the same defendant, *John Philip Reynolds v Raymond John Carroll*, (1984) another odontologist also invoked this principle:

[Witness] But bearing in mind [...] the fact that no two dentitions are identical, I think it is highly unlikely that other teeth would have produced marks with that same pattern.

[Counsel] So you say that there are no two dentitions the same? [Witness] Yes. Well, I would say that [...] no two dentitions are identical. (Committal Hearing Transcript)

Recently, this assumption of uniqueness has come under fire from several sources, the most prominent being the 2009 National Academy of Sciences report (National Research Council, 2009) which concluded that 'in most forensic science disciplines, no studies have been conducted of large populations to establish the uniqueness of marks or features.' Such criticism had previously been voiced by other academics, who not only question the so-called 'proof' of individuality posited by forensic practitioners, but the entire absurdity of claiming such a notion (Cole, 2009; Kaye, 2003a, 2009; Saks and Koehler, 2008). Like that of other disciplines, the notion of uniqueness in odontology often relies on tenuous proof such as anecdotal evidence and experience. Furthermore, literature regarding the uniqueness of the dentition<sup>1</sup> that has been cited in the courts since the mid 1970's (for example, in *People v Smith*, 1981; *People v Milone*, 1976; and *R v Lewis*, 1987) when critically examined all suffers major flaws, including the application of false logic, faulty reasoning, and erroneous mathematics, which call their conclusions of support for the uniqueness proposition into question.

---

<sup>1</sup> The earliest of which could only be traced back to 1974, and not the 1960's as suggested by the aforementioned Australian odontologist.

## The Ideological Origins of the Uniqueness Principle

The concept of ‘uniqueness’ has more the qualities of a cultural meme than a scientific fact (Mnookin, 2001). ‘Meme’ was a term first coined by Richard Dawkins in the 1970’s (Dawkins, 1976), has been refined by several others since (see Wilkins, 1998), and describes a ‘unit of cultural transmission’; a piece of ‘thought’ transmitted from person to person. In this context, one explanation for the origin of the parallel tenet ‘nature’s infinite variation’ was that it then provided support for the existence of God. The reasoning of the 17<sup>th</sup> century philosopher and polymath Gottfried Leibniz represented that of the time: God is infinite and God created the Earth, therefore nature, as God’s creation, reflects His infinity (Leibniz and Rescher, 1991; See Stevenson, 2005). To deny nature’s infinite variation was to deny God’s infinity – a claim that would have been considered heresy. A similar line of reasoning gave rise to the maxim ‘nature abhors a vacuum’; God, with his infinite power, would not have created ‘nothing’; therefore a vacuum simply could not exist as it would imply the absence of God.<sup>2</sup>

More recently, 20<sup>th</sup> century philosophers and sociologists have posited that there is a fundamental human desire to see oneself as unique, stemming from a primitive emotional need. Snyder and Fromkin suggested in their 1980 essay that humankind is most emotionally satisfied when we perceive at most only an intermediate level of similarity relative to other people (Snyder and Fromkin, 1980). The level of personal emotional satisfaction supposedly falls dramatically when persons consider themselves too similar to others, thus humans have a fundamentally strong desire to *want* to believe in uniqueness that is difficult to dislodge, even in the absence of definitive proof.

Moenssens has stated that the belief in the individuality of forensic traits rests on what he terms the ‘Snowflake Syndrome’ (Moenssens, 1999a). Demonstrating another instance of the pervasiveness of the cultural meme, snowflakes are often singled out as the ‘paradigm of uniqueness’ (Thornton, 1986) and are often cited by others as the archetypal analogy to the concept of individualisation in the forensic sciences (see Inman and Rudin, 1997). In the late 1800’s, a young farmer invented a novel way of photographing ice crystals and began documenting snowflakes. Just before his death, he declared that he had never seen any two

---

<sup>2</sup> This was questioned in the 17<sup>th</sup> century after Torricelli’s experiments involving the early mercury barometer, and the existence of a vacuum was definitively proven possible by Blaise Pascal in the late 1640’s (Close, 2009). In the face of this irrefutable evidence, a different line of reasoning was then hastily adopted in order to explain God’s role in all of this. The reasoning was now that of *course* a vacuum could exist, because to claim that it couldn’t was to claim that God’s powers would be somehow limited; He can create anything He sees fit, including the creation of ‘nothing’ where ‘something’ previously existed.

alike, after taking 5,381 photomicrographs. This gradually morphed into the concept that no two snowflakes could *ever* be alike. Perhaps realising that evidence of a mere 5,400 snowflakes, is somewhat meagre to stake such a claim on, society has sought other ways in which to justify this belief. Most people appear to believe that no two snowflakes are alike because of the enormously large number of possible ways to arrange the approximately  $10^{15}$  molecules in the average snowflake (Thornton, 1986),<sup>3</sup> which by contrast, dwarfs the number of snowflakes predicted to have ever existed.<sup>4</sup> To say, as Thornton does, that this number is ‘virtually’ beyond human comprehension is somewhat of an understatement.

Yet the mathematics and logic behind this evaluation is deeply flawed. Just because such arrangements of molecules are mathematically possible does not imply that they are physically possible, and almost certainly some arrangements simply could never exist, simply due to the restrictive nature of inter-molecular interaction in solid crystals. A further problem with the mathematical gymnastics employed by this attempt at providing proof of uniqueness is that similar arrangements would undoubtedly be perceived as exactly the same arrangement at the observer level (Kaye, 2003a). In fact, atmospheric researchers often claim that it is not impossible for two snowflakes to look the same (albeit highly unlikely), but most still hold to their tenet by explaining that on a molecular level, they would be revealed to be different (Roach, 2007). Uniqueness at a molecular level is hardly a useful foundation for the forensic scientist concerned with fingerprints, teeth, or bullets. Two snowflakes that *were* supposedly visually identical were found in 1988 by a researcher who was documenting snowflakes for the National Center for Atmospheric Research, yet this does not appear to be widely known<sup>5</sup> (Henson and Politovich, 1995; Kruszelnicki, 2006; Schmid, 1988).

Fingerprints have been held as a paradigm of uniqueness since the early 20<sup>th</sup> century, and have even been used as ‘proof’ of claims of forensic uniqueness such as that of fingernails merely by anatomical association (see also Starrs, 1999; Stone and Wilmovsky, 1984). The high esteem of fingerprints is reflected in the modern terminology DNA ‘fingerprinting’. Few realise that this belief in the uniqueness of fingerprints, first suggested in the late 19<sup>th</sup> century, represents a continuation of a particular notion that fitted the appropriate cultural

---

<sup>3</sup> The incredible figure of  $10^{2250}$  possible arrangements.

<sup>4</sup> A mere  $3.12 \times 10^{31}$  snowflakes.

<sup>5</sup> Or accepted.

ideology of the time,<sup>6</sup> and through social and cultural forces has come to be accepted by future generations — it is not the result of robust scientific evidence (see Cole, 2001).

## The Evidence for Uniqueness

The evidence cited for uniqueness includes the anecdotal and experiential, biological, and mathematical, yet all of these approaches suffer major disadvantages that result in little faith being able to be afforded to their conclusions. A critical discussion of the so-called ‘evidence’ for uniqueness is necessary in order to ascertain the true standing of this supposedly fundamental tenet of forensic identification.

### Anecdotal Evidence and Experience

Fingerprint examiners, odontologists, handwriting experts, firearms and toolmarks experts, have all at some stage argued that uniqueness exists because as yet, in the history of their discipline, no two objects have been found to have an exact duplicate. These claims of ‘never observing an identical match’ may be true, but it relies on the assumption that every examiner remembers the details of every object ever examined, and, even if only subconsciously, has ‘compared’ all of the objects he has happened to examine with one another. Such a proposition is highly dubious, and relies on claims and ‘observations’ that have neither been recorded nor compiled (Cole, 2009). More importantly, attempts to affirm the uniqueness proposition using anecdotes and experience rests on pure inductive reasoning. Several authors have cited induction as being inappropriate for application in forensic science research (Faigman, 2007; Meuwly, 2006), noting that ‘science’ arising from pure inductivism gave us the practices of blood-letting and phrenology. As an approach to scientific methodology, induction fell out of favour during the epoch of Sir Francis Bacon in the 16<sup>th</sup> century. David Faigman illustrates the citation of Baconian methods in forensic science by describing how the former chief of the FBI’s latent fingerprint unit, Stephen Meagher, once described his approach as being experiential and inductive:

*I laughed when he said that – because it’s a very Baconian kind of idea that you are just going to be an inductivist. It was just my reaction to the fact that Meagher actually said something as silly as that. Part of it is a paradigm difference. I think that ours is the 21<sup>st</sup> century paradigm and forensic science’s is the 16<sup>th</sup> century paradigm. (Faigman, 2009)*

---

<sup>6</sup> Galton’s work on the individuality of fingerprints originally began with an attempt to classify individuals by racial origin, using their fingerprints as a marker. This research was eventually abandoned when Galton could find no discernable link between fingerprints and race.



Pure inductivism leads to the ‘induction problem’, which can be simply stated as ‘knowledge can never exist from purely inductive reasoning’. This is partly because induction necessarily invokes a positivistic approach, where the theory rests on the assumption that the future will resemble the past, yet there is no logical reason to assume this (McLachlan, 1995). The work of Hume demonstrated that there can never be any certainty in induction, even for those inductions that have never been disproved by the occurrence of a contrary (Russell, 2004). Accumulation of positive instances simply cannot lead to a conclusion of certainty.

As an illustration of the failure of inductive reasoning in the use of anecdotal and experiential evidence, consider the well-known tale of the coelacanth. This rare species of fish was thought to have been extinct since the end of the Cretaceous period, some 80 million years ago. This thinking prevailed from the year 1836, when Agassiz described and named the coelacanth from the fossil record (Thomson, 1991). For the next one hundred years it was presumed to have been extinct, as no one had ever come across a living specimen, however, a live coelacanth was caught off the coast of South Africa in 1938. Several living coelacanths have since been caught off the coasts of East Africa and Indonesia, (Erdmann et al., 1998); the reasoning that the species was extinct simply because no one had ever found a live one was seen to quickly fail.

Judges in the United States are becoming increasingly aware of this somewhat flawed attempt at bolstering identification conclusions via the use of ‘experiential’ evidence. Judge Albin, delivering the majority opinion of the appellate court in *State v Fortin* (2007) noted that the original trial judge was not satisfied with the expert’s conclusion that the bitemarks, or even the pattern of infliction of multiple bitemarks, was unique, based on their experience alone, despite the claim that they had reviewed ‘thousands’ of cases in the course of their professional lives. The trial judge requested production of ‘a compilation of the sexual assault and homicide cases with human bite marks on victims, or a reasonable sampling of [such cases] from their experience’ – in other words, some form of data to support their reasoning. This decision was later upheld by the appellate court.

The experts in *Fortin* follow the general argument seen for ‘proof’ of uniqueness via observational experience, which follows the lines of: ‘ $n$  number of samples have been observed, and none of them are the same. Therefore, one can assume that no sample will ever be the same as any other.’ McLachlan, a philosopher, has directly addressed the use of this principle in fingerprint testimony, where he noted that evidence from a mere sample of human beings gives us no reason at all for doubting the existence of an identical set of

fingerprints somewhere in the world. Additionally, he noted that there is no mechanism which *prevents* the occurrence of two people sharing the same fingerprint, no ‘check and balance’ to ensure that once one fingerprint pattern has manifested itself in human friction ridge skin it will never appear again (McLachlan, 1995). This reasoning is clearly applicable to any forensic characteristic, including bitemarks.

Using an example proposed by Saks and Koehler (2008), it can also be mathematically demonstrated that the use of experience fails to advance the uniqueness argument to any significant degree even in significantly smaller populations than that of the entire world. Suppose that *were* 100 pairs of individuals who happen to have indistinguishable arrangements of the anterior dentition from each other. Let us say there were 100,000 people in total in a particular town. This means there are  $[100,000(99,999) \div 2] = 4.99 \times 10^9$  possible ways of pairing two people’s dentitions in order to compare them. Lets say that 10 forensic odontologists each compared 100 dentitions a day for ten years, so that between them, they managed  $3.65 \times 10^6$  comparisons – over one hundred years of ‘collective’ experience. The chance that they would miss all one hundred matching pairs is approximately 93%.<sup>7</sup>

It is obvious to most people that the lower the incidence of a trait (i.e. the lower the probability of observing that trait – ‘*p*’), the less likely it is that it will appear in a sample. Given that sample sizes are finite, this presents a problem when *p* is small and *N* is large, as the chance of *not* observing a match within the sample then becomes very large. As the sample size is limited by human endeavour, we quickly reach the realisation that proving uniqueness this way is unlikely to be successful. The problem is thus: we want to prove that something is so rare that it could be considered unique, but because it is so rare it is more likely than not that an observance would be missed in any reasonable sample size. This can be partially addressed via the use of probability theorem and the product rule, and its usefulness in this regard will be discussed later in this chapter.

### **Knowledge about the process of formation**

Another reason for claiming that certain physical characteristics are unique stems from knowledge about how these characteristics are formed. In actual fact, such reasoning technically relies on the *lack of knowledge* about how these characteristics are formed. Fingerprint examiners often argue that because friction ridge formation is induced by the stresses and strains experienced by the foetus *in-utero*, which are random and infinite, it is likely that they subsequently produce a random, infinite variety of friction ridge patterns

---

<sup>7</sup> See Saks and Koehler (2008) for a more detailed mathematical description of how this is calculated.

(Budowle et al., 2006). Such reasoning begs the question rather than answering it, as it provides no better grounds for assuming that the causal process itself is infinitely variable. Additionally, McLachlan (1995) again noted that there is no logical reason why the same effect cannot be caused by different processes. To make such a claim regarding the variety of forces acting upon friction ridge skin *in-utero*, it would be imagined that some form of study should be conducted in order to assess the scope of variation in such variables as pressure, fluid dynamics, skin tension, temperature and so on, before concluding that such variables are infinite, or even extremely large. Such study appears yet to be conducted (Cole, 2009).

A similar argument occurs in bitemark analysis. It is well known that the spatial arrangement of the dentition is partly influenced by local environment factors, such as the natural forces exerted by the oral musculature, and other external forces such as oral habits; tongue thrusting, thumb sucking and so on. There is no logical reason to assume that because these forces remain largely unquantified in terms of their scope and their relative influence on the position of the teeth that they are infinite, and that this invariably means that the arrangement of the teeth is also infinite. Without studies demonstrating the magnitude and variety of forces acting on the dentition, assuming they are random and infinite is pure conjecture.

### Studies

A modern solution to the induction problem was proposed by the early 20<sup>th</sup> century philosopher of science Karl Popper, who gave renewed credence to the deductive method of reasoning (Popper, 2002). Pure deductivism involves the concept of falsification, which alas has also been criticised as a method of attaining knowledge (Stove, 1982), but nonetheless, refinement of Popper's work has given rise to the modern theory of attainment of scientific knowledge, via the hypothetico-deductive method. This involves the use of a cyclic model incorporating hypothesis formation, testing, and re-formulation (see Gimbel, 2011) and is preferable to pure inductivism because it specifically attempts to *disprove* the theory by continually attempting falsification, whereas the inductive method relies only on the tenuous accumulation of positive instances. While the hypothetico-deductive model results in more meaningful proof, the quality of this proof is still open to criticism when these studies make unrealistic assumptions regarding the model used to test the theory, or draw conclusions outside the realm of logical inference.

### **Observational Studies**

Forensic practitioners have relied on observational studies as evidence for the individuality of the frontal sinuses from as early as 1935 (Mayer, 1935).<sup>8</sup> Yet as McLachlan stated, evidence from a random sample of human beings does not necessarily give us reason to believe that a particular characteristic is unique. He illustrates this by inviting us to consider a random sample of, say 400 human beings. This is likely to demonstrate that none of them have the same mother; but we know that to conclude that not one person on the Earth shares the same mother defies common sense.

In other studies cited as evidence for the uniqueness of the frontal sinuses, Harris examined 32 individuals (Harris et al., 1987), and Asherson studied the radiographs of 74 monozygotic and heterozygotic twins. All of these studies concluded that the frontal sinuses were *always* different (Asherson, 1965). This is not disputed for the samples studied, but extrapolation to the world population from such a small sample seems to be nothing more than a great leap of faith. Similarly, a study that concluding ‘that palatal rugae were unique and identification could be based on their comparison’ was based on the ability of four dentists to match 25 post-orthodontic maxillary casts with their pre-orthodontic counterpart (English et al., 1988). The study by Kieser and colleagues (Kieser et al., 2007), on the individuality of the dentition considered only 33 maxillary casts and 49 mandibular casts from a total of 50 individuals. Kieser replied to criticism of this small sample size by Bowers (Bowers, 2007b) in stating that the study does not attempt to single-handedly ‘prove’ the uniqueness argument, but contributes to the gradual acquisition of evidence ‘point[ing] towards the uniqueness of the anterior dentition.’ (Kieser and Bernal, 2007) This is a fair reply, however one needs to consider the size of the sample in relation to the tens of thousands of 17-20 year olds that have post-orthodontic occlusions as a result of treatment in New Zealand alone<sup>9</sup> before deciding how much weight this adds to the uniqueness argument.

Rational thinking suggests that the larger the sample size  $n$  then the more faith one can have in such a conclusion. Just when  $n$  is large enough to allow extrapolation to a population  $N$  is

---

<sup>8</sup> Mayer additionally suggests that Dr Poole’s ‘...special study of the sinus prints of the insane [...] will also throw additional light on th[e] important problem [of insanity]’; presumably in addition to providing a means for identification. Such a comment throws more doubt on the scientific method of the day than it does anything else.

<sup>9</sup> See (Marriot et al., 2001). Their survey conducted in 1999 indicated that the average load of an orthodontist in New Zealand was 371 patients, based on a survey sample of 56 out of a total of 71 practicing orthodontists at that time. 77.9% of these were considered children (not defined, but let us conservatively assume that this means under 18). This means that approximately 20,500 patients under the age of 18 were receiving orthodontic treatment. Of course, not all of these would have full fixed appliances, but when considering the large number of dentists who perform similar services in New Zealand, it is easy to arrive at numbers in the tens of thousands when considering the number of 17-20 year olds who would have experienced similar orthodontic treatment in New Zealand alone.

a matter of debate, however mathematicians and statisticians have developed generally accepted tools to assist with this decision. The calculation of sample size is based on a statistical equation that will depend on how certain one wants to be that the induction would hold true. This level of ‘certainty’ is often referred to as the confidence level, and is a well-known concept in scientific studies. Most studies consider a confidence level of anything less than 0.95 unsatisfactory, although sometimes a confidence level of 0.9 is used. Studies that have used this calculation to determine sample size allowing for a confidence level of 0.95 arrive at a sample size of approximately  $n=385$  in order to feel confident above  $N = 100,000$ .<sup>10</sup> However, this statistical tool was developed for determining sample sizes for surveys, and is not applicable for instances when the likelihood of observing a trait is rare as there is too great a risk that *no* positive instances would be recorded in a sample.

Courts, or at least lawyers, have recognised the futility of the observational study with regard to proving identification of a suspect, if only from a philosophical rather than a mathematical perspective. In *R v Carroll* (2000), counsel for the defence pressed the prosecution witness on this issue. The odontologist, in an attempt to bolster his claim that no other dentition could possibly have caused the mark on the deceased baby’s leg, carried out a pre-trial ‘experiment’ involving a ‘random’ (although not truly so) selection of individuals for comparison with the mark.

[Witness] ... we took 25 sets of similar models of people from similar ages and subjected them to the same analysis, and, indeed, we did that to other people who could have conceivably have been involved, including Mr and Mrs Kennedy, and we were unable to demonstrate any relative match at all with any of those sets of teeth. (Trial transcript, p587)

Defence counsel, in his cross-examination, then argued that this proved nothing, a notion to which the witness eventually agreed;

[Counsel] But of course, none of those people were suspects, were they? — [Witness] No, they weren’t.

So really, aren’t they just simply randomly selected people from the public... — Absolutely.

...who happened not to have killed Diedre Kennedy? — That’s correct. And who were sharing the same age as the suspect at the time that the crime was alleged to have been committed, yes.

But obviously, there’s hundreds of thousands of people across the State who would fit in that category, isn’t there? — Absolutely.

---

<sup>10</sup> Above  $N=20,000$ , the sample size does not increase significantly due to the limits of the equation. The asymptote occurs when  $n = 385$ , theoretically allowing for  $N$  up to  $\infty$ . Note that these numbers also rely on an assumed confidence limit of 0.05.

Right. I'm just struggling with what you actually say you proved by doing that? — Oh, we didn't prove anything [...]

[...]

[Counsel] What that means though, at the end of the day, is just that simply none of those 25 people matched up with that bitemark? — [Witness] Absolutely.

[...]

[Counsel] So whether you can get 25 or 25000 others with fewer points of, you say, concordance, that still doesn't prove anything in relation to the identification of Mr Carroll, does it? — [Witness] It was never intended to prove anything. It was simply intended to run the experiment to demonstrate the validity of what we were doing. The same reasoning is true of all bite mark analyses. (Trial transcript, p621)

The expert witness eventually agrees that this sort of observational study (or experiment), whereby matches are searched for in a given population, fails to advance the science behind identification from bitemark analysis. Of course, mathematical sampling is fraught with difficulties in applying specific theories to a large population, even for straightforward studies involving surveys. The task of adequate sampling becomes even more difficult when the traits of interest are rare, yet there are other methods that enable researchers to give more weight to the results than could otherwise be placed on simple frequency data.

### ***Combinatorial Modelling***

One such method involves the use of combinatorial functions. We have already seen that small sample sizes defy extrapolation to a large population, particularly when the likelihood of a match itself is considered rare. Researchers can abrogate this issue, at least to some extent, by relying on statistical modelling of *combinations* of traits that are known to exist in a population. Observance of the trait itself (ie the pattern of the dentition as a whole) might be rare, but the observance of each individual characteristic that makes up the trait (ie the position of each individual tooth within a dentition) as a whole is probably not. By examining each individual characteristic, and applying combinatorial theorem in order to then develop a probability for observing the trait as a whole (the trait being made up of different 'combinations' of characteristics), modelling in this fashion reduces the mathematical error introduced by the likelihood of 'missing' an observance in a simple observational sample. This is the basis for statistical analysis of DNA frequencies in the population and has also been attempted<sup>11</sup> in areas such as fingerprints and forensic odontology.

---

<sup>11</sup> With arguably less believable results, as we shall see.

One of the main concerns with the application of this technique to forensic characteristics is that the derivation of the probability that a particular characteristic will occur is often questionable. Many early studies simply assumed base-rate probabilities without attempting to verify the frequency of the trait in the population. Galton's work (Galton, 1892) is often lauded in the fingerprint community as proof of the individuality of the fingerprints. He calculated that the probability of a specific ridge configuration occurring in the population was  $1.45 \times 10^{-11}$ . As the population was approximately 1.6 billion at the time (16 billion fingers), he concluded that the odds of finding another finger with the same ridge detail were approximately one in four. Numerous scholars criticised this crude estimation from as early as 1930, because in deriving this final probability, Galton simply estimated the frequency of particular ridge details in the population and made no attempt to experimentally verify these assumed frequencies (Cole, 2001; Pearson, 1930; Roxburgh, 1933; Stoney (in Lee and Gaensslen, 2001); Stoney and Thornton, 1986). Galton similarly estimated that the probability of observing a particular pattern type, such as a loop or arch, was 1 in 16. Likewise, he assumed that he could predict ridge patterns, given a partial print, with a frequency of 0.5, and that the probability of the number of ridges entering a particular area of the print was 1 in 256. These were the key numbers he used to arrive at his final probability of  $1.45 \times 10^{-11}$ , however, *none* of these numbers were based on actual population data.

Similarly, the flawed assumption of the presence of ridge minutiae (ridge endings, bifurcations and so on) being *equally* probable abounds in other attempts to 'prove' the uniqueness of fingerprints (see Stoney and Thornton, 1986). This assumption fails to realistically describe the situation, as is easily demonstrable by population frequency data. Additionally, most of these studies have also assumed that distribution of minutiae is random across the surface of the friction ridge skin, and therefore have modelled their analyses on a random probability distribution. This assumption too was proven false in the late 1980's by Stoney (1988), and has been recognised as contributing to significant over-estimates of the likelihood of 'uniqueness' for fingerprints (Chen and Moon, 2008). Several other studies (Champod and Margot, 1996; Pankanti et al., 2002) use complicated mathematical and computer-generated models in order to calculate the possible number of combinations of fingerprint ridge minutiae. The key feature of these studies is that they calculate tiny probabilities from large numbers of possible combinations. The fact that small probabilities are predicted perhaps gives the study an aura of credibility, but the failing of all of these experiments is that the authors do not attempt to verify their model assumptions with that of data derived from the population. It is therefore impossible to ascertain to what degree their model represents the true distribution of such traits in the real world.

Two early papers in forensic odontology make similar, flawed assumptions. Rawson's attempt at proving the individuality of the human dentition initially assumed that the probability of each tooth occupying a particular position was 1/6 (Rawson et al., 1984). Keiser-Nielsen's analysis relies on the assumption that the likelihood of a tooth being missing or restored is exactly the same for each tooth (Keiser-Nielsen, 1975, 1977, 1980). It is well established that these assumptions do not hold. For example, population studies clearly demonstrate that third molars are more likely than canines to be congenitally absent, incisors are more likely to be missing than canines due to trauma, and so on.<sup>12</sup>

Approaches that rely on assumed probabilities are essentially using combinatorial theorem to calculate the number of *possibilities* of combinations, and the problem arises when researchers then equate this with the *probability* of any one of these combinations appearing in the population. Keiser-Nielsen's paper is considered one of the seminal works in forensic odontology, and appeared in various forms in the late 1970's. Its earliest appearance was as a paper at the 1975 Continuing Education Course in Forensic Odontology (Keiser-Nielsen, 1975). It then appeared in the now discontinued journal *Forensic Science* (Keiser-Nielsen, 1977), and again later in a textbook (Keiser-Nielsen, 1980). Keiser-Nielsen applied the reasoning that if the number of possible combinations of twelve or sixteen 'concordancies' used for fingerprints are so enormous, then the presence of 32 teeth gives us a similar if not better starting point for concluding the same thing about the dentition. He calculates that the maximum number of possible combinations of 16 restored teeth is 601 080 390, using a simple combinatorial formula that assumes the remaining 16 teeth are intact.

Keiser-Nielsen stated 'whenever 16 teeth are missing in the mouth of a given person, dead or alive, then his actual and personal combination is one out of more than 600 million possible combinations'. This falsely implies that the chance of observing one particular arrangement of teeth in a person with 16 teeth is greater than one in 600 million, which is fallacious in the extreme. The mathematics is the same as that used to invoke the argument for the existence of identical snowflakes. Simply because there are 600,000,000 ways to 'arrange' 16 restored teeth says nothing about the probability of encountering any such combination in the population. Keiser Nielsen has quantified the *possibilities* regarding where a tooth might be restored [or missing], but has completely failed to address the issue of *probability*. This study cannot be used to infer uniqueness in the population simply because it uses no population data whatsoever to support its claims.

---

<sup>12</sup> For some of the latest population data, see Brown and Goryakin, 2009; Lesolang et al., 2009; Starr and Hall, 2010; Upadhyaya and Humagain, 2009.



Attempts at proving the uniqueness of impressed toolmarks have fallen into a similar misconception, with practitioners quoting from flawed, highly theoretical analyses of the variability of toolmarks as support for their claim that no two marks are identical (see Nichols, 2007). As an example of this type of analysis, relying on the resolving power of the microscope and a shape characterisation of the forensically significant marks that could theoretically present on the surface of a hammer, Stone (2003) calculates the number of possible positions for point, line and curve characteristics. Arriving at astronomical figures, invoking the product rule, the implication is that no two marks could ever be the same because of the incredibly large number of possibilities. Criticising Stone for being too *conservative*, Collins modifies these calculations (Collins and Stone, 2005) and concludes that on practical level, toolmarks are unique, but again fails to realise that arrival at a certain number of possibilities says nothing about the relative frequency of occurrence on an actual tool face.

### ***Combinatorial Modelling from Population Data***

Alas, the use of population data still fails to alleviate doubt regarding the conclusion that any particular characteristic is unique. Data used in any model is ideally derived from the population via some form of sampling. Therefore, a fundamental concern in any data derived from a survey is the risk that the samples are not truly random or representative of the target population. Representation of the entire world is a very difficult goal to achieve in a finite sample. MacFarlane's study on the quantification of the anterior dentition is based on figures derived from a sample of 200 emergency patients presenting at the Glasgow Dental Hospital (MacFarlane et al., 1974). Such a sample is not truly random, and would not necessarily accurately represent the population of Glasgow itself,<sup>13</sup> let alone that of the Earth.

Yet just because a study has gathered data from the population in a truly random way still does not necessarily give the study more weight. Rawson and colleagues' study (Rawson et al., 1984), cited by many authors as proof of the individuality of the dentition (Drinnan and Melton, 1985; Pretty and Sweet, 2001c)<sup>14</sup> and among the first to use population data, conclude that they have 'demonstrate[d] ... the uniqueness of the human dentition beyond any reasonable doubt'. Unfortunately, the authors negate any benefit this data may have had by failing to use it in a mathematically sound way. In this study, the positions of each of the

---

<sup>13</sup> Despite this flaw, a calculation of the incidence of a particular dental 'signature' derived from this study arrived at the incidence of 1:10,000 and was used without objection in a Scottish court in the late 1970's (MacDonald and Laird, 1977).

<sup>14</sup> Although both of these sets of authors correctly conclude that it did not address the question of the uniqueness of *bitemarks*. Pretty and Sweet notably state 'Rawson has proven what his article claims' regarding his proof of uniqueness of the *dentition*. Analysis of Rawson's work reveals several causes for doubt that this is the case.

twelve anterior teeth recorded in 397 wax bites were recorded. The number of positions observed was defined by the polar co-ordinates  $x$ ,  $y$  and angle  $\emptyset$ . The number of positions for each tooth ranged from 150 (for the upper right canine) to 239.9 (for the upper left lateral incisor). The authors then calculate, using the product rule, that the probability of finding two sets of dentition with six teeth in the same position for the maxillary arch is  $2.2 \times 10^{13}$  and for the six mandibular teeth,  $6.08 \times 10^{12}$ . The total number of positions for all twelve teeth is therefore deemed to be  $1.36 \times 10^{26}$ . Even their conservative estimate, using the minimum number of positions as 150, results in a figure of  $(150)^6 \approx 1.14 \times 10^{13}$  for each arch. This might be true if the probability of encountering a tooth at any one of these positions was equal, but their own data itself proves that it is not, and so this mathematics is fundamentally flawed.

Rawson and colleagues fail to correctly use their frequency data to calculate the probability of a tooth *actually occurring* in any given position. For example, the range of  $x$ -coordinates for the centre point of tooth Participant #6 was given as  $136 \leq x \leq 161$ , allowing for 26 distinct possible tooth positions along the  $x$ -axis<sup>15</sup>. Therefore, using the reasoning applied in their concluding calculations, the ‘probability’ that a tooth would occupy one of these  $x$ -axis positions is given as  $1/26$ , or  $0.03846$ .<sup>16</sup> However, their graphical data does not support this. At  $x = 150$ , the data shows a frequency of no less than 63 occurrences. Therefore, the probability of a tooth occupying the polar coordinate  $x = 150$  should be more correctly calculated as  $63/397$ , or  $0.1587$ . Considering the probability of the tooth occupying  $x = 161$ , using the frequency data they provide (the frequency is 1)  $P(x=161) = 1/397$ , or  $0.00252$ . Thus the probability of a tooth occupying any given position is *not equal for all positions*, and thus the probability of observing any particular tooth in any give position depends on total number of positions available *and* the relative distribution of frequencies. They have calculated the first variable (number of possible positions) as  $1/26$  and attempted to use that to calculate the overall combinatorial probability. They have completely ignored the second variable, that of the distribution of the frequency, which suggests that the likelihood of observing a tooth in one position is actually far greater than observing it in other positions. Their results are no better than Keiser-Nielsen’s in that they have essentially only calculated possibilities and equated this with probability.

---

<sup>15</sup> It is unclear in the study as to whether the digits 136...161 in the figure represent the  $x$ -axis coordinates, or simply line numbers in the computer program used to graph the frequencies. For clarity, I shall assume that it represents the  $x$  coordinate, but even if it doesn’t it is irrelevant as the data clearly suggests that there were at least 26 positions observed.

<sup>16</sup> How they derived the number of possible positions for this tooth along the  $x$ -axis as only 5.0 remains unexplained.

Of course, even if Rawson had calculated the true probabilities from his population data, he should not make assumptions about the type of distribution of the frequencies, which determine the formula used to relate the two variables. Without conducting a proper analysis of the frequency data, one cannot assume that the distribution is a Normal one. Other distribution models, such as the Binomial, Gaussian or Poisson distribution may more adequately describe the frequency data, in which case the formula for calculating probability is different. This can only be determined by further analysis of the data.

This is one failing, among several others, in the *50-K* fingerprint study<sup>17</sup> (see Wayman, 2000) where it was simply assumed that the minutiae characteristics followed a binomial distribution without verifying that this was the case by analysing the population data. This study was conducted in response to a challenge to the reliability of fingerprint evidence (*United States v Byron Mitchell*, 1999) where the FBI commissioned Lockheed-Martin to prove that fingerprints were unique, and it has been widely cited by the fingerprint community as achieving this aim. It was the first time a large, real-print database was used in order to address the issue of uniqueness in this discipline.<sup>18</sup> The study essentially ‘scored’ the relatedness of each fingerprint with every other print (including itself) using a series of computer programs, and concluded that ‘...the probability of a non-mate rolled fingerprint being identical to any particular fingerprint is less than 1 in  $10^{97}$  (1 followed by 97 zeroes). [T]hus given the population of the earth, the chance of observing any two rolled fingerprints being identical is approximately 59 in  $10^{88}$ .’ (Cited in Kaye, 2003a).

The *50-K* study has been critiqued in some detail by Kaye (2003a). Even the appellate court noted this incredibly large figure with some scepticism, however it was not merely the astronomically large number of  $10^{88}$  that cast doubts on its validity. In the course of the matching process, three unusual scores were obtained. It turns out that these three prints were actually from the same individuals as three other prints in the database, which were inadvertently included in the sample. This presented as three scenarios where the *same finger* was rolled twice and appeared as two separate entries in the database. The scores for these prints were clearly in the range that the researchers considered as indicating that they were from *different* fingers. The logical conclusion to be drawn from this was that the fingerprints of the same finger could be more dissimilar than prints from different fingers – and therefore, one should realistically conclude that fingerprints from different fingers could *not* be unique! This finding rendered the entire study ‘worthless for documenting the

---

<sup>17</sup> This study, credited to Meagher, Budowle and Ziesig became known as the *50-K* Study because it utilised a database of 50,000 fingerprints.

<sup>18</sup> The FBI’s own fingerprint database.

individuality of fingerprints' (Stoney, in Lee and Gaensslen, 2001), however, this did not stop the authors from making such a claim.

Similarly, Christensen (2005) noted that no quantitative testing of the occurrence of frontal sinus features had ever been carried out, and conducted a study involving 808 frontal sinus radiographs. 305 of these were replicates (not simply 'duplicates'), enabling the inter-individual versus intra-individual variation to be compared. Using a method that involved tracing the outlines, scanning the resultant tracings, and converting the resultant closed curve to an ordered set of data points (a technique known as elliptic Fourier analysis), Euclidean distances between pairs of outlines were calculated. The results demonstrate that the mean distance between individuals was significantly larger than that of the same individuals at a p-value of less than 0.0001. However, there were instances where different individuals displayed less Euclidean distances between them than between replicates of the same individual, and vice versa. This is demonstrated graphically in some overlap of the cumulative probability density curves. Concluding that the 'frontal sinus is distinctly quantifiably different at a highly significant level' is reasonable phraseology, and the author does not make any comment on the 'uniqueness' of frontal sinuses, however, the results of this study actually *disproves* any such theory that the frontal sinuses are unique, at least using this method, due to the overlap noted between inter- and intra-individual variability. Bush and colleagues have recently established the same proposition to be true for the dentition in that not only does the same dentition produce variable bitemarks in skin, (Bush et al., 2009, 2010a) but the variation between the bitemarks on human skin from the same dentition can be greater than that seen between bitemarks from different dentitions (Bush et al., 2011).

### ***Distribution Models***

The most mathematically valid studies for characterising the existence of forensic characteristics in the population rely on so-called 'generative' distribution models. In a generative model, the distribution of traits is 'learned' by a computer program from a known data set. Ideally, this should then be verified against a second data set, in order to ensure that the modelling is applicable to other populations. Probabilities can then be calculated using derivations from this (verified) distribution model.

While this is a more sound technique than simply assuming the data fit a Binomial or Poisson distribution, the extrapolation to uniqueness from these results will still involve a 'leap of faith' (Stoney, 1991). As Saks and Koehler (2008) point out, this is because the probability of the same series of markers (1,2...n) appearing more than once in the

population should technically be zero in order for a trait to be considered truly unique. Mathematically,  $P(1,2...n) = 0$  is impossible to achieve using product-rule formulae, as the calculation of such a probability necessarily involves use of multiplication, based on frequencies of the individual characteristics occurring in the population. No number exists that when multiplied by another number is zero, apart from zero itself. The probability of observing such a characteristic either alone or in combination with others is always *greater* than zero, which implies that there is always the probability that a duplicate exists, no matter how small. The ‘leap of faith’ occurs when the practitioner ‘rounds down’ this probability to zero and then claims uniqueness.

Another more pervasive concern with probability models in general is that they generally rely on the assumption that each individual trait is independent of any other, thus allowing use of the product rule to calculate the likelihood that two or more of the features would occur in combination. This is also the subject of much debate. No attempts at quantifying the independence of traits has been introduced for fingerprints, bitemarks, frontal sinuses, or even toolmarks or firearms. Some studies simply assume independence, such as the earlier fingerprint studies, whereas others recognise the existence of the problem, but suggest inadequate or arbitrary ways of dealing with it (See MacFarlane et al., 1974). Even in DNA analysis, it is accepted that complete independence of alleles is unlikely, however, DNA studies have allowed the best-estimation of a constant  $q$  that is incorporated into the calculation of the random match probability for DNA samples<sup>19</sup> in order to compensate for this fact (National Research Council, 1996). While this is not a perfect solution, it goes at least part-way to ensuring a more realistic probability estimate. At this stage, the influence of dependence remains largely unstudied (and thus unquantified) for the dentition, fingerprints and most other forensic markers.

### **Quantifying Uniqueness: The problem of numbers**

Kaye criticises Saks and Koehler for upholding this as an untenable standard that is not followed even in the most rigorous of scientific disciplines (Kaye, 2010), however even arriving at a very *low* probability of encountering a match via this form of statistical modelling does not ‘prove’ anything. So the question arises as to how low should a probability be before we can consider something to be unique? Saks and Koehler describe the common fallacy that people assume when the probability of sharing a characteristic is given as less than 1 in 6.7 billion, the approximate population of the Earth, then this proves

---

<sup>19</sup> This constant  $q$  is generally given as 0.01 for the general population, with figures anywhere from 0.001 to 0.03 used for various population sub-group calculations. This is deliberately conservative – the true value of  $q$  is estimated to be between 0.0006 and 0.001 for any population (Budowle et al., 2001).

that no one else on the Earth could possibly have the same characteristic. In other words, uniqueness is proven when  $\Pr(x|x) \leq 1.49 \times 10^{-10}$ . The fallaciousness of this reasoning is given in the familiar ‘birthday problem’. Using somewhat more fathomable numbers, the birthday problem asks ‘What is the probability that two people in a room share the same birthday, when there are N number of people in the room?’ Assuming that the probability of someone having a birthday on any given day is 1 in 365,<sup>20</sup> then most people are surprised to learn that the probability of encountering someone with the same birthday as them in a room of only 23 people is greater than 50%. For 99 people, the probability exceeds 99%, or 0.99. So even though the number of possible birthdays (365) exceeds the number of people in the room by a factor of three, there is almost a 100% chance that someone in the room will share the same birthday.

Using similar mathematics, if we consider a machine that is capable of printing tickets at random, numbered from 1 to 100, and there are only 10 tickets printed, the probability that two customers will have a ticket that has the same number is actually close to 40%, far greater than intuition would suggest. To assume that uniqueness exists simply because the number of people on Earth exceeds the total number of possible combinations of characteristics is fallacious for the same reasoning. Rawson and colleagues (Rawson et al., 1984), in addition to their poor mathematical treatment of their data, committed this error when they claimed that as few as 5 matching teeth would be needed to be confident of an identification, because the number of possibilities of combinations for five teeth ( $150^5 = 75.9 \times 10^9$ ) exceeded the number of people on earth (at that time  $4 \times 10^9$ ).

Using Bayesian methods, Champod (2000), calculated the probability of a person actually *being guilty* of a crime, given a ‘unique’ trait frequency of 1 in 6 billion. However, using this very low probability of there being an identical match in the entire world, there is still only a 50% probability that this person actually committed the crime. In recognition of this fact, others have argued that it is safe to assume ‘uniqueness’ only when the probability of encountering a duplicate characteristic exceeds the reciprocal of the population by several orders of magnitude (but compare with Balding, 1999; Kaye, 2009; see also Balding, 2005). Philosophically, this makes little sense, as by its own definition, ‘uniqueness’ implies that there will *never* be a repetition of the same arrangement of friction ridge skin, dentition, toolmark patterns and so on. In other words, the possible number of combinations should potentially be infinite.

---

<sup>20</sup> In reality it is different for each day, due to seasonal patterns in reproductive habits, and we discount the leap years.

One should technically consider the total number of people ever to have lived on the Earth,<sup>21</sup> in addition to the number of people yet to be born,<sup>22</sup> in order to assert that a particular characteristic is truly unique (McLachlan, 1995). By implying that these former and yet to be born persons need not be considered in the equation implies that ‘uniqueness’ only holds for a certain population size, leading to the oxymoronic notion of ‘observable uniqueness’. The reasoning that ‘uniqueness is true only for numbers less than  $10^x$ , but not for greater than  $10^y$ ’ makes no sense. Uniqueness cannot be a spectrum, it is a binary concept by its own definition, and therefore must either exist for all numbers less than infinity, or not at all. While Kaye argued that ‘uniqueness’ may indeed be justified when the match probability is infinitesimally smaller than the reciprocal of the population, he is still sceptical of such claims, because as this chapter has demonstrated, it is so difficult to establish that the models used to arrive at such probabilities are realistic or accurate (Kaye, 2010).

Further complicating the search for mathematical ‘proof’ of uniqueness is the fact that the probability of a particular characteristic *existing* in a population (i.e. the figure derived by Kieser-Neilsen, Rawson and others) does not equate to the probability of *seeing such a characteristic again* following its actual observation. It is the latter which is the all-important forensic question. What is important to realise about this semantic difference is that the probability of observing  $x$ , after it has already been seen once, is *always* greater than the probability of seeing it, had it not been seen before (Weir, 1999). This is easily demonstrated:

The probability of not observing a particular trait in a given population can be given by  $P(x = 0) = (1 - p)^n$ , where  $p$  is the probability of any one person having the trait and  $n$  is the size of the population. The probability of observing one instance of the trait in a population is  $P(x = 1) = np(1 - p)^{n-1}$ , and the probability of observing more than one instance is  $P(x \geq 1) = 1 - (1 - p)^n$ . We want to know what the probability of observing the trait is, given we have already found one example of it in the population. This is represented by  $P(x > 1 | x \geq 1)$ . Conditional probability theorem gives us;

$$P(A|B) = \frac{P(A) \cap P(B)}{P(B)}$$

<sup>21</sup> Estimated to be approximately  $1 \times 10^{11}$

<sup>22</sup> An unquantifiable number.

where  $P(A) \cap P(B)$  represents the intersection of events A and B; the probability of observing incidences occurring only *more than* once. As we do not know what this is, it must be calculated by subtracting the probability of observing *just one* instance from the probability of observing *one or more* instances,  $P(x \geq 1) - P(x = 1) = P(x > 1)$ , so we have:

$$P(x > 1 | x \geq 1) = \frac{1 - (1 - p)^n - np(1 - p)^{n-1}}{1 - (1 - p)^n} = P(x|x)$$

Let us assume that the probability of observing a particular characteristic  $p$  is calculated to be one in 12 million, or  $8.33 \times 10^{-6}$ . Let us also assume that the population of interest consists of only 1 million people, and hence the probability of occurrence exceeds the reciprocal of the population by a factor of twelve.<sup>23</sup> What is  $P(x|x)$ ; the likelihood of observing another person with characteristic  $p$  out of this population, given that we know at least one exists? The answer is surprisingly high: 5% or 0.05 – very much larger than one in 12 million.

Of course, this equation would only be valid for features that follow a binomial distribution. A more accurate probability distribution for forensic purposes is the Poisson distribution, often used for modelling systems when the probability events are considered rare (although this is still not necessarily representative, either). Sparing the reader from the mathematical derivation of calculating probability of observing  $x$  given that it has already been seen once, it can easily be verified that for the Poisson distribution, using the same numbers from the example above, we arrive at  $P(x|x) \approx 0.041$ . Again, these numbers are purely arbitrary, and the use of either the binomial or Poisson distributions for forensic traits such as fingerprints may or may not be appropriate, however, the point is made that the probability of observing a characteristic in a population, having already observed one occurrence, is always significantly *greater* than the probability of the trait actually occurring. This is in accordance with the laws of mathematics and yet is counter-intuitive to most people. It is easy to conclude that when the probability of observing a particular characteristic are low, the likelihood that a second person would share that characteristic would be even lower, however to do so falls into an intuitive fallacy not supported by the mathematical analysis.

---

<sup>23</sup> Recall earlier comments regarding those who advocate that uniqueness can be safely assumed when the probability of observing the trait is orders of magnitude greater than the reciprocal of the population. The numbers used here are smaller than the population of the Earth for ease of demonstration, but still fulfil this ‘requirement’.



## The (Il)logic of the Uniqueness Literature

More often than not, the conclusion of uniqueness enjoyed by most studies attempting to prove the notion simply fails to follow from the data that they present. Of course, several studies often cited as ‘proving’ the uniqueness of a forensic marker actually make no such claim. Rather, the proponents of the theory have themselves inferred a uniqueness proposition, based on misinterpretation of the study’s results or conclusions. Pretty and Sweet (2001c) noted that authors frequently cite the Sognaes identical twin study (Sognaes et al., 1982) which concludes that identical twins do not have the same anterior dental arrangements. This may be the case, however, the inference from this study that the dentition is unique may represent the folly of the reader rather than the authors. Sognaes and colleagues themselves do not make such an explicit claim, carefully limiting themselves to the conclusion that ‘even so-called identical twins are not identical’.

Yet consider another study on the quantification of the anterior dentition by MacFarlane, MacDonald and Sutherland (1974), in which the authors concluded that ‘...that even if two individuals had similar tooth status, rotations and arch displacement of teeth, this does not mean that they would produce identical bitemarks. Such factors as the length of incisal edges, exact angulation and spacing of teeth would probably combine to ensure that even apparently similar teeth produce unique bite marks.’ This conclusion represents an unwarranted leap from the authors’ initial aim of quantifying the spatial relationship of the anterior dentition. The authors present no data on the effects of tooth angulation, incisal edge length or spacing on the appearance of a bitemark, as these features ‘require more study to be evident as a distinctive feature of a bite mark’ — yet they conclude that these features would render bitemarks unique. Furthermore, this conclusion is counter-intuitive to the example they give in the introduction of the paper, where they note that an ‘incisor with short length might be misconstrued as a missing tooth when noting the “absence” of a particular feature on a bitemark’. This should suggest that the possibility of more than one dentition ‘fitting’ the bitemark is *more* likely, not less, so that bitemarks are in fact *not* demonstrably unique

Many people assume that twin studies lead us to be able to draw conclusions about uniqueness. This is a misrepresentation of the purpose of twin studies. Twin studies are useful for determining the heritability of traits by using controls where the genetic influence is supposedly the same. Any observable differences between monozygotic twins could therefore reasonably be assumed to be due to environmental factors, for if the trait were solely determined by genetics, the twins should exhibit exactly the same characteristics.

Thus if twins demonstrated identical dentitions, we could say that the spatial arrangement of teeth is entirely dependent on genetic heritability.

The major premise in any twin study, although often not stated, is that '*twins are genetically identical*'. The minor premise that would follow Sognnaes and colleagues' study (Sognnaes et al., 1982), and as the results demonstrate, is; '*twins do not have identical dentitions*'. The only logical conclusion that can be drawn from this is; '*Therefore, the dentition is not entirely dependent on genetic factors*'. A conclusion such as '*Therefore, the dentition is unique*' bears no relation to the premises that go before it, and therefore it fails as a logical argument. According to the laws of logic, there is no syllogism that can lead to the generalisation of all cases from premises based on specific observations. Such reasoning would be purely inductive, and hence we return to the familiar induction problem.

At best, we can only conclude from these studies that people who do not share 100% of their DNA with another person (i.e. non-twins) will not exhibit the same arrangement of the dentition. Unfortunately, the results of Sognnaes detract markedly from even this modest hypothesis. Sognnaes et al proved that twins did not have identical dentitions, and therefore the (unmentioned) inference should be that the spatial arrangement of teeth is not entirely dependent on genetic heritability. The claim that Sognnaes and colleagues offer about their results is that 'comparison of the bitemark pattern of monozygotic twins *could* offer a clue to the degree of uniqueness that can be attributed to an individual dentition'. But despite their optimism, this study gives us no clues about the degree of uniqueness that could be attributed to the dentition, due to the authors' failure to quantify inter-person versus intra-twin variation, which had it been conducted, might have actually provided us with evidence that twins may have *more similar* dentitions than non-twins. This also might have even provided clues as to the relative influence of genetics versus other factors in the position of the teeth. Yet even if this was attempted, it still does not logically lead us to the conclusion that the dentition is in any way 'unique', it simply leads us to a statement regarding the amount of similarity between twins and non-twins.

Jain and colleagues (Jain et al., 2002) conducted a twin study using a generative computer model of fingerprints comparing the index fingers of 100 pairs of identical twins. It is worth noting that there are several robust aspects to this study – the distribution was generated by using a large random data set, which was then compared to a second set in order to test the applicability of the model, and the data set used two prints of the same finger in order to allow for intra versus inter-print variation. Additionally, the threshold for a 'match' was determined via a Receiver Operating Characteristic curve, which has become a widely

accepted decision model for determining threshold points that minimises the risk of false positives while maximising the determination of true positives (Phillips et al., 2001). The study's conclusions also logically followed from its results. While the study concluded that twins did indeed have prints that were not identical, the intra-pair variation was no more than the intra-class variation seen in the general population. In other words, the variation between identical twins was similar to that seen amongst non-twins who had the same Level 1 classification.<sup>24</sup> Consequently, the 'similarity' between twins was because they shared Level 1 features. Despite the fact this was a well-designed study, it proves nothing as far as the 'uniqueness' of fingerprints is concerned, it simply discusses the *similarity* between twin and non-twin's fingerprints. This study was *not about uniqueness*, and just tells us that twins' fingerprints are no more similar or variable to each other than they are to the rest of the population.

The most recent attempt to prove the individuality of fingerprints was a commissioned study submitted to the US Department of Justice in August of 2009 (Srihari, 2009). They firstly conclude that 'the similarity of twin fingerprints is different to that of prints between the same finger.' In other words, *twins do not have identical fingerprints*. The second conclusion is that 'identical twins share the same similarity among their fingerprints as fraternal twins', i.e., there is *no difference between the relative similarity of identical twin fingerprints and the relative similarity of fraternal twins' fingerprints*. The results from comparing prints from the same finger/different print set also suggest that there is a difference in similarity observed when comparing a print image to an identical image, versus a print image to a non-identical image but rolled from the same finger. None of these conclusions are startling, they have essentially already been surmised reasonably convincingly by prior literature, and in fact were already generally agreed upon in the field. The fact remains that just like the Sognnaes and Jain study, neither of these conclusions regarding twins advances the argument for uniqueness.

Firearms examiners claim to rely on the premise that no two gun barrels will leave the same mark on a fired bullet or cartridge case (Bonfanti and De Kinder, 1999) in order to support their identification. This proposition arises from studies of consecutively rifled barrels that demonstrate differences in either the markings transferred to a bullet (Lardizabal, 1995), or on the internal barrel surface itself (Freeman, 1986). While this may be true, the evidence for this relies on relatively few studies with very small sample sizes – in one such study, Hatcher

---

<sup>24</sup> Level 1 features are those used to classify fingerprints, such as loops, whorls, arches, and so on. Level 2 features are the minutiae such as bifurcations and ridge endings, which are usually used to make an identification after the level 1 detail has been assessed. Level 3 features are things like pores and ridge widths, which are rarely used in identification.

compares only two gun barrels before concluding that *all* consecutively rifled barrels are unique (Hatcher et al., 1957). The assumption is that *'barrel markings from consecutively manufactured or rifled guns are more likely to be similar than those from guns that are not'*. This is reasonable, but it relies on the prior assumption that the possibility of two *non-consecutively* manufactured barrels sharing the same pattern is zero. The fact that consecutively rifled or manufactured barrels do not leave identical marks on bullets or cartridges is useful for advancing the argument that 'examiners can distinguish between bullets fired from two guns that were consecutively manufactured', but this does not directly address the concern that another gun could possibly have left the same pattern of marks. This latter concern has not yet been the subject of any large-scale study, and yet is the same problem of the 'random' occurrence of two fingerprints, or dental arrangements, being identical. The fact that two consecutively manufactured barrels do not share the same characteristics does not necessarily address the possibility of a random match, and represents the exact same problem of trying to prove uniqueness for any other forensic marker.

### Study Design

Many of the studies attempting to 'prove' uniqueness are designed, whether intentionally or not, to favour that outcome. Studies such as Reichs' work (Reichs, 1993) on the individuality of the frontal sinuses arrive at astronomical figures simply by using a scoring model that allows a large number of combinations. Consider the alternative approach: one could design a scoring system to describe a particular forensic characteristic as being one of four possible configurations. Performing this analysis at three different locations means that the number of possible combinations is  $4^3 = 64$ . By the same reasoning employed by the studies above, one could conclude that it was highly likely that the characteristic was not unique, as the study 'proves' there are only 64 possible combinations. It is unlikely that this would be taken seriously.<sup>25</sup>

As another example of how studies can be designed to demonstrate particular outcomes, consider the 50-K experiment, supposedly proving the uniqueness of latent prints. This experiment used the now discredited practice of using 'sub-set' inked prints (that is, portions of existing prints were simply cropped to provide 'partials') to simulate latent images (see Sparrow, 1994). No attempt at distorting the image, as might occur in a real latent print was made, and so the fact remains that the second part of the 50-K study continued to compare

---

<sup>25</sup> Indeed, no one who wanted to convincingly prove uniqueness would design such a system, because using 4 features does not give rise to anywhere near 'enough' combinations – hence the rationale for developing more complicated methods of classification in order to ensure the study arrives at a sufficiently large number – an example of 'results by design'.

identical images when comparing ‘matching’ prints. This results in unfairly high ‘match’ scores that favour results toward the conclusion that no fingerprint is the same as any other.

The 50-K study also represents another pervasive problem in the uniqueness literature. Any experiment designed expressly for use in the courts should be carefully scrutinised for evidence of bias, and Epstein reveals some startling revelations regarding the nature of the 50-K study, which was prepared with court presentation in mind. He has stated that as the tests proceeded, Lockheed ‘repeatedly provided the preliminary results to Steven Meagher, the FBI fingerprint examiner who arranged for the testing to be done, in order to make sure that the FBI was pleased with the data being produced and the manner in which it was presented.’ (Epstein, 2001). The 50-K study, which has never been published, is a good example of that prepared expressly for litigation, and previously warned against admission without reservation by court authorities.<sup>26</sup>

## The Irrelevance of Uniqueness

In an interview with one published critic, who wished to remain anonymous, it was stated:

*I’ve written and said that the question of uniqueness is not related to the notion of scientific reliability so many times, that I don’t know whether its intentional or unintentional that fingerprint experts continue to answer “but all fingerprints are unique.” I say “...that isn’t the question I’m asking...”, and they still answer “...but all fingerprints are unique...!” It’s like a broken record. (Anonymous (1), 2009)*

What this author has argued is that when considering the implications of uniqueness to their logical end, it can be seen that the entire notion of whether forensic markers are unique or not is, in fact, irrelevant. This irrelevance relates not only to the fundamental goal of identification, but also to the practice of forensic analysis and the subsequent question posed by the courts.

## The Irrelevance of Studies to Forensic Practice

The very nature of the type of study used to advance the uniqueness argument often renders its results impractical to apply to real-world situations. This favours the outcome for those

---

<sup>26</sup> ‘One very significant fact to be considered is whether the experts are proposing to testify about matters growing naturally and directly out of research they have conducted independent of the litigation, or whether they have developed their opinions expressly for purposes of testifying.’ (*Daubert v Merrell Dow Pharmaceuticals Inc. II*, 1995)

who believe in the uniqueness proposition, but rarely represents a realistic model of how analysis of the forensic trait is carried out. The *50-K* experiment used a model that failed to represent the situation that the fingerprint examiner would encounter – that of comparing two *identical* images to see if they match. Fingerprint examiners already agree that fingerprints taken from the same finger are never visually the same, and hence they would never end up comparing identical images in actual case work, even when encountering the strongest examples of a ‘match’.

The latter half of the most recent effort to demonstrate the uniqueness of fingerprints (Srihari, 2009) focuses on the development of a mathematical model that allows the probability of a ‘random match’ to be calculated. Ignoring for a moment the issue of random sampling and independence, of ultimate concern in this study is that at no stage is the definition of ‘similarity’ given, or whether it is comparable to the level of discrimination achievable by a human examiner. The study demonstrates that a computer is able to discriminate between fingerprints to a level such that the likelihood of encountering a ‘duplicate’ is close to  $1.2 \times 10^{-23}$ , but whether a human being would be able to discriminate between two fingerprints at the same level as suggested in this study depends on whether the level of intra-print variation is sufficiently low such that two prints from the same finger would not be mistaken for two prints from different fingers. Without using data derived from actual human performance to model the computer comparisons, the study has little practical significance.

Returning to odontology, the study by Kieser and colleagues (Kieser et al., 2007), a geometric-morphometric analyses of 82 randomly selected post-orthodontic casts from 50 individuals ranging in age from 17 to 20 years, is also mathematically sound. The study concluded that there were clear differences in the anterior dentition in both shape and form and this was demonstrated by figures depicting the two most similar and the two most dissimilar dentitions in a procrustean superimposition. However, it is unclear whether the superimposition of the two most similar dentitions would appear different without the benefit of the geometric-morphometric analysis, which is almost certainly *never* carried out in bitemark analysis.

Thomas-Johnson and colleagues (Thomas-Johnson et al., 2008, 2009) demonstrated that the probability of encountering a dentition, at least in the Detroit area, with the same thirty-six individual parameters (as measured in their study) approaches the ‘one in a trillion’ category. While this high number renders the possibility of witnessing a duplicate dentition to be relatively low, one of the most recent studies to consider the individuality of the dentition

illustrates very well the overall irrelevance of these low probabilities forensic practice. Miller and colleagues (Miller et al., 2009) demonstrated that none of the ten mandibular dentitions considered physically unique in their study actually translated these features reliably onto cadaver skin such that they could be definitively associated with only one dentition. This raises important questions about any study model that does not consider realistic parameters relevant to the practice of the discipline – that is, the individuality of the dentition does not translate to individuality of bite marks, which are the ultimate issue in bite mark analysis.

### **The Irrelevance of Uniqueness as a Tenet of Identification Theory**

Some forensic practitioners have avoided the uniqueness debate by utilising the term ‘individualistic’ to refer to the property by which each and every person has their own distinct physical characteristics. The process of matching these characteristics to latent marks is thus termed ‘individualisation’, and cleverly avoids the use of terms such as ‘unique’, ‘identify’ and ‘identification’. Kaye has argued that ‘individualisation’ and ‘uniqueness’ are separate phenomenon, and the two terms should not be used interchangeably (Kaye, 2009). ‘Local individualisation’ refers to identification of a mark having come from a source within a defined subset, such as a particular city, town or other geographical area. ‘Universal individualisation’ refers to the identification of a mark having come from a set that includes the entire population of the world. He also defines the term ‘special uniqueness’, implying the quality of object *Y* that has left mark *y* from a set of objects (..., *X*, *Y*, *Z*, ...). He compares this with the term ‘general uniqueness’, which implies special uniqueness for all of the object *Y*’s in the set. This means that there is more than one object *Y*, and each one leaves a similar mark *y*, but each mark is distinguishable from the marks made by the other object *Y*’s. This is akin to what fingerprint examiners and other pattern-matching disciplines claim, and is much more difficult prove.

As yet, no data exists that verifies the general uniqueness proposition, but as Kaye has noted, this is not a necessary condition for individualisation, at least for DNA evidence, because very rarely does the set consist of every human being on the planet. The existence of other evidence helps one to narrow down the set of people that might have committed the crime, and thus the term ‘individualisation’ may be appropriate in the context of a finite set – ie, ‘local individualisation’. The issue then is *not* ‘is this feature unique’ or even ‘are all forensic features unique’, but ‘Is there sufficient evidence to demonstrate that this latent mark originated from this source?’ Alas, this too runs into difficulties when pursuing the implication of local individualisation to its logical end. Forensic practitioners are not generally aware of this other evidence – and therefore have no justification for reducing the

set to a local one. This becomes important when considering the effects of bias and context effects on forensic practitioner decision-making. Assessment of evidence external to that of the simple ‘matching’ of latent marks to forensic characteristics is properly carried out by a judge or jury, and thus individualisation becomes the task of the trier of fact, not the forensic practitioner. Consequently, the expert should not only avoid terms such as ‘unique’, it being an irrelevant and unprovable notion, but also ‘individualisation’, which is an issue solely for the trier of fact (Kaye, 2009). This concept has been supported by others in the forensic identification field (Champod and Evett, 2001; Stoney, 1991).

Uniqueness is invoked as a necessary assumption for the certainty behind forensic conclusions of a ‘match’, but this is somewhat absurd, as the very process by which a match is declared *must* result in some degree of uncertainty as a direct result of the uniqueness proposition itself that ‘no two objects are ever the same’. According to Kwan (1977), ‘qualitative’ identity can be established when a set of properties agrees in two objects, hence they belong to the same set. Therefore, the stronger the match between an object and its impression, the more likely it is derived from that object. This implies that only an exact match between object and impression would definitively indicate that the object left the mark, but this too is demonstrably false. Consider how tools change over time – a tool may not match the mark it made because it might have been subject to modification, via legitimate use or otherwise, but still have made the questioned mark. Conversely, the absence of qualitative identity does not prove that the tool did not make the questioned mark either, for the same reasoning that the tool face may have changed over time.

Requiring an exact match would set an unattainable standard in forensic science, and what Kwan’s example demonstrates is that conclusions regarding identity of source from the relevant ‘agreement’ between source and object properties can only ever lead to a probabilistic conclusions – there is a point at which the number of matching characteristics can only make it *more likely than not* that the tool was the source of the mark. The concept of ‘uniqueness’, whether it exists or not, has no bearing on the fact that absolute certainty of identification is unattainable in pattern-matching forensic analysis, and hence cannot be the basis for individualisation. The uniqueness proposition is simply a justification to ‘round up’ from an unspecified level of uncertainty to being 100% certain. Cole (2009), via a different epistemic path, considered pursuit of the implications of uniqueness to its logical end, has also reached a similar conclusion that uniqueness is *not* the basis for individualisation.



### The Irrelevance of Uniqueness to the Courts

Cole has argued that the condition of uniqueness also bears no relationship to the fundamental questions posed by the legal system (Cole, 2009). He has argued that the question of whether a particular forensic assay is accurate is far more important than that of whether a particular trait might be considered unique. If we consider the random match probability of a bitemark to a dentition to be  $1.0 \times 10^{-12}$ , under several author's propositions, we may be justified in deciding that this particular dentition is unique. However, it also becomes apparent that the error rate for matching bitemarks, due to observer effects, poor quality of the mark, and so on, might be 0.1%. In other words, this (very conservative) error rate assumes that the result is wrong about 1 out of every 1000 times.<sup>27</sup> The issue of 'uniqueness' in the context of a remote possibility of a match becomes irrelevant, because even if the person's fingerprint *is* unique, the probability that the practitioner has falsely matched the print is a billion times larger, and thus has far more relevance to the weight that should be given to this evidence.

Some judges and lawyers have realised that uniqueness, or even proof of 'high degrees of variability' take a sideline to more important issues that result from an 'identification' of a suspect by matching with a latent mark. In the 2000 *Carroll* trial, defence counsel pursued this line of reasoning when questioning the witness about his attempts to demonstrate this huge degree of variability, as described earlier in this chapter;

[Counsel] What that means though, at the end of the day, is just that simply none of those 25 people matched up with that bitemark? — [Witness] Absolutely.

So in the context of this it proves absolutely nothing? — If we'd taken 10,000 people and run the same experiment we would say exactly the same. It proves nothing, and it proves that 10,000 people didn't match the bite.

So why do it? — We did it because we wanted to demonstrate the fact that if we took a group of random casts that we would not expect to find concordance. [...] It adds weight to the conclusion that with the degree of concordance that we've been able to find here, there is a real meaning that can be interpreted from that.

I suggest to you that it does nothing of the kind, because that still leaves open the subjective opinion that the amount of concordance that you've got is – what you say you've got in this photographs [sic] – is capable of identifying the suspect? — That particular determination [...] is always going to be a matter of opinion. (Trial transcript, p622)

---

<sup>27</sup> It is, of course, likely to be much higher, and this demonstrates another key point, in that we don't even know what it *might* be, because virtually no good-quality studies have been published that attempt to quantify examiner error rate for bitemarks.

Judges in the United States have also followed this tack when deciding to exclude or limit forensic evidence. Gertner J noted in *US v Green* (2005) that ‘...even assuming that some of these marks *are* unique to the gun in question, the issue is their significance, how the examiner can distinguish one from another, which to discount and which to focus on, how qualified he is to do so, and how reliable his examination is.’<sup>28</sup> The idea of a gun leaving marks on a bullet unique to that particular firearm was noted, but dismissed as avoiding the fundamental issues. From the prospect of ensuring admission of forensic testimony in court, studies proving the uniqueness of any particular characteristic are unlikely to prove helpful.

## Conclusion

Reliance on cultural memes fails as proof of the uniqueness tenet in any sense. As Thornton wisely concluded in his original article exploring the possibility of two snowflakes the same existing; ‘unlike snowflakes, gun barrels are not made in clouds, and the proof of uniqueness of other objects must be based on yet other grounds’ (Thornton, 1986). Philosophers such as McLachlan specifically avoid arguing for the falsity of the proposition of uniqueness itself, instead restricting themselves to the argument that the theory, even if it is true, can never be known to be such, and that there is no such reason for believing such a theory, as one could never exist. The dictum of philosophy tells us that uniqueness is not an establishable proposition, and belief in such represents ‘a highly dubious snatch of metaphysics...a philosophical assumption rather than what it is presented as being – a hard headed scientific conclusion’ (McLachlan, 1995).

Nonetheless, forensic practitioners have tried to prove that various features are unique. Studies attempting provide convincing evidence that any feature or combination of features is unique face insurmountable issues, most of which cannot be overcome using modern-day experimental techniques. Stoney recognised nearly twenty years ago that attempting to ‘prove’ uniqueness using statistics was a ‘ridiculous notion’ (Stoney, 1991), yet research continues to this day. For those that have attempted to prove uniqueness, analysis of their methodology fails to allow much faith in their conclusions. Even considering these efforts in combination, the evidence for the individuality of any characteristic remains minimal at best when considering their collective flaws.

These issues become secondary with the realisation that ‘uniqueness’ is not relevant to either the theory or practice of forensic identification, nor is it relevant to the fundamental forensic

---

<sup>28</sup> Emphasis added.

and legal questions asked by the courts. Identification, or ‘individualisation’ if one prefers, is the task of the judge or jury and not the odontologist. Directing odontology research resources towards proving uniqueness is therefore pointless, and diverts resources away from more useful projects (Cole, 2009). To attempt to prove uniqueness also puts the cart before the horse when more basic issues such as the meaning of the term *match*, the quantification of level of certainty and the standard of practitioner performance – all of which will influence how we perceive uniqueness – are yet to be agreed upon. Kwan, who undertook a purely logical analysis of identification in his 1977 thesis ‘Inference of Identity of Source’ realised that there can never be a certain conclusion regarding identity from latent marks, but there exists a threshold, or series of thresholds, at which point the likelihood of identity is proportionately higher or lower. Determination of these thresholds becomes a key issue, and these need to be determined via a series of reasoned, valid experiments. Additionally, the effect of phenomena such as bias, error, variation in examination technique, standards and other considerations will also influence the relative certainty of a conclusion, and these too should be described and quantified in order to provide the trier of fact with a reasonable estimation of the reliability of any forensic analysis. Some aspects of these phenomena will be explored in later chapters.

Uniqueness of the dentition simply doesn’t matter all that much; mistakes and misidentifications are not made because someone has an identical dentition to someone else in the world. They are made because of guesswork, poor performance, lack of standards, bias and observer error to name but a few examples. While many forensic disciplines rely on such a notion as ‘proof’ of their infallibility, the question of ‘uniqueness’ is an entirely separate issue to one of ‘reliability’. There remain few valid reasons to continue this fruitless search for what remains a philosophical ideal. There is no valid reason to quote ‘uniqueness’ as a tenet of bite mark theory, and as nothing more than idealism, it consequently fails to advance the reliability of the discipline to any extent.

## CHAPTER 6

## A Critical and Contextual Evaluation of Bitemark Analysis

---

Having established that proof of one of the central tenets of forensic identification, that of uniqueness, actually bears little relation to the overall validity or reliability for not only forensic odontology, but for any forensic science discipline, we now turn to a review of other scientific literature regarding the reliability of bitemark analysis. While courts in Australia appear to principally concern themselves with qualifications, experience, and education, these are recognised by most to be crude indicators of the true validity of an expert's testimony. Qualifications and education vary greatly in their content, length of course, relevance, and standing in any particular scientific community, as well as more generally. Experience is difficult to parse in any detail and can easily be over-rated by the trier of fact, as very rarely is any effort made to distinguish between 'ten years of experience', or 'one month of experience, repeated one hundred times', or even 'one month of experience, spread over ten years' (Faigman et al., 2007). Statements related to relevant experience, such as 'I have never seen another similar instance in my 26 years as a practitioner' are difficult for counsel to follow on from during examination or cross-examination, as is then eliciting the extent and quality of that experience (Thornton et al., 2007).

Witnesses who testify using such justifications for their opinion remove themselves from the scientific method, and science in general, despite the notion that forensic practitioners are 'obliged to follow the scientific method in performing examinations and formulating conclusions' (Klinkner, 2009; Peterson and Murdock, 1989). Saks has argued that expert evidence that is devoid of science has no place in the courtroom (Saks, 2002), and represents bare opinion, *ipse dixit*. Such evidence fails to provide the trier of fact with an appropriate means of assessing the expert's testimony, and therefore provides them with no means to assign the appropriate weight to such evidence.

Critics of forensic science, such as Professor David Faigman, have chastised the forensic identification sciences, inclusive of fingerprints, firearms, toolmarks and bitemarks, as 'failed' sciences with little or no research basis, relying largely upon the subjective judgement of practitioners (Faigman, 2007). Labelling these and several other disciplines as

‘anecdotal sciences’, he has stated that these disciplines contain a substantive degree of subjective judgement in their application, do not test their hypotheses in any serious manner, and only look for confirmation of their practice while rationalising, or ignoring, contradictory evidence. Faigman draws a parallel between forensic science and both phrenology and blood-letting, two areas of science where practices continued for decades despite no evidence they ever achieved what they claimed they could, modelled on practices of simply ‘what had gone before’.

Saks and Risinger (2003) noted that the forensic identification sciences differ in what most consider ‘science’ to be, with these disciplines having never been subjected to any formal validity testing. They have claimed that there are numerous problems with the research that has already been undertaken by forensic practitioners, including by having placed propositions beyond empirical reach by relying on subjective means of measurement; using research designs that cannot generate clear data on individual practitioner competence; having manufactured favourable test results by giving poor practitioner performances additional help or resources to reconsider their judgements; refusing to share data with researchers who wish to re-analyse or further analyse results; encouraging over-stated interpretations of data; burying unfavourable results in reports where they are least likely to be noticed; and generating disclaimers that the data cannot be used to infer the false positive error rate it actually reveals.

In another publication, Saks (2002) wryly noted that very *absence* of the type of data needed to evaluate criminal case expert testimony seemed to be the rule, rather than the exception. Faigman made an interesting analogy between the ‘forensic science’ approach to gathering data and that of real-estate appraisers when he related an elucidating anecdote. He once asked a real estate appraiser about how he arrives at a particular monetary value for a property, to which the answer was; “Well, we have comparisons in the neighbourhood, and we have this methodology...” which was then explained. Faigman then asked; “Do you go back, after you have sold the house, and you have some data, to see if you are right?” to which the appraisers reaction was one of bewilderment: “Why would I do that? I know I’ve got it right, because ... well, we have comparisons in the neighbourhood, and this methodology...” (Faigman, 2009). This example illustrates a blind faith in the hypothesis, methodology and conclusions without actually conducting any formal testing that both real estate appraisers, and by way of comparison, forensic scientists seem to rely on. In doing so, both disciplines seem to ignore the more important question of whether the ‘standard’ theory or ‘accepted’ methodology is actually yielding the *correct* result.

It has been said that forensic scientists in general have failed to consistently appreciate the implications of the scientific method (Thornton et al., 2007). The scientific method, generally described as progressing through the stages of research, formulation of a hypothesis, testing, analysing results, and then modifying the hypothesis where necessary, has the advantages of in-built evaluators, such as the calculation of error rates, as part of this process, in addition to other benefits, including its impartiality (See Gauch, 2003). While academics note a tension between ‘science’ and ‘forensic science’, there are some who disagree that the traditional ‘scientific method’ is appropriate for forensic science to follow, due to its unique position straddling the disciplines of both science and law. These contenders exploit this relationship, and note that forensic science is required to engage in the kind of scientific inquiry that cannot be modelled after the traditional natural sciences (Pyrek, 2007). Nordby wrote (in James and Nordby, 2009) that forensic science operates outside the carefully controlled environment that the traditional sciences endure, and thus ‘the law-covering model of natural science accounts for expectations of scientific certainty, which no forensic science allegedly approximates’.

While it is true that the conditions appropriate for using traditional scientific models of research<sup>1</sup> are not applicable to the forensic sciences, conditions and assumptions appropriate for experiments in physics are not appropriate for experiments in chemistry either, and vice versa. Each of these disciplines has conducted independent research in order to ascertain the boundaries of external influences, which enabled scientists to then measure and subsequently account for them. Claiming that forensic science does not enjoy the pristine conditions of experimental science, and thus is not directly comparable, avoids a reality that has plagued all scientific research — a reality that has already been managed by experimental scientists through careful and deliberate hypothesis testing.

Nordby (in James and Nordby, 2009) noted that ‘in the forensic sciences, we reason from a set of given results (a crime scene, for example) to their probable explanations (hopefully, a link to the perpetrator).’ While the forensic scientist makes deductions from given data in order to draw conclusions, these deductions often have failed to subsequently undergo the rigorous testing, re-formulation and refining that those associated with traditional science would. There is a tendency to accept the theory without questioning its underlying basis, simply because it appears to fit one, or perhaps several, examples. Nordby acknowledges that the forensic sciences are never one hundred percent foolproof, and briefly discusses the American scientist and philosopher Charles Sanders Pierce’s notion of ‘contrite fallibilism’ as a cornerstone of his own personal philosophy. While such notions are admirable, they

---

<sup>1</sup> For example those used in chemistry, biology or physics.

contribute nothing to the underlying scientific knowledge base. Having an awareness of ‘how much [one] does not know, and having the humility to acknowledge the possibility of making mistakes’ is nowhere near as useful to either the legal or the forensic community as quantifying just how far wrong one might be.

Most critics simply claim they ask for nothing more than ‘evidence-based forensics’ (Edmond, 2008b; Cole, 2007). The concept of evidence-based practice has been long accepted in other scientific disciplines, such as medicine. While evidence-based medicine is not without controversy, it has been enormously influential in health care, and it has proven a powerful metaphor suitable for importation into other areas of technical decision-making (Cole, 2007). Evidence based practice is typified by repeated, rigorous data gathering, rather than relying on rules, singular observations, or custom. This model has led to the development of systems to stratify evidence, and there are several systems to choose from, with subtleties dependent on the nature of the empirical question. Essentially, they all classify evidence on a graded scale from high to low value, with anywhere from three to five main levels. What is of interest is that the lowest level of evidence in virtually all of the classification systems is that of ‘expert opinion, without specific critical appraisal, or based on bench research’ (Oxford Centre for Evidence-based Medicine, 2009). The highest level of evidence (almost always Level 1, or A, in such classification systems) is generally that obtained by a systematic review of existing studies that are based on randomized controlled trials and studies validated in different populations.

A key feature of these high-level studies is that they involve a control group as well as an experimental group. Just because a study is designated ‘randomised and controlled’ does not mean that it automatically gains status in the level-of-evidence hierarchy. Studies that fail to clearly define comparison groups, fail to measure outcomes in the same objective way, fail to identify confounders, or fail to test ‘blind’ fall into the poor-quality category, and consequently, the relative level of evidence assigned to the inferences drawn from these results falls. The evidence-based model has a systemized approach to look beyond the numerical results of the study in order to ascertain the true meaning of its conclusions.

This practice has been notably absent from many forensic science studies. Even considering the few empirical studies carried out in light of the ‘evidence-based’ model of research now well recognised in medicine and science, few of them meet the standards expected for high-level evidence, where randomisation, good controls, blinding, narrow confidence intervals and consistent reference standards are considered the norm. There have been few ‘blinded’ performance studies conducted to date in odontology, and confidence intervals are generally

wide in odontology studies due to the small sample size.<sup>2</sup>

A number of larger-scale studies have been conducted using judicial casework as validation for a particular forensic technique, but these subsequently fail to ensure independent verification of the true source of any particular mark (Cole, 2005, 2007). As such, these implicit tests invariably fall into a very low-level category of evidence due to the non-independent nature of the reference standard.<sup>3</sup> Many studies that have attempted to bolster the argument for forensic examiner accuracy, such as those run by the Collaborative Testing Service and the American Society of Crime Laboratory Directors, have been criticised as having many of the features of poor-quality studies, including no control over the conduction or timing of the tests, uncertain sample sizes, and varying difficulty levels (Cole, 2007). Haber and Haber (2003) also add that these tests often only assess a small proportion of the examiners typical job duties, and thus do not represent accurate descriptions of examiner proficiency in court. They describe one organisation's test procedures as being 'un-interpretable' due to numerous flaws in the design of the assessment.

Critics of forensic science, including the courts, have noted a somewhat lesser standard in the quality of forensic science literature compared to mainstream science. One judge commented that '...the court found...these articles [published in several forensic journals] to be significantly different from scholarly articles in such fields as medicine or physics, in their lack of critical scholarship.' (*United States v Starzecpyzel*, 1995; see also Giannelli, 2009). There is no question that there is an established body of literature concerning the theory and application of forensic odontology. Publications such as the Journal of Forensic Science, Forensic Science International and the Journal of Forensic Odontostomatology have published articles on bitemark analysis for many years. However, not all of this literature should be judged equal. Pretty and Sweet (2001c) reviewed all of the English-language articles concerning bitemarks from 1960 to 1999 retrieved by a MedLine search. From the 120 papers considered, case reports accounted for 28% of these, while empirical research accounted for only 15%. Review articles accounted for a *greater proportion* of the literature than the number of research articles published in the same period. Only 44% of all the literature published had been cited since publication, with most citation accounted for in the aforementioned review articles. It would not be unreasonable to conclude that more than half of the articles published regarding bitemarks are never considered seriously by other

---

<sup>2</sup> The largest experiment to date that has been reported in the literature regarding practitioner performance was conducted by the ABFO in 1999 (Bowers, 2007a; see also Arheart and Pretty, 2001), and consisted of only 32 participants.

<sup>3</sup> Just because the results of a judicial enquiry have been derived from an adversarial process still doesn't mean that the conclusion – that of guilt or innocence – is actually the correct one.



researchers in the field. This raises serious questions regarding the type and quality of the literature being produced in this discipline.

While many critics publish in high-ranking journals, many comprehensive responses issued by members of the forensic community either fail to be published in journals of equal status, or fail to categorically disprove such criticisms. By way of illustration, consider the response to Saks and Koehler's now infamous article, *The Coming Paradigm Shift in Forensic Identification Science* (published in *Science*, currently the third most cited journal in the world).<sup>4</sup> As of June 2010, there were 23 published citations of this article. 21 of the articles that cited *The Coming Paradigm Shift* were published in peer-reviewed [scientific or law] journals, all of them supportive of Saks and Koehler's original text. Of the remaining two, both of which rebutted Saks and Koehler's claims that much of forensic science was in need of a proper research foundation, one appeared in the *Crime Lab Report*, the 'internet publication' of a self-described 'independent organization that analyses media coverage, industry trends, and public-policies related to forensic science laboratory testing and its application within the criminal justice system' (Crime Lab Report, 2008) and the other appeared in the *News of the California Association of Criminalists*. Another four brief replies to Saks and Koehler's article were published in a forthcoming issue of *Science*, in the Letters section (Harmon et al., 2006), however, from a total of twelve references that were cited between them in support of their rebuttals, four of these references were by the same author, and three others were actually US Government sponsored reports.

Edmond (2008b) makes a number of valid points about the philosophy of journal publication and the peer-review process, concluding that these and other 'ingredients of popular representations of science and medicine...are unlikely to provide the kinds of discriminating criteria that might guide meaningful assessments of expert evidence'. This may be true in terms of usefulness for a lay juror or judge, however, this is not necessarily true for the scientists and academics, who form the majority opinion in criticising forensic science — it is a quintessential part of scientific culture that hypotheses, theories and methods pass through the filter of the general scientific community via publication in scientific journals. If the forensic science community intends to address such criticisms, it needs far more scientifically valid, research-based evidence than currently exists in order to counter the growing pile of irrefutable scientific literature exposing its flaws.

---

<sup>4</sup> Source: ISI Journal Citation Reports 2008-2009 (Science Citation Index) (accessed 30 Sep 2009 2008).

## **Literature Concerning the Collection and Analysis of Bitemark Evidence**

Collection and analysis of bitemark evidence has been the subject of many case report articles, with reports of techniques including impression taking and model fabrication of the bitemark in plaster of Paris (Benson et al., 1988), silicone, or polyvinylsiloxane (Blackwell et al., 2007), photographing of the bitemark in situ, and excision of the portion of flesh containing the mark (Sweet and Bastien, 1991). Methods of examination and analysis have also included radiographic assessment of both the bitemark itself (Graham, 1973) and the impression of the suspects teeth in wax. A solarisation or equidensity printing technique was advocated by Rawson (1976), and other authors have recommended using infrared lighting (Gustafson, 1966), ultraviolet lighting (Cameron et al., 1973), narrow band lighting (West et al., 1992) and ordinary lighting (transillumination) (Dorion, 1987) in order to gain maximum information from the bitemark. Rao and Souviron (1984) advocated dusting and lifting the bitemark using fingerprint powder. West and Frair (1989) described a method of videotape analysis, and others have more recently described various 3-D laser scanning and computer shape-analysis techniques (Lasser et al., 2009; Martin-de-las-Heras and Tafur, 2009; Nambiar et al., 1995).

The current ABFO Bitemark Guidelines (American Board of Forensic Odontology, 2009) endorse many of these methods (see Figure 6.1), despite later literature suggesting that many of them are highly unreliable. Very few of these methods have anything other than one or two published articles suggesting their use, with many of these doing so on the basis of a single application of the technique in a case study. Determining a standard method, based on research that demonstrates its use involves the lowest possible accuracy rate, is an essential prerequisite for the generation of any kind of data that might indicate bitemark analysis' usefulness (see Beecher-Monas, 2008).

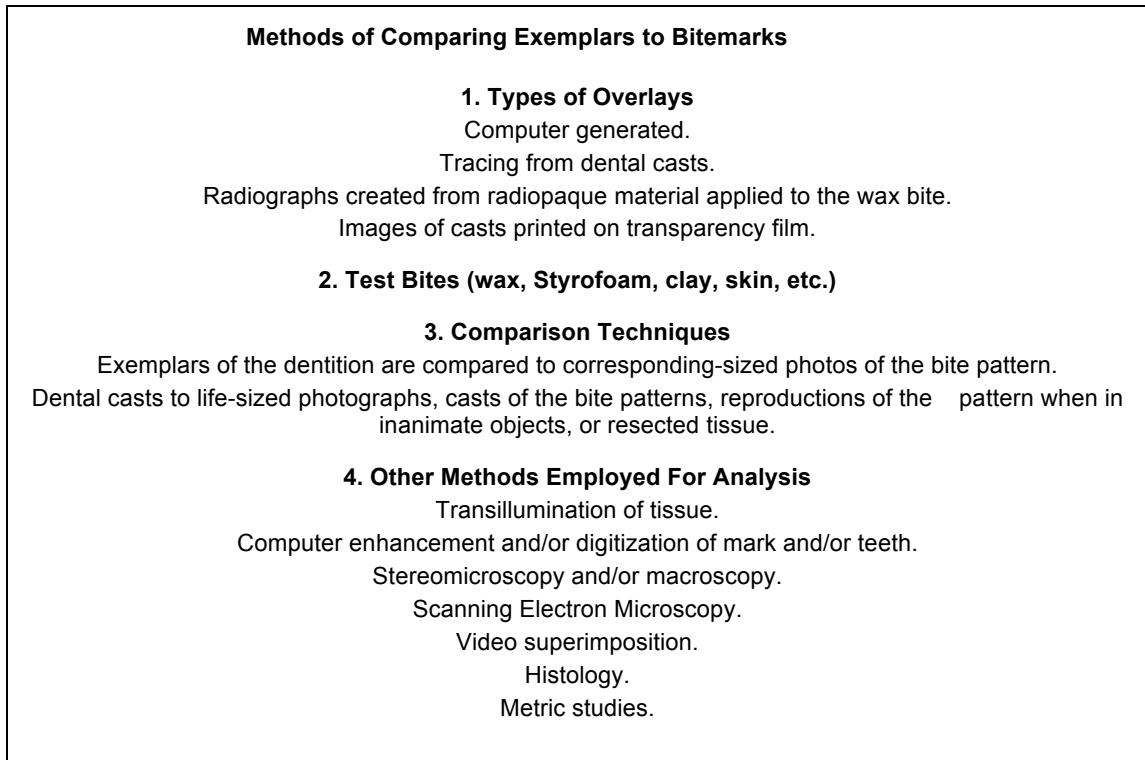


Figure 6.1 - ABFO Endorsed Methods of Bite Mark Analysis<sup>5</sup>

This wide variation in potential methodologies employed by the odontologist has a detrimental effect on the ability of researchers to determine an overall accuracy rate of bitemark to dentition comparisons. In 1971, DeVore (1971) concluded that there was a large margin of error in using bite mark photographs and unsecured excised skin as a methodology of comparing marks to a suspect's dentition. He concluded that bitemarks were 'useless' for positive identification purposes unless the exact position and condition of the body was known at the time of biting.

Whitaker (1975) simultaneously assesses the accuracy of model photograph and plaster model comparison with bitemark photographs, wax bites, and bitemarks in pig skin. The highest degree of accuracy was obtained when the two examiners compared impressions in wax with models of the subjects, where 98.8% of impressions could be accurately matched. Approximately the same degree was obtained when matching stone models of the wax bites against the original study models. When photographs were compared with photographs of the study models, an accuracy of 96% was achieved providing the examiners used a quantitative method of analysis, involving the measurement of arch curvature, tooth widths, tooth angulations and spacing. When no measurements were taken and a purely subjective approach was used, the accuracy fell to 68%. Furthermore, after using photographs of the

<sup>5</sup> From the ABFO Diplomate Reference Manual (American Board of Forensic Odontology, 2009)

bitemark taken one hour and twenty-four hours after it was made, the accuracy fell to 35% and 16% respectively, even when measurements were used.<sup>6</sup>

Recent literature has supported the use of digitally produced overlays of the suspects dentition over hand-drawn (from either wax impressions or models), photocopied, or radiographic outline methods. Despite the plethora of 'recommended' techniques, the most common comparison method appears to be using overlays, either hand-drawn from a wax bite, model or a photocopy of models, or a digitally produced overlay from a scanned model (Bowers and Johansen, 2003; Martin-de las Heras et al., 2007; Pretty and Sweet, 2001a). These overlays are then directly compared with the bitemark itself, by 'overlaying' them on the bitemark (either a scaled digital image or a life-sized photograph), and attempting to line up the digitally-produced outline of the dentition with the features present in the bitemark. In a web-based survey of 72 odontologists, 27 of whom were ABFO Diplomates, Pretty (2003) established that 63% of respondents use some form of overlay method for bitemark comparison, with 34% of these using a non-digital method for its production, such as hand-tracing, photocopying or radiographic techniques.

According to a study by McNamee and colleagues (McNamee et al., 2005), digitally-generated overlays still provide at best moderate inter and intra-examiner reliability results regarding incisal area and subsequent tooth shape measurements. In their assessment, the reliability scores for incisal area representation by digitally produced overlays made by practitioners using two different digital production methods ranged from 0.260 to 0.651, where a score of 1 indicates excellent reliability, and 0 indicates poor reliability. Assessment of tooth position via the {x,y} centroid coordinates gave far better results, with the mean reliability scores for the centroid measurements calculated to be 0.947 and 0.997 for one digital technique, and 0.964 and 0.993 for the other. These results suggest that the overall shape of the entire *arch form* is satisfactorily reproducible in both an inter- and intra-examiner sense using digital methods, but the relative dimensions of individual tooth outlines are not.

One point worth noting with regard to the design of this study is that the intra-examiner reliability was assessed by providing duplicate casts for comparison amongst the set of casts given for overlay production. The authors note that the occlusal surfaces of the posterior teeth were altered, in order to 'mislead the examiners into thinking the casts were different', however, this clearly did not work, as some examiners commented that they had noticed a

---

<sup>6</sup> This suggests that retention of the three-dimensional data obtained via impression of the dentition in the medium is likely to allow significantly more accurate comparisons.

few cast images were duplicated. The extent to which examiners compared and then altered their overlays when this was recognised cannot be ascertained as the participants were not supervised. Additionally, as all of the overlays were produced over a relatively short period of time, the effect of familiarity with the technique, allowing greater confidence and ability to be imparted to those overlays generated after the initial few, must be considered. The overall performance of practitioners may also depend on both the order of the assessment of the cast images and whether the division of labour over one or several time periods. These factors lead one to suspect that the reliability scores are probably weighted towards being more accurate than they really are.

Sweet and Bowers (1998) established that computer based methods of overlay production were superior to other overlay production methods in terms of accurate reproduction of incisal area and tooth rotation, and this study has been cited by others as affirmation of the digitally-produced overlay technique as the most reliable (Kouble and Craig, 2004). Sweet and Bowers conclude that 'the computer-based method produced digital images of dental study casts ... were found to be very accurate', but this needs to be read in the context of comparison to other methods of overlay production. In their study, the digital overlays were considered the 'gold standard', and the other four methods compared to the digital overlay itself, and not the models of the suspect's teeth. Thus, technically, it is a comparison of only the latter four methods, as it appears that the digital overlay was taken as the reference standard, and so naturally has the highest accuracy score when compared with itself. The radiographic method appears to be the next best technique for determining area, and the photocopy technique the next best for determining tooth rotation. Hand-drawn overlays from wax bites were third in both categories, and hand-drawn overlays from models were fourth in both categories. Unfortunately, the radiographic method was also the poorest in representing tooth rotation, and the photocopy method the poorest in representing area, thus indicating that neither of these two techniques are particularly desirable as analytical methods compared to the digital or hand-drawn overlay method. The study did not compare the digital method with any raw data, such as that obtained from the actual casts, and thus it cannot attest to the overall accuracy of the method as a whole.

At least one study has demonstrated that repeat measurement errors tend to be greater for the dentition of scanned casts compared to digitisations of a wax bite (Bush et al., 2011). This suggests that the digital overlay might be better produced from a wax bite rather than a scanned model. There is a comparison here with the practice of fingerprint and toolmark examination, where in both of these pattern-matching disciplines, matching is carried out comparing latent mark with latent mark, and not latent mark with source. Suspect tools are

applied to surfaces and the resultant marks are compared with the latent mark. Ballistic evidence is compared in a similar way – bullets are fired from the suspect gun and they are compared with the suspect bullets. The gun itself is generally not examined internally in an effort to make the rifling and barrel patterns ‘fit’ the suspect bullet. Johansen and Bowers briefly discuss the advantages of making a digital overlay from a wax bite of the model, as compared to a scanned image of the cast in their original publication (Bowers and Johansen, 2003), and note that the philosophy of letting the ‘teeth do the talking’; by using the actual marks made by the teeth to fabricate the overlay is a good one. On the other hand, they suggest that where a good quality bitemark is apparent that demonstrates individual characteristics of certain teeth, fabrication of an overlay from a direct scan of the dentition (allowing a ‘compound overlay’, which demonstrates features of the dentition within each hollow-form outline of the teeth) is the better method. This is, however, based on their opinion and not on any research into qualifying how useful such a compound overlay may be.

### **Metric analysis**

Odontology texts often refer to the advantages of using metric analysis in order to analyse potential matches between bitemarks and dentitions (Bowers and Johansen, 2003; Dorion, 2005; Stimson and Mertz, 1997). However, emerging research suggests that metric analysis has only limited use in correlating dentitions to bitemarks. Early research suggested that bitemarks in skin are likely to be proportionally larger than the respective dentition rather than smaller (Jakobsen and Keiser-Nielsen, 1981), and although the research basis for this is tenuous, it does make practical sense. A classification of the types of distortion encountered in bitemark analysis was offered by Sheasby and MacDonald (2001). They introduce the terms primary and secondary distortion, however do not attempt to quantify any of these. They conclude that the degree of distortion present in a bite mark is variable and affects arch size and shape, and that ‘an exact match in arch size is fortuitous and unpredictable, questioning the significance of this criterion’. Distortion of bitemarks on skin is likely to vary considerably depending on factors such as the nature of the bite itself, the time elapsed since the bitemark was made, the elastic properties of the skin in various areas of the body, the posture of the body part during the bite, and distortion introduced during the photographing and recording process.

More recently, experimentally derived results have shown that distortion in both negative *and* positive dimensions occurs with bitemarks (Bush et al., 2010a). The tolerance of these measurements of distortion needs to be properly determined using a larger dataset of knowingly matched bitemarks and dentitions, however, it has been demonstrated that

distortion of up to 3.5 mm when measuring inter-canine width is not unusual in both the negative and positive directions. Similarly, mesio-distal width of teeth has been demonstrated to vary by as much as 30% of the actual incisal edge measurement, and rotation of teeth by up to 80% of the individual angulations. Further complicating these findings is that these measurements vary between individual teeth in the *same* dentition.

### **3-dimensional analysis**

Another technique for bitemark evidence collection and analysis is the taking of impressions of the suspected bitemark. Numerous techniques have been described for analysing the three-dimensional properties of bitemarks (Blackwell et al., 2007; Flora et al., 2009; Martin-de-las-Heras and Tafur, 2009), yet there seems to be little use of these techniques in actual casework. While texts recommend the taking of impressions of bitemarks in order to preserve evidence, there is little elucidation on what to do with the resultant impressions and/or models of the resulting bitemark. Clearly three-dimensional information is more likely to yield more convincing information about the relative ‘match’ between a suspect dentition and a latent mark, however, exactly how this analysis is to proceed remains somewhat vague in the literature.

## **Interpretation of a Suspected Bitemark Injury**

Most forensic medicine textbooks warn of the dangers of interpretation of cutaneous wound patterns (see Gall and Payne-James, 2011). Reports in journals have demonstrated that marks from teeth can be easily confused with injuries from other objects such as bottle-tops (James and Cirillo, 2004); glass bottles and defibrillators (Clement, 2011); a variety of other medical equipment (Dorion, 2005); pathological lesions such as ringworm, lupus, pityriasis rosea and other dermatoses (Gold et al., 1989; Riviello, 2010); not to mention a myriad of other inanimate objects such as the heel of a shoe, childrens’ toys and jewellery (Clark, 1992). Generic bruising has been demonstrated to sometimes resolve in a ‘ring’ formation, where the centre of the bruised area heals first, leaving a ring-shaped pattern of bruising with the passage of time (Hunt, 2007). Objects that bear little resemblance to the shape of the healing bruise may in fact cause injuries that could be easily mistaken for a bitemark.

An attempt to establish the basis for identifying ‘abusive’ bites on children revealed that there is little literature on the subject of identifying potential bitemarks. A review of 149 articles on the subject of bitemarks in children established that the literature regarding the identification of injuries as bitemarks lacked case-control studies and involved relatively small numbers of subjects (Welsh Child Protection Systematic Review Group, 2008).

Furthermore, the physical evidence examined in these studies was generally poor, with all except five articles lacking any confirmation that the injuries considered indeed were (or were not) bites. The majority of the literature resorted to limiting the description of bitemarks to very general terms; ‘a 2–5 cm oval or circular mark, made by two opposing concave arcs, with or without associated ecchymosis’. Importantly, the authors note that there were no studies validating the ‘much-used BAFO guidelines for bitemark analysis’ (Kemp et al., 2006).

While experienced odontologists may feel they can determine what is and what is not a bitemark, there is no supporting literature regarding this subject. Arch width or inter-canine distance is commonly cited as a means of determining whether the biter was an adult or child (Dorion, 2005; Levine, 1984) however orthodontic studies confirm that there is considerable variation (and overlap) in these measurements for any given population that is dependant on factors such as race, development, sex and general individual variation (Acharya and Mainali, 2009; Barsley and Lancaster, 1987; Bishara et al., 1997; Radmer and Thomas Johnson, 2009)

In the only study of its kind, Whittaker and colleagues (Whittaker et al., 1998) established that there was no significant difference in the ability of senior or junior forensic odontologists to distinguish between an adult or child’s bitemark (using case-work photographs), although both forensic odontology groups out-performed police, social-workers, and general dentists, with AUC<sup>7</sup> of 0.693. These results suggest that training is more important than experience when it comes to determining the origin of bitemarks, yet the more fundamental question of whether odontologists can accurately deduce whether an injury is indeed a bitemark or not remains unanswered.

Gall and Payne-James note that the ageing of injuries is a particularly contentious practice that is open to a significant range of opinion (Gall and Payne-James, 2011). Bitemark injuries typically manifest as bruises, or cutaneous surface wounds ranging from ill-defined abrasions to more discrete lacerations associated with the incisal edges of teeth. Most studies regarding the ageing of bruises and cutaneous wounds focus on histochemical and histopathological features (Hernandez-Cueto et al., 2000). Enzymatic markers such as esterases, phosphatases and peptidases began to be used in the early 1960’s as indicators of wound timing as it relates to the peri-mortem period (see Raekallio, 1972; Thornton and

---

<sup>7</sup> AUC is often referred to as the ‘accuracy score’. The closer this score is to 1, the better the decision making ability of the group, and is represented by the area under the ROC curve. It does not represent the ‘error rate’ associated with the study. A score of 0.5 would indicate no better ability at decision making than random guessing.



Jolly, 1986), in recognition of the fact that simple macroscopic observation by forensic practitioners was failing to allow even a modest determination of wound age. Newer immunohistochemical techniques are attempting to characterise the appearance and relative concentrations of pro-inflammatory cytokines and other proteins at different stages of wound healing (Dressler et al., 1999; Kondo et al., 1999; Kondo, 2007; Sato and Ohshima, 2000). Yet even some of these studies have produced conflicting results (Cecchi, 2010). Notably, in a comprehensive review of wound assessment procedures published in 2000, while the determination of the age of the wound was classified as an all-important step in wound examination, the author failed to find one article that advocated simple visual determination in order to provide an estimate of wound age (Ohshima, 2000).

The micro-morphological progression of cutaneous wound healing is well documented (see Williamson and Harding, 2004). While macroscopic recognition of generalised inflammatory responses, such as those seen in cutaneous wounds and bruising, is considered relatively easy, there are a large number of local, exogenous and systemic factors that influence the progression and resolution of such a response (Grellner and Madea, 2007). Microscopic morphological changes have been more closely related to immunochemical marker appearance and concentration, but many of these changes are not visible on a macroscopic level (Cecchi, 2010). Additionally, due to the multitude of factors that affect the rate and progression of healing (including on an immunohistochemical level, (see Betz, 1994) as well as a macroscopic one), without seeing the initial wound it is not realistically possible to determine the age of a wound with any degree of accuracy.

Baum and Arpey (2005) discussed the macroscopic changes associated with cutaneous wound healing and their correlation to cellular and molecular events. In their article, they noted the key macroscopic changes as visible to the naked eye for cutaneous wound healing. Specifically, they noted that the presence of ‘inflammation’, involving the classic heat, swelling, erythema and pain, can last from 15 minutes to 6 days. Granulation tissue begins to appear at approximately 4 to 7 days, with wound edge re-approximation occurring anywhere between 4 and 14 days. Decreased redness of the wound, associated with a decrease in capillary density and subsequent scar formation can take weeks to months. They note that while the healing of wounds is classically divided into inflammation, proliferation and maturation, none of these phases correspond to a precisely defined period of time and they all overlap to some degree.

Even using sophisticated microscopy, combined with histochemical and immunohistochemical techniques, the ageing of bruises has been shown to be highly

unreliable (Bariciak et al., 2003). Dailey and Bowers (1997) conducted a literature review on the ageing of bitemarks, focussing on the bruising aspect of injury, and found evidence of at least 19 variables affecting the appearance of a bruise. Langlois and Gresham (1991) studied 89 subjects, and found that the only reliable conclusion that could be drawn from pure macroscopic evaluation of a bruise was that the presence of a yellow discolouration, when viewed on the victim (and specifically *not* from a photograph) indicated it was likely to be more than 18 hours old. These authors concluded it unlikely that a bruise could be reliably aged from appearance alone. To further complicate matters, there is evidence to suggest that both inter and intra-observer agreement regarding the colour description of a bruise itself is also poor. A study by Munang, Leonard and Mok (2002) found that complete agreement between doctors and nurses on the colouration of bruises when viewed directly on the victim was only 21%. No agreement at all was found in 39.5% of cases. Even the intra-observer variability was of concern, particularly when comparing the results of viewing the bruise directly as opposed to a photograph. When the same observers were shown photographs of the bruises they had earlier examined, nearly one-quarter of bruises (24%) were reported in complete disagreement with their original assessment. Complete agreement occurred in only 31% of bruises. Maguire and colleagues (Maguire et al., 2005) conducted a review of the data concerning bruising in children, and concluded that a bruise could not be accurately aged from either a clinical assessment or a photograph. They further suggested that ‘any clinician who offers a definitive estimate of the age of a bruise in a child by assessment with the naked eye is doing so without adequate published evidence’.

Few, if any studies have attempted to characterise the age of bruises and wounds using purely macroscopic observational methods. Ageing of bruises and wounds via qualitative observation is perhaps best described as an educated guess. That is not to say that odontologists cannot necessarily determine an older, or healing bitemark from one that is relatively fresh. As clinicians, odontologists have specific training and experience in pathology and trauma, and it is likely that most, if not all, could make an educated guess as to whether a bitemark was recent or not. However, the extent to which this can be further distilled into discrete time periods other than perhaps as ‘occurring within a few days’, as opposed to ‘an older [or healing] wound’, is not yet supported by the research literature.

## **Effect of Quality of Injury on Assessment**

Pretty (2007) developed a scale to characterise the forensic significance of bitemark injuries, correlating this with the severity of the mark, in an attempt to conceptualise the quality of a bitemark and hence its potential usefulness. He advocated that neither the least severe

(involving bruising) or the most severe bites (involving complete excision of tissue) provide much useful forensic information. The marks that yield the most information tend to involve varying degrees of both, and thus the middle section of the scale represents the most forensically significant injuries. Bitemarks that have characteristics described by Pretty as ‘very obvious bruising with small lacerations associated with teeth on the most severe aspects of the injury’, or ‘numerous areas of laceration with some bruising, some areas of the wound may be incised’ represent categories three and four of the six-category scale, and thus are said to have the highest forensic significance. This is postulated due to the higher number of class and individual characteristics of the dentition likely to manifest in such injuries, when compared to more severe episodes of injury involving loss of tissue, or lesser injuries where only bruising is present.

In one of few studies to consider the effect of evidence quality on practitioner performance, Bowers and Pretty (2009) in a post-hoc analysis of 49 bitemark cases where there was expert disagreement, found that the vast majority of these ranked at 2 or below on the bitemark severity and significance scale. The development of this scale represents a useful tool in development of a decision matrix regarding the relative strength of conclusions that can be drawn from bitemark injuries, however, the main concern is the inter-examiner reliability for practitioners using this scale in order to define the quality of a bitemark.

Avon and colleagues (Avon et al., 2010) conducted a study on practitioner accuracy in relating dentitions to bitemarks by using three dentitions to make 18 bitemarks (6 bitemarks per dentition), and then asking each participant to relate one of the three dentitions to each of the bitemarks. Practitioner accuracy varied widely, even among bitemarks made by the same dentitions. For example, considering those bitemarks made by dentition A, correct responses by ABFO diplomates ranged from as little as 5% for mark number 1, to over 80% for mark number 6. These were for marks that were made by the same dentition — hence the differences in accuracy can only logically be due to the differences in quality of the bitemark. Unfortunately, the study does not provide examples of these bitemarks, so it is impossible to determine what features a ‘good quality’ bitemark, or at least one that resulted in substantial practitioner accuracy, had compared to one of poorer quality. While this was not the intent of the study, its results clearly demonstrated that the quality of the bitemark itself has an effect on the overall ability of an odontologist to relate the mark to a dentition.

Aside from the inherent quality of the bitemark itself, one must also consider the quality of the presentation of such evidence. The manner in which the mark has been photographed, whether there is inclusion of a suitable scale, the lighting, focus, angulation and other

technical features can all influence whether a bitemark is likely to yield a reliable result or not during the comparative phase. Where analysis proceeds on the basis of a photograph of a bitemark, as most anecdotally do, the current literature on bitemark analysis<sup>8</sup> is fundamentally in agreement regarding details such as the need to include a two-dimensional scale, such as the AFBO No 2, and to photograph the image as close perpendicular to this scale as possible so as to minimise distortion (Bowers, 2004; Dorion, 2005). It is recommended that an orientation photograph, one that includes the surrounding anatomical features, is taken to provide anatomical context for the injury. Multiple images from several angles, taken in both black and white as well as colour are recommended, particularly for curved surfaces, and this corresponds with literature from other fields regarding the need to accurately photograph patterned injuries such as forensic pathology, forensic medicine and dermatology (Dolinak et al., 2005; Schosser and Kendrick, 1987; Shkrum and Ramsay, 2007). There is no reason to suspect that bitemark analysis does not require the same standards. The problem lies in the quality of the casework material presented to the odontologist. Anecdotal evidence suggests that the majority of bitemark cases are handed to forensic odontologists well after the incident has occurred, meaning that the quality of any photographic evidence is usually lacking, having been taken by inexperienced or non-informed crime scene technicians, doctors, nurses or police. Despite poor quality evidence, many odontologists proceed with analysis, and it is here that a fundamental problem regarding the reliability of any such conclusions based on such inferior information is encountered.

## **Reporting – Agreement in Language and Semantics**

Another criticism many forensic science areas face is ensuring communication of their results without error, exaggeration, or misleading the trier of fact (McQuiston-Surrett and Saks, 2008). This basic ability is essential in order to allow the trier of fact to accurately assign appropriate weight to the testimony in accordance with the true meaning of the results. In Britain, the House of Commons Science and Technology Committee Report has also highlighted the need for clear and consistent terminology in forensic science evidence presentation, particularly statistical evidence (House of Commons Science and Technology Committee, 2005). Edmond noted that the language employed by forensic scientists in the court should be linked to testing, frequencies and publicly accessible data sets, and it is often the failure of such testing that leads to language problems in the first place (Edmond, 2008b). Some have argued that by following the so-called ‘scientific method’ in forensic

---

<sup>8</sup> And that of forensic imaging.

science research would allow proper evaluation of the evidence by universally recognised standards, and thus the expert witness could avoid resorting to vague, unquantifiable statements such as ‘probable’ and ‘possible’. Mnookin (2008) has then suggested that claims of ‘absolute’ and ‘positive’ identification could then be replaced by more modest and meaningful claims about the significance of the term ‘match’.

The use of conclusionary terms in order to neatly summarise the odontologist’s findings appears to be rather inconsistent. In Australia, it was agreed upon in 2010 via the Australian Society of Forensic Odontology that each State would adopt the Interpol standard for reconciliation reports for identifications. The terminology was as follows (Interpol, 2011):

<i>Identification</i>	There is absolute certainty the PM and AM records are from the same person
<i>Identification probable</i>	Specific characteristics correspond between PM and AM but either PM or AM data or both are minimal
<i>Identification possible</i>	There is nothing that excludes the identity but either PM or AM data or both are minimal
<i>Identity excluded</i>	PM and AM records are from different persons
<i>No comparison</i>	No comparison can be made

Figure 6.2 — INTERPOL standard terminology for DVI Odontology Identification Reports

These terms, however, were meant to be applied to ante-mortem and post-mortem reconciliation reports, and are not necessarily applicable to bitemark analysis reports. There is currently no standard terminology for bitemark comparison results in Australia, and most odontologists appear to resort to these terms when making conclusions about the identity of a biter.

The American Board of Forensic Odontology lists the following terminology as an agreed-upon way of expressing bitemark analysis *conclusions* as published in the ABFO Reference Manual (American Board of Forensic Odontology, 2011):

The Biter
The Probable Biter
Not Excluded as the Biter
Excluded as the Biter
Inconclusive

**Figure 6.3 ABFO Endorsed Terminology for Bitemark Analysis Conclusions**

Aside from the addition of the words ‘the biter’, they are also semantically equivalent to the terms used for identification of remains. They further note that ‘all of these opinions are stated to a ‘reasonable degree of dental certainty’. There is no definition offered for this phrase, nor why it should be used, or why experts should attest to it. This phrase has been met with scepticism from the courts and legal academics alike, with both camps noting that it is vague and meaningless. The American Law Institute noted that the ‘reasonable-certainty standard provides no assurance of the quality of the expert's qualifications, expertise, investigation, methodology, or reasoning’ and thus rejects an expert’s need to testify to such (American Law Institute, 2010). Evidence scholars such as Professor Faigman have also noted that the phrase ‘within a reasonable degree of medical certainty’ has no empirical meaning and is ‘simply a mantra repeated by experts for purposes of legal decision makers who similarly have no idea what it means’. (Faigman, 2010).

Others have claimed that ‘because the phrases, “reasonable degree of [medical] probability,” “reasonable degree of [medical] certainty,” and the standards they purport to embody are so poorly understood and inconsistently used, courts should prohibit their use. Because these phrases are so inconsistently used, their value is small, while the likelihood that the finder of fact will apply a definition different than the expert witness is high. In other words, use of the phrases is confusing and not helpful to the finder of fact’ (Abbott and Magnusson, 2008). One query often raised is regarding the use of the term ‘medical’ (or ‘dental’) certainty. Is there some empirical difference between medical certainty and other forms of certainty? Forensic practitioners have yet to satisfactorily explain the difference.

The ABFO also specifically states that ‘terms assuring unconditional identification of a perpetrator, or without doubt, are not sanctioned as a final conclusion.’ (American Board of Forensic Odontology, 2011). If an odontologist concludes that person *X* was ‘the biter’, then what other possible meaning could this have, other than an identification of the perpetrator? What ‘the biter’ then means is anybody’s guess.

The ABFO guidelines also issued a statement regarding the terminology to be used for describing the *likelihood of an injury being a bitemark* as follows:

<b>Bitemark</b>	Teeth created the pattern; other possibilities were considered and excluded. <i>Criteria:</i> The pattern conclusively illustrates: a) classic features; or b) all the characteristics, or c) typical class characteristics of dental arches and human teeth in proper arrangement so that it is recognizable as an impression of the human dentition.
<b>Suggestive</b>	The pattern is suggestive of a bitemark, but there is insufficient evidence to reach a definitive conclusion at this time. <i>Criteria:</i> General shape and size are present but distinctive features such as tooth marks are missing, incomplete or distorted or a few marks resembling tooth marks are present but the arch configuration is missing.
<b>Not a Bitemark</b>	Teeth did not create the pattern.

Figure 6.4 ABFO Guidelines on Terminology for Assessing Bitemark Injuries

These terms are equally nebulous. There is mention of ‘classic features’, ‘typical class characteristics’ and ‘characteristics’ of dental arches and teeth, but no elucidation as to what these features of characteristics might be or how they might appear in a bitemark. The ABFO definition of ‘not a bitemark’ seems to simply state the obvious ... ‘teeth did not create the pattern’. It is difficult to see how these ‘explanations’ offer anything to help odontologists, lawyers or jurors understand what they mean.

Guidelines for terminology for describing the likelihood of an injury being a bitemark were published via the website of an ABFO Diplomate on his website, last updated in the year 2000 (figure 6.5). A similar chart is offered for a description of terms used to link a bitemark to a suspect (figure 6.6).

<b>Ordinate Ranking of Terms</b>	<b>Connotation</b>
definite positively	No doubt in my mind it is a bitemark
reasonable medical certainty highly probable	Virtual certainty; allows for the possibility of another cause, however remote
probable	More likely than not
possible similar to consistent with conceivable may or may not be cannot be ruled out cannot be excluded	Such a mark could have been produced by teeth but not necessarily and could have been created by something else; no commitment to likelihood
unlikely inconsistent improbable	Less likely than not
incompatible excluded impossible	No doubt in my mind it is not a bitemark; represents something else
indeterminable shouldn't be used insufficient	Pattern shows insufficient characterization to comment on teeth as a cause

Figure 6.5 — Terms to indicate that an injury represents a bitemark



<b>Terms</b>	<b>Connotation</b>
reasonable medical certainty extremely probable high degree of certainty	Virtual certainty; no reasonable or practical possibility that someone else did it
very probably probably most likely	More likely than not
possible consistent (with) can't exclude	Could be; may or may not be; can't be ruled out
improbable	Unlikely to be the biter
ruled out excluded exculpatory could not have; did not eliminated dissimilar no match; mismatch incompatible not of common origin	Not the biter
inadequate inconclusive insufficient	Insufficient quality/quantity/specificity of evidence to make any statement of relationship to the biter
evidence has no probative (forensic) value unsuitable (should not be used) non-contributory non-diagnostic	Of no evidentiary value

Figure 6.6 Degrees of Certainty Describing The Link Between the Bitemark and Suspect

Bell noted that ‘using numbers and percentages to represent opinions is inappropriate unless a specific statistical analysis on a case has been done’. This is good advice, but use of terms such as ‘most people’ still implies some sort of statistical or numerical probability, even without explicitly saying so, and so use of this terminology doesn’t really solve this issue. The connotations on the right hand side still lack formal definitions and are equally vague in terms of their standard definition. Another concern with these tables is the ability of odontologists to actually agree to use these terms, let alone agree on their definitions. More importantly, these terms continually reference the notion of ‘certainty’, when this issue actually bears no relation to ground truth. These connotations, while perhaps assisting in standardisation of terminology, are fundamentally inappropriate to use in this manner.

In order to limit the problem of never ending semantic subtleties, fingerprint examiners have limited their conclusionary scope to one of three statements (Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST), 2011). They claim that doing

so is inherently *more* reliable than allowing individualised opinions, by providing what are essentially black and white criteria that are not subject to interpretation by individual examiners or by courts. The Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST) maintained in 1997 that ‘Friction ridge identifications are absolute conclusions. Probable, possible, or likely identifications are outside the acceptable limits of the science of friction ridge identifications’ (Scientific Working Group on Friction Ridge Analysis Study and Technology, 1997).

<b>Individualisation</b>	Individualization is the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.
<b>Inconclusive</b>	An inconclusive conclusion occurs when an examiner is unable to individualize or exclude due to an absence of complete and legible known prints (e.g., poor quality fingerprints and lack of comparable areas), or when corresponding features are observed but not sufficient to individualize. Likewise dissimilar features may be observed but not sufficient to exclude. In such an instance, the inconclusive conclusion means that the impression needs to be reexamined using clearly and completely recorded known impressions. In either case, the inconclusive conclusion means that the unknown impression was neither individualized nor excluded as originating from the same source.
<b>Exclusion</b>	Exclusion is the decision by an examiner that there are sufficient features in disagreement to conclude that two areas of friction ridge impressions did not originate from the same source. Exclusion of a subject can only be reached if all relevant comparable anatomical areas are represented and legible in the known exemplars. Notes and reports shall clearly state if the exclusion refers only to the source or the subject.

Figure 6.7 SWGFAST Endorsed Fingerprint Examiner Conclusions

Firearms and toolmark examiners in the United States have endorsed a similar terminology, allowing for a fourth category of ‘unsuitable’ (that is encompassed in the ‘inconclusive’ category that fingerprint examiners use):<sup>9</sup>

<sup>9</sup> As an aside, I found the SWGGUN guidelines to be somewhat amateur when compared to the relevant SWGFAST documents. Spelling errors, inconsistent (or entirely absent) formatting, incomplete sentences, lack of any form of version control or dating system, and statements that do not make any linguistic or scientific sense abound in these ‘guidelines’. The policy on ‘Criteria For Identification’ ends with guideline 2.2.3: ‘Currently the interpretation of individualization / identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience’. It is unclear as to how exactly this assists firearms and toolmark examiners to justify their conclusions when it admits subjectivity, fails to elucidate what the ‘scientific principles’ it is based on are, and then contradicts this by saying that interpretation of identifications is based on training and experience.

<b>Identifications</b> [ <i>sic</i> ]	Agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool.
<b>Inconclusive</b>	<ol style="list-style-type: none"> <li>1. Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification.</li> <li>2. Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.</li> <li>3. Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.</li> </ol>
<b>Elimination</b>	Significant disagreement of discernible class characteristics and/or individual characteristics.
<b>Unsuitable</b>	Unsuitable for microscopic examination.

Figure 6.8 — AFTE Glossary – Range of Conclusions Possible When Comparing Toolmarks

Forensic document examiners have used a nine-point scale as endorsed by the American Board of Forensic Document Examiners, which has controversially been disallowed in court on several occasions due to its lack of scientific underpinning:<sup>10</sup>

1	Identification (a definite conclusion that the questioned writing matches another sample)
2	Strong probability (evidence is persuasive, yet some critical quality is missing)
3	Probable (points strongly towards identification)
4	Indications [that the same person] did [create both samples] (there are a few significant features)
5	No conclusion (used when there are limiting factors such as disguise, or lack of comparable writing)
6	Indications [that the same person] did not [create both samples] (same weight as indications with a weak opinion)
7	Probably did not (evidence is quite strong)
8	Strong probably did not (virtual certainty)
9	Elimination (highest degree of confidence)

Figure 6.9 Standard Terminology for Expressing Conclusions of Forensic Document Examiners  
(ASTM Designation E 1658-04.)

According to Professor Moenssens in an article he wrote in 1999, research that sought to

<sup>10</sup> See *United States v Rutherford* (2000), *United States v Van Wyk* (2000), *United States v Hines* (1999), and *United States v Santillan* (1999)

validate statistically every one of the nine levels of document examiner opinions that the ASTM standard provides was underway (Moenssens, 1999b). None of this research appears to be yet available, and as demonstrated in Chapter 4, evidence that has referred to this nine-point scale has actually been excluded in US courts, due to the failure of the profession to validate its application.

The problem faced by the SWGFAST and SWGGUN scales, and including any scale endorsed by forensic odontologists along similar lines, is that until research is conducted that establishes under what circumstances identifications or exclusions can be made (i.e. what the threshold limit for any given conclusion is), then the scale simply becomes a ‘best guess’ exercise as to the likelihood of any particular dentition or object having made the mark. There is yet to be any research that clarifies the thresholds at which these conclusions become valid as there are no scientifically defined benchmarks. There is therefore little *scientific* basis for these conclusions in pattern analysis disciplines. They remain an expression of practitioner certainty not based on objective or scientific standards.

## Error and Accuracy Rates in Bitemark Analysis

Error rate is often regarded as one of the four *Daubert* criteria courts are obliged to consider when assessing expert evidence. However, the US Supreme Court gave no more specific direction other than ‘the court ordinarily should consider the known or potential rate of error and the existence and maintenance of standards controlling the techniques operation’ (*Daubert v Merrell Dow Pharmaceuticals, Inc.*, 1993). This has been criticised as a vague and meaningless way to assess the reliability of any scientific evidence without further clarification (Cole, 2004b; Jonakait, 1993), and *Daubert*’s minimal elucidation has proved to be unhelpful to most courts, who appear to have ‘had great difficulty with the concept of error rate’ (Faigman et al., 2007).

Faigman and colleagues noted that there are several issues concerning error rate. Firstly, they ask what sort of error is one concerned with as there is a clear distinction between the ramifications of false positives and false negative results. Errors may also be random, due to factors such as sampling techniques, or systematic. Systematic errors tend to introduce bias, whereas random errors generally cancel each other out when the process is cumulative, but present problems when validity relies on individual results. Secondly, they noted that the existence of error itself is not a genuine concern, as all applied science has some error rate associated with it, the larger issue being what percentage of error, or mistakes, is acceptable.

The courts have yet to define what this threshold might be, and perhaps rightly so, as the costs are different in different legal contexts.

Budowle and colleagues (Budowle et al., 2009) suggest that ‘the most critical error would be a false association’, however this is not necessarily so, as the consequences of a mis-identification are different, depending on the circumstances of the case. Is the false association of a fingerprint of person suspected of committing burglary and sentenced to three months in prison worse than the false exclusion of a fingerprint of a person who murdered and raped a young girl, and subsequently went on to commit more crimes of a similar nature? Consideration of such difficult moral questions is not helpful for the forensic expert. As noted by Faigman and colleagues (Faigman et al., 2007), the legal system has so far avoided grappling with this question given the difficulty of a blanket approach to such emotive moral issues. One advantage of being able to describe such characteristics as error rate lies in the difficulty of then being able to attack the discipline on the basis of claims of infallibility. Polski believes that many of the criticisms of fingerprint and other pattern evidence would become void by presenting answers about the reliability of the testimony as a probability, akin to DNA. He has stated; ‘I don’t think it would affect the weight of the identification one bit, but I do think it would make it much more difficult for those people who want to attack that absolute identification to attack that kind of statement; after all, that is a scientific way of looking at things, because in science, nothing is absolute.’ (in Pyrek, 2007, p. 268).

In reality, information about error rate is gained by proper application of the first *Daubert* principle, that of testing. The modern scientific method includes the analysis of error as an integral part of the empirical testing of the hypothesis. It is therefore perhaps a symptom of this lack of testing for forensic pattern identification sciences that many of them can then offer very little data on the accuracy of their techniques.

Several commentators have noted that there is really no such thing as an error ‘rate’, only the likelihood of a given error occurring. Jonakait (1993) noted that not all error data carries the same weight, and that information regarding only that particular technique in use in the case at bar should be considered. Additionally, he noted that the term theoretical or ‘potential’ error rate has no meaning, and that only data regarding error that is empirically derived from a technique’s practical use has any value in assessing its reliability. Furthermore, *Kumho*’s ‘task at hand’ requirement specifically requires the judge to assess reliability only as applied in a particular case, and thus a requirement for a generic ‘error rate’ established for any particular discipline, or even technique is perhaps questionable under this auspice. After all,

just because all fingerprint examiners are wrong 1% of the time does not mean that this *particular* examiner in this *particular* case has a one in 100 chance of also being wrong. However, some form of assessment of error *should* be necessary, as it has long been held as a principle of expert testimony that the expert must furnish the trier of fact with sufficient information to enable assignment of the relative weight of their evidence. Despite the perception that the mandate for ‘error rate’ arose from *Daubert*, the requirement for an assessment of the likelihood of error has always been encompassed via the more general principles of expert evidence.

There are many types of errors that may affect the likelihood of the observed conclusion being true or false. Dror and Charlton (2006) characterise these errors as one of three main types: human error, accounting for fraud, negligence and competence; instrumentation and technological errors, defined by the limits of the analytical technology used, including random errors such as chance or equipment break-downs; and methodological errors, encompassing the inaccuracy of algorithms used to detect a particular effect; and the influence of psychological and cognitive errors on a particular result. The most commonly assessed aspect of error in odontology addresses the issue of competence and the accuracy of our analytical algorithms, yet there are other sources of error relevant to the practice of forensic odontology that have been largely ignored by the literature.

### **Error Rate and Proficiency Tests**

Authors do acknowledge that it is difficult to measure the true accuracy of forensic analysis. Crime lab proficiency test data obtained in the late 1970’s indicated very poor performances, resulting in a questioning of the reliability of forensic laboratory work (Jonakait, 1991). This study was criticised by several authors soon after its publication, largely for using data that was already 13 years old at the time of writing and seemingly ignoring the technological advances that had eventuated in the interim, but it was also supported in its conclusion by others (Letters Forum, 1991). If error rates were indeed as high as 2%, as suggested by Peterson based on similar data (Peterson and Markham, 1991), up to 1400 people per year may have been erroneously convicted, or erroneously exonerated, on fingerprint evidence alone (Koppl, 2005). A fingerprint proficiency study conducted in 1995 found over 30% of answer sheets containing at least one false identification, resulting in one in five of the participants potentially providing damning evidence if the proficiency test represented actual case work (Grieve, 1996). Dror notes that the relative rarity of errors, and under what conditions they occur at a practical level in forensic science remains unclear, however with proper research and subsequent systematic application, these errors can be reduced and minimised (Dror and Charlton, 2006).

Most studies attempt to quantify error by conducting proficiency-type tests, despite the National Research Council Report on DNA evidence (National Research Council, 1996) having observed that the ‘objective of ... proficiency testing is to improve laboratory performance by identifying problems that need to be corrected. ...[It is not]... designed to measure error rates’. Koehler (1997) regards the fact that proficiency tests are not designed to measure error rate as irrelevant, when such data can easily be extracted from such results, yet he and others have noted that these proficiency tests are usually non-blinded, relatively easy internal tests that tend to be favourable towards low false positive error rates. Despite this focus on proficiency testing, it is worth noting that courts in the United States and elsewhere in the world (Australia included) have not yet required experts to include the results of proficiency testing when calculating random match probabilities or likelihood ratios for DNA evidence, for example.

Critics of proficiency testing as a means of measuring error rate claim that the use of a statistic derived from past performance invariably gives the ‘wrong value’ for prediction of future events. As the National Research Council report commented, ‘error rates are therefore in a state of flux’ and cannot meaningfully be described by a single statistic (National Research Council, 2009). As Koehler has also noted, this is true for virtually any statistic, and this does not necessarily mean that the data is not useful for predicting future performance. Furthermore, failure to consider a non-zero error rate places undue probative weight on the random match probability (Koehler, 2007). Another claim regarding the inadequacy of using statistics derived from proficiency testing is that the error rate for the ‘industry’ is not necessarily representative of single cases, and thus cannot be used to describe the error rate associated with a single, particular case. This notion has been well described in behavioural science research, where frequency statistics should often be discounted if they are not sufficiently case-specific.

Others note that a large number of proficiency tests would be required in order to derive a meaningful statistic with a narrow confidence interval. Even when errors are not made, proficiency testing at the individual laboratory level would need to be extensive to produce error rate estimates that carry meaningful confidence limits. Until such extensive testing has been completed, Koehler advocated the use of industry-wide error rates when estimating error rate. Even though this might be perceived as being ‘unfair’ to those laboratories that perform above average in such tests, he noted that not all analysts can possibly have a rate above that of the average, and in cases where this is claimed, the analyst should be able to verify this with data. If no data is presented, then the industry-wide rate should be used (Koehler, 1996).

Proficiency testing requires well-designed and properly conducted tests in order for their results to be meaningful to the practice of forensic identification. Cole (2004b) noted that the tests undertaken by fingerprint examiners since 1983, while providing some data on overall (industry) error rate are difficult to interpret due to a number of flaws. These tests were mostly conducted by mail, were unsupervised and untimed. The relative difficulty of each of the tests was also difficult to establish relative to casework.<sup>11</sup> These proficiency tests are often non-blinded, and it cannot be established who exactly takes them. Anecdotal reports have indicated that sometimes these tests are assigned to novice examiners, or involve ‘committees’ who collaboratively complete the test (Peterson and Markham, 1991). Overall, the false positive rate has declined since 1995, however, this may be due to a reduction in the difficulty of the test, or a more thorough approach to its completion by laboratories.

Firearms and toolmark proficiency testing, like fingerprint testing, often focuses on the false positive and fails to appreciate realistic conditions. A large study conducted by Hamby, Brundage and Thorpe (2009) involving 507 participants from ten different countries over ten years concluded that ‘the fact that there were no actual errors shows that the test procedure used to ascribe bullets fired from consecutively rifled barrels is reproducible on a worldwide basis’. However, the design of this study failed to allow any direct relevance to forensic practice to be drawn. Participants were given two known bullets fired from each consecutively rifled Ruger pistol. They were also given 15 ‘unknown’ bullets, which they had to match to the correct barrel, with the possibility of up to three bullets being from the same barrel. They were given a finite set from which to make their associations, and the study does not address the likelihood of a bullet being falsely matched to an eleventh, (unseen) consecutively rifled barrel. If there was inclusion of a bullet that was fired from an eleventh, unseen, gun, how many false associations would have occurred then? The study involving a closed set of barrels from which to associate bullets does not often represent real-life forensic conditions.

### **Proficiency Tests and Bitemark Analysis**

Because of the existence of several techniques for analysing bitemarks, it is difficult to quantify a ‘generic’ error rate for the discipline. Rawson and colleagues (Rawson et al., 1986) found that the average accuracy of odontologists in correctly matching bite marks to the dentition, when using the American Board of Forensic Odontology’s scoring system, was

---

<sup>11</sup> Although anecdotal evidence suggests that they were relatively easy – see Somabat (2001) and *US v Llera Plaza II* (2002).



only 66%. This system was first introduced in 1981, but was subsequently withdrawn in the late 1980's due to its inherent unreliability (Bowers, 2007a).

Many odontology studies, such as the 4<sup>th</sup> ABFO Bitemark Workshop (1999) (See Arheart and Pretty, 2001) did not specify the use of any particular technique, and thus assumed a cross-section of contemporaneous techniques was used. The mean accuracy of 32 examiners, calculated via a Receiver Operating Characteristic curve, was 86%, however, the authors note that this result should be interpreted with caution, as there is a very wide confidence interval due to the small number of cases that each odontologist was asked to examine (each participant was only asked to assess four cases). It is also difficult to assess whether the difficulty of the cases accurately represented those encountered in practice. They report individual practitioner 95% confidence intervals as wide as 0.24 – 1.00. Some practitioners obtained a 95% confidence interval from 1.00 – 1.00, due to the zero error rate (100% accuracy) obtained during the analysis of the four cases and seven dentitions, however this does *not* indicate the 'error rate' is zero, with a 95% confidence interval. If we consider the ABFO study as simultaneously providing practitioners with a maximum of 22 separate 'tests', each asking 'does this model of the dentition match the bitemark?',<sup>12</sup> according to the mathematics described by Koehler (1996),<sup>13</sup> we still cannot be certain that the error rate is actually zero. We can only be 95% confident that a practitioner who correctly assigns all of the dentitions as 'matching' or 'non-matching' has an error rate somewhere between 0 and 13%. Unfortunately, the ABFO data is not provided in such a way as to enable calculation of an overall error rate for the discipline, but it can be ascertained that some practitioners achieved a 'perfect' accuracy score.

Bowers (2007a) has noted that two of the four ABFO studies failed to generate publishable information regarding the accuracy of forensic odontologists in determining the source of a bitemark, citing a 'combination of poor experimental design and limited analytical prowess by the profession.' He noted that the results of the 4<sup>th</sup> ABFO experiment, as described above, fail to accurately capture error information, as the *lowest* level of accuracy actually attainable by the test design was as high as 71%. This occurs because even if an examiner got a case completely wrong, the fact that he failed to erroneously implicate the remaining 5 dentitions (6 dentitions, minus the correctly matching one) means that these were all considered as

---

<sup>12</sup> It is not established whether once a dentition was assigned, it was considered in the analyses of the other bitemarks, (i.e. sampling without replacement) hence the number of 'tests' was likely to be 22 (that is, 7+6+5+4) as once a dentition was identified to a bitemark, it was probably not considered as a candidate for the other marks. Even if it was, increasing the number of 'tests' to 28, the 95% confidence interval for accuracy lies between 0 and 10%, assuming that no errors were recorded.

<sup>13</sup>  $(1-e)^n \leq \alpha$ , where  $e$  = upper bound error,  $n$  = number of successful reports, and  $\alpha$  = 1 – confidence level

‘correct’ non-matches. His analysis of the 1999 ABFO study demonstrates that the median false negative rate (where the true biter was excluded) was 22%, but the median false positive rate (where an innocent suspect was erroneously identified as the biter) was 63.5% – in other words, forensic dentists are more likely to be wrong than they are right.

It seems intuitively odd that experts should have an accuracy rate less than that achieved by pure chance, and analysis of Bowers’ statistical treatment of the results reveals that this too is probably incorrect, and in fact goes too far in proclaiming inaccuracy. The participants were not asked to either positively identify or positively exclude the suspect, they were asked to rate the likelihood of the suspect being the biter on a Likert-like scale from 1 to 7.<sup>14</sup> To consider categories 1-3 as ‘positive’ does not accurately represent the intention of these categories. A significant proportion of conclusions fell into category 3 – defined as ‘possible’ – which implies that there must be some value to the alternative hypothesis ‘possibly *not*’. Thus someone who has identified a bitemark as ‘possible’ has not necessarily made a positive identification, and so it cannot strictly be an erroneous conclusion – it should not necessarily be considered a ‘false positive’.

Likewise, ‘inconclusive’ does not realistically represent a definitive negative identification either, and the impact of the assignment of a bitemark to the inconclusive category when the suspect did indeed bite the victim should not be considered a ‘false negative’, as the bitemark evidence alone might truly have been inconclusive. Thus, at least the ‘inconclusive’ and ‘possible’ categories are reasonable conclusions that comfortably qualify the uncertainty of the bitemark evidence and do not necessarily warrant consideration as a true ‘error’.<sup>15</sup> Furthermore, it is difficult to parse the true meaning of these results because the ground truth, or ‘gold standard’ in three of the four cases was taken as the result of the original examining odontologist and it is not certain that they were correct. The authors vouch for their accuracy by stating that these three cases were successfully litigated, however this still ignores the fact that there is propensity for error on the part of original examiner. The baseline level of truth regarding a potential match was not technically known, it was only assumed from the verdict in each case. There is an inherent danger in equating judicial outcomes as known truths, as has been demonstrated by the number of convictions from bitemarks where suspects have later been exonerated by DNA or other evidence. The lack of any suitable definition for the rating criteria 1-7 also makes the conclusions of each

---

<sup>14</sup> The scale for ascertaining the link between the suspect dentition and the bitemark was described as 1=reasonable medical certainty, 2=probable, 3=possible, 4=improbable, 5=incompatible, 6=inconclusive, and 7=non-diagnostic.

<sup>15</sup> The quality of the bitemark itself might not have allowed a distinctive conclusion to be drawn.

participant difficult to assess. The overall poor design of the ABFO study has greatly hindered any authors' attempts at extracting meaningful data.

Pretty and Sweet (2001a) conducted a study of inter and intra-examiner variability in bitemark analysis using digitally produced overlays of a suspects dentition. Overall the more experienced forensic odontologists, represented by the cohort from the American Board of Forensic Odontology, had the better accuracy (83.2%), but the cohort with supposedly less forensic experience, that of the American Society of Forensic Odontology, had the lowest false positive rate (11.9%) and also the highest false negative rate. The group with no forensic experience, represented by general dental practitioners, had the lowest false negative rate, but also the highest false positive rate (22.0%), indicating a tendency to erroneously identify someone as the biter. These results should be interpreted with caution, as the study employed a 'forced decision model', requiring a dichotomous positive/negative answer that is not usually employed in forensic casework, at least by the prudent odontologist. Moreover, this study suggests that training and experience have little effect on the application of overlay to bitemark identifications, as there was no significant difference detected between any of the measured values of all three cohorts, although the authors note that more detailed questionnaires would be required to correctly identify the variables regarding experience and training. The authors' conclusion that the study 'satisfies the requirement of *Daubert* in relation to determining error rates and other quantifiable values' is somewhat optimistic. The authors do not address the main issue that the data highlight, in that it appears that board-certified, supposedly trained and well-experienced odontologists are *still* wrong nearly one in five times when attempting to make an identification based on bitemarks.

Blackwell and colleagues (Blackwell et al., 2007) obtained an ideal, ROC-derived false positive rate of 15% in matching models of bitemarks made in wax to the models of biters using a laser scanning technique and morphometric analysis of the model and bitemark images. The main concern with this study is that it represents a novel technique not commonly used in routine bitemark analysis. Additionally, the authors included the premolars in the analysis, data that is often missing in forensic bitemarks but may contribute to the ability of the practitioner to discriminate between dentitions. A separate study, using another computer-generated model where known models were used to generate bitemarks<sup>16</sup> and then an algorithm developed to relate these models back to the imprints they created, recorded an accuracy rate of 88% (Flora et al., 2009). This was substantially better than the manual matching score by odontologists (two odontologists scored 73 and 60% respectively,

---

<sup>16</sup> Using a lipstick-coated model to provide an imprint on a foam surface

resulting in an average score of 67%), however, the computer algorithm considered three-dimensional data, whereas the odontologist only considered data in two dimensions (such as is encountered in the overlay comparison technique). Again, of note in this study is not the accuracy obtained by the computer model, but the relatively poor accuracy obtained by the odontologists.

In Avon and colleagues' study (Avon et al., 2010) they establish that the overall 'critical' error rate (i.e. false association) for Board-certified odontologists was 0% for one bitemark, 2.5% for another, and 16.9% for a third, giving a mean error rate of 6%. While this seems impressively low, it should be noted that the first two figures quoted here were for marks with which a definitively matching dentition was provided. For the third bitemark, no matching dentition was provided (i.e. the third dentition did not actually make any of the marks given to the participants to analyse). This means that unlike the previous two cases, there was no ability for the odontologist to see a comparison between a 'matching' and a 'non-matching' dentition – and hence the participants were required to make a judgement call on whether the association between bitemark 3 and the third dentition was strong enough to claim identity. Nearly one in five times, the Board-certified odontologist made the *incorrect* call. It is this error rate that has the most relevance to current bitemark matching practices, and suggests that the risk of false association is significant in even the more experienced practitioners. This rate was even higher for participants who were not board-certified but were members of various forensic dental associations.<sup>17</sup> The risk of false association where bitemarks were presented with a non-matching dentition, but without a true match for comparison, was reported as high as 56.3% in this study.

In Avon and colleagues' study, errors characterized as 'non-critical' involved false negative decisions – where the bitemark did match the dentition provided, but no match or an inconclusive result was recorded, however, to characterise these as non-critical seems dismissive. While it may be comforting to know the difference between a false negative and false positive 'error' rate, particularly when the false positive error rate appears low, the most fundamental error rate calculations need to consider the number of times the odontologist simply 'gets it wrong' in order to provide a general indicator of the reliability of the science. Avon's study suggests that the 'overall' error rate is as high as 35.3% for board-certified odontologists, and up to 43.7% for other practitioners of bitemark analysis. It should be noted that the authors of this study considered an 'inconclusive' decision as a non-critical error when it was reached for a bitemark that did actually match one of the dentitions provided. This is probably unfair as it increases the numerical error rate while not directly

---

<sup>17</sup> Yet all had still had training and experience in bitemark cases.

being associated with practitioner performance. The ‘inconclusive’ decision is most likely to have been reached due to the poor quality of the evidence, rather than poor performance by the examiner. This is borne out in the results, which demonstrate that bitemark A had a much higher incidence of ‘inconclusive’ when compared to bitemark B, suggesting that the quality of bitemark B was superior from an evidentiary perspective. This is acknowledged in the discussion section of their paper, although the authors fail to comment on the significance of the quality of the bitemark and its relationship to practitioner performance.

## Conclusion

There appear to be a number of key issues faced by bitemark analysis when faced with criticism regarding its scientific basis. The ability of odontologists to accurately associate bitemarks with dentitions is questionable, and this stems from the fact that there is so much variation in fundamental processes that lead to these conclusions, such as methods of evidence collection, methods of analysis, the significance of certain patterns in the injury, and the actual meaning of any conclusions drawn. Given these concerns, it is little wonder that practitioners struggle to find much agreement amongst themselves in this field.

Qualitative assessment of bitemark injuries is also fraught with danger, with evidence suggesting that the ageing of wounds is virtually impossible with any degree of accuracy. Furthermore, studies on distortion also suggest that those features considered ‘cardinal’ hallmarks of dentitions, such as arch width, may not be accurately represented, particularly on human skin.

The literature concerning the accuracy of bitemark analysis is thin and of dubious quality, and it is here that much of the criticism regarding the theory and practice of this ‘science’ lies. Proficiency studies give some indication of the reliability of bitemark analysis, which at best appears to be mediocre, however, given the wide variation in techniques for analysis; the language used to describe a match, and the quality of evidence used in these studies, it is difficult to say how reflective of practitioner accuracy during actual case work these ‘error rates’ are. The next chapter attempts to put some of this research into context, with a review of bitemark practices and casework drawn from contemporary Australian odontology practice.

## CHAPTER 7

# Bitemark Analysis in Contemporary Australia

---

One of the principal difficulties encountered in establishing the reliability of bitemark analysis lies in the variable nature of the processes by which conclusions regarding identity are reached, and in the subsequently variable nature of these conclusions themselves. In order to establish these primary baselines as relevant to Australia, data was drawn from two sources: casework since the year 2000 from two major odontology centres; and through a series of interviews with practising odontologists from around Australia as described in the introductory chapter. The purpose of this exercise was to establish baseline practices for bitemark analysis that could then be critically assessed.

In order to garner some representation of the nature of bitemark casework in Australia, all case files involving bitemark analysis were reviewed from two major centres in Australia that were presented between January 2000 and May 2010 (table 7.5). All five major odontology centres were contacted,<sup>1</sup> however, only two were able to provide access to historical case files. This was not due to unwillingness on the part of the other centres, but was a reflection of the fact that the police usually retained the original files, and bitemark casework was thus often scattered across multiple police centres, in different divisions (such as homicide, sexual assault, child protection), making the task of consolidation of these records beyond the limits of one author. Regardless, this exercise represents the largest consolidated bitemark review conducted in this country, with over 119 files being examined.

Interviewees were drawn from the members of the Australian Society of Forensic Odontology (AuSFO). While AuSFO is more than 50 members strong, very few of these members are regularly involved in casework, and so potential interviewees for establishing bitemark analysis practices were limited. Those members of AuSFO known to be involved in at least permanent, part-time work, or known to have had significant experience in forensic odontology were contact by the author, and 15 members agreed to be interviewed for the

---

<sup>1</sup> Western Australia, Queensland, New South Wales, South Australia and Victoria

purposes of this research. While this number is relatively small, it still represents the majority of practising odontologists in Australia.<sup>2</sup>

Odontology studies that attempt to characterise practitioner agreement have mostly used ‘forced decision’ models in order to present dichotomous, or at least limited, options for participants to choose in an effort to focus their responses. A conscious decision was made not to employ this model in this thesis, as the process of bitemark analysis rarely results in simple yes or no answers, and so this would be unlikely to reflect the nature of decision-making and bitemark reporting in this country. Odontologists were free to express their opinions using terminology of their choosing, and to narrow their options for the purposes of this study would decrease the applicability of its findings to actual practice.

The relatively few number of practising odontologists in Australia prevents large-scale statistical analyses of these results due to the limited sample size. While kappa scores are a convenient way of expressing examiner agreement, they become less significant as the sample size decreases. This is partly the reason for maintaining a largely qualitative approach as used in this study. Fleiss’s kappa scores ( $\kappa$ ) represent a convenient way of expressing inter-examiner agreement across multiple raters ( $n > 2$ ) for several items, and despite the fact that the small sample size means that these scores in general should be interpreted with caution, they still provide useful insight into patternicity in examiner agreement. Fleiss’s kappa score was calculated for some of the data that follows in order to summarise numerically the relative level of inter-examiner agreement in both the comments regarding the origin of the injury, and assigning severity and significance scores. Weighted kappa scores were not deemed appropriate for this study due to the use of a qualitative interval scale rather than a pure ordinal scale (Sim and Wright, 2005). Generally, kappa scores above zero indicate that there is more agreement between practitioners than would be expected in accordance with random chance.

## Results

### Experience in Forensic Odontology

Individual experience level in the field of forensic odontology varied from 4 years to 35 years. The median level of experience was 19 years, with the mean being 22 years experience.

---

<sup>2</sup> There are less than 20 members of AuSFO routinely engaged in permanent part-time forensic casework totaling more than one day per week.

<b>N Years Experience</b>	<b>N Odonts</b>	<b>%</b>
<5 years	1	7
6-10 years	3	20
11-15 years	0	0
16-20 years	4	27
21+ years	7	46

Table 7.1 Experience in Forensic Odontology

### Post-graduate qualifications in Forensic Odontology (at time of interview)

All except four of the participants in this study had the same basic level of training – a graduate diploma in forensic odontology – and some had additional higher degrees. The highest level of post-graduate qualification attained (specifically in forensic odontology) ranged from nil to doctoral-level. The total number of participants who had completed at least a Graduate Diploma in Forensic Odontology was 11, with one of these participants later completing a Master’s level qualification. Three participants had non-forensic related Master’s degrees, and one participant was in the process of completing a PhD.<sup>3</sup>

<b>Post Graduate Qualification</b>	<b>N Odonts</b>	<b>%</b>
None	2	13
Graduate Diploma	10	67
Masters	2	13
Doctoral	1	7

Table 7.2 Qualifications of Participants

### Courtroom Experience in Presenting Bitemark Evidence

Courtroom experience does not necessarily translate into casework experience, as very little bitemark casework ever proceeds to court, yet less than half of the participants had *ever* testified in court on bitemark analysis. Of those that had, only one person had testified more than 5 times in their career.

<sup>3</sup> These particular individuals’ qualifications are not reflected in table 7.2.



N times testified	N Odonts	%
0	9	60
1	3	20
2-5	2	13
6 or more	1	7

Table 7.3 Court Experience

The relative proportion of odontologists who have had significant courtroom experience in the area of bitemark analysis was very low. Despite this, the majority of participants could almost certainly be deemed ‘experienced’ odontologists, with only three having less than ten years in the field.

### Case Type

Anecdotally, requests for bitemark analysis by law enforcement agencies in Australia are comparatively rare. Interviewees all expressed the relative infrequency with which they were asked to provide opinions on bitemark cases, and most noted that they were even more rarely asked to match a bitemark to a suspect. More often than not, they were asked to provide an opinion on whether or not the mark could have been made by human teeth:

*Its not common to be asked to identify someone from a bitemark. They want to know if it is a bitemark or not, or if it is human or not. Then they can add it to the list of information they have about the injuries on the person. For example the child protection unit, they usually aren't interested in getting me to say who the person was, but they are often interested in whether it was an adult or a child, or whether they could have bitten themselves. (Participant #1598)*

*For example, in one case ... I didn't have to say who did it, but I just had to say which was upper and which was lower. In order to corroborate a victims story. That's a much more common scenario. (Participant #7239)*

Concerning the nature of cases that are associated with bitemark analysis, the review of case files revealed that child abuse represented the largest proportion of case type associated with bitemarks, with n=45 (37.5%). Assault and sexual assault represented the next highest proportion, with 48 cases involving an allegation of either assault (n=24, 20.2%) or sexual assault (n=24, 20.2%). 6 cases involved analysis of a bitemark on a deceased human body, with only 4 of these cases officially deemed homicides. Homicide represented only 3.4% of all bitemark cases encountered in a ten-year period. Formal reports were raised in 77 (65%)

cases. Only two of these cases proceeded to trial, neither of which required testimony by the odontologist. 13 cases involved animal bites, with the majority of them occurring on inanimate objects, such as surfboards and scuba equipment. Only three of these cases involved a bite on human skin.

Crime	N Cases	% Cases
Child abuse	45	37.8
Domestic Violence	2	1.7
Assault	24	20.2
Homicide	4	3.4
Sexual assault	24	20.2
Other	4	3.4
Animal bites	13	10.9
Unknown	5	4.2
Mark On Deceased	6	5.0

Table 7.4 Case Types Associated with Bitemarks

## Quality

Interviews with odontologists revealed that most were generally dissatisfied with the material quality of evidence presented to them for analysis:

*We often see photographs of poor quality, without scale markers, strange angles, perspectives, and they are rarely ever taken of the injury in the position the person was during the assault. (Participant #7249)*

Another participant claimed similar difficulties with photographic images:

*Often the police don't take the photo perpendicular to the injury, there is no scale, or it's not an ABFO scale, bad lighting ... and sometimes there is only one photograph of one bitemark when there are clearly others on the body. (Participant #7744)*

The retrospective case data supported these concerns. While 101 out of 119 cases involved the presentation of colour photographs (84.9%), with only one case having a black and white photograph, only 51 of these 102 cases (50%) had an ABFO No. 2 scale included in the photographs. 33 of the remaining photographs (32.3%) had some other form of scale, usually a linear rule (in two cases, this was a generic plastic 30cm ruler). The remaining cases appeared not to have any form of scale included in the photographs available for

review. Direct examination of the bitemark itself by the odontologist occurred in 67 out of 119 cases (56.3%), including some cases where physical evidence in the form of an inanimate object was presented. This practice was generally believed to be useful by the odontology community, particularly given the poor quality of photographic records.

*Sometimes you just have to be there. It's like looking at photos of oral lesions, I know I've taken photos and, particularly if it hasn't been taken right, they tend not to look anything like they do in the mouth. Sometimes it just doesn't always represent what you think you are looking at.* (Participant #9037)

Another aspect to consider regarding the quality of the evidence is the quality of the injury itself. During the review of casework, severity scores were assigned by the author to all bites deemed to be of human origin, in accordance with a scale developed and validated by Pretty (2007). 30 out of 113 bitemarks were deemed unsuitable for assignation of a severity score.<sup>4</sup> Severity scores were assigned to 83 separate bitemarks, using both the written descriptions and pictographic representations (as published by Pretty) as a reference. The majority of bitemark evidence appeared to fall in the 'lowest forensic significance' category, with 42% of bitemarks rating a 1 on Pretty's scale. Only 24% of bitemarks were at the highest forensic significant range, with scores of 3 or 4.

<b>Severity &amp; Significance Score</b>	<b>N Cases</b>	<b>% Cases</b>
1	35	42.2
2	25	30.1
3	13	15.7
4	7	8.4
5	3	3.6
6	0	0.0

Table 7.5 Forensic Significance and N Cases

### Method of Analysis of Bitemarks

When asked, most odontologists expressed a preference for using a hand-tracing on acetate, or generation of a digital overlay using Photoshop as a means of comparison with a dentition. Other methods included a purely qualitative description of the relationship between tooth marks and a model of the dentition; creating a contour map of teeth by photographing a model submersed in ink; and creation of a clear acrylic replica of the

<sup>4</sup> Due to either their presence on an inanimate object, or the absence of a photographic record of the injury.

dentition that is then placed directly over a life-sized photograph of the bitemark. Several odontologists commented that they would often use a supplementary method, such as a metric analysis of intercanine distances, or individual tooth lengths, in order to strengthen their conclusions.

Primary Method of Analysis	N Odonts	%
Digital overlay	5	36
Acetate	7	47
Other	3	17

Table 7.6 Methods of Analysis

Analysis of the retrospective casework data generally agrees with this trend, although this is clearly dependent on the practitioners that are at the sites where this casework was undertaken. In 27 out of 119 cases the method of analysis of the injury was unable to be ascertained. Of the remaining 92 cases, the most common method of analysis of the injury was characterised as simply ‘qualitative’ (n=61, 66.3%). This method involved a visual assessment of the bitemark, where no measurements were taken, and no tracing produced. This method was generally used when the odontologist had been asked to provide an opinion on whether the mark could be a bitemark, whether it is of human origin, or whether it could be an adult or child. Use of a digitised version of a photograph, as per Bowers’ and Johansen’s method (Bowers and Johansen, 2002) was the next most commonly used technique, with n=22 (24%).

Method of Comparison	N Cases
Digital overlay from photograph (Bowers & Johansen)	22
Modified digital overlay from photograph	0
Hand-traced overlay from photograph	0
Injured tissue excised for examination	0
Quantitative measurements compared	3
Other method (specify)	6
Qualitative only	61
<i>Not applicable</i>	9
Method of analysis unknown	18
TOTAL	119

Table 7.7 Methods of Analysis

There were no instances of odontologists using hand-tracing methods or excising the bitemark at these two centres. There were 10 instances of odontologists using a direct comparison method, where acrylic templates or the stone models were manually superimposed over the injury, although this supplemented digital overlay analysis in four of these cases. Metric analysis, where physical measurements were taken of various aspects of the injury, proceeded in only three cases.

### **Consistency of Opinions**

Interview participants were shown six images of bitemarks, all of them taken from actual casework presented to one odontology centre in Australia. They were asked to assume that these were photographs given to them by a law enforcement agency for their initial comments. No contextual information was given to participants, and they were free to express as much or as little opinion as they liked regarding the image before them. An overview of odontologist's comments concerning each of these images is given below.

#### ***Image A***

The first image consisted of a latent mark considered to be of low severity and significance in accordance with Pretty (2007). The image had significant specular reflection from the flash used to light the area, and the scale consisted of a clear plastic 30cm ruler that was neither placed parallel to the injury nor in the same plane as the centre of the mark.



**Figure 7.1 — Image A**

There was a central area of ecchymosis consistent with rupture of the small vessels beneath the skin. The photograph was taken in extreme close-up, and it was difficult to deduce what part of the anatomy the bitemark was on. The photograph was also not taken perpendicular

to the centre of the image. The lower left portion of the mark appeared to be out of focus with the remainder of the image.

Most odontologists commented on the poor quality of this image:

*There are a few problems here; its on a curved surface, and there is distortion...* (Participant #8547)

*There is a problem in that its just this one area, the scale is to one side, you don't know what angle its been taken from.* (Participant #1624)

*The problem we going to get with analysing this is the scale. Its on a curved surface, and no attempt has been made to take the photograph from different points around that curvature, so scaling becomes a problem with all but the central axis draw horizontally through the injury. If that was the only evidence I had, I wouldn't consider that one photograph to be sufficient for analysis.* (Participant #5195)

*I would say it's a bad photo. We haven't got a very good scale. You've got a curved surface but a flat scale. There isn't enough contrast. It could also be pathology.* (Participant #9037)

*It's a relatively poor bitemark for image analysis* (Participant #7249)

Despite there being general agreement that the image quality was poor, several participants offered further comments. The central area of ecchymosis drew a wide variety of opinion regarding its origin, with some describing it as a scraping or tongue thrust bruise:

*I'd say it possibly could be a bitemark... perhaps a scraping or a tongue thrusting in the centre.* (Participant #1598)

Most attributed it to 'sucking':

*There is an area around [sic] the injury that could be due to some suction force.* (Participant #6487)

*This is someone that's had a bit of a suck, hasn't he.* (Participant #7239)

*[There's] some bruising from suction in the centre, possibly.* (Participant #6828)

Yet others attributed it to the artefact created when the tissue is squeezed between the upper and lower dentition:

*It's got the characteristic bruise in the middle, which is caused by compression of small surface vessels as the jaws close.* (Participant #5195)

There was frank disagreement as to whether the image even represented a true bitemark or not:

*On class characteristics, yes it's a bitemark and it is likely to be human. (Participant #5195)*

*It could also be pathology. You can't assess something like that from a photograph. It could be a suction cup mark. (Participant #9037)*

*I'd say its possibly a bitemark, and its possibly a human bitemark. (Participant #1624)*

*I'd say it could be anything from a bottle top to a bitemark. (Participant #2737)*

*I'm not sure that it even is a bitemark. (Participant #1660)*

*It's a bitemark, yes. (Participant #7249)*

*I don't think it's a bitemark, I think it is a bit tiny. Even a child with deciduous dentition, I think its too small. I'm just looking at the ruler, its only 4 cm. Its very small. (Participant #1786)*

### **Image B**

This image consisted of a medium-range photograph of a decedents chest area, with two marks that were considered bitemarks by the initial investigating team. These marks would be considered of moderate significance in accordance with Pretty's scale. They overlap slightly and are on two different angles to one another.



**Figure 7.2 — Image B**

There is an ABFO No. 2 scale included in the image, although it is not being held parallel to either mark. It does not appear to be on the same plane as the centre of the marks themselves, and the thumb of the person holding the scale is obscuring part of the orientation

circle in the bottom left corner. There are significant other injuries in the immediate vicinity that overlap portions of the bitemarks, including what appear to be artefacts from medical equipment that interfere with the continuity of the lower mark. The lighting towards the left superior aspect of the photograph, including over the superior aspect of the suspected bitemarks is poor and not consistent with the remaining two thirds of the photograph.

Some odontologists noted the difficulties associated with this photograph, although fewer commented on this issue than for Image A:

*Distortion with the ruler is a bit of a problem with this. For a preliminary report I would say that it looks to be a bitemark, but for analysis the problem is the angulation of the photograph, the distortions across the slide, the ruler is set forward.* (Participant #1624)

*The scale is adequate, but there are a couple of problems. The scale hasn't been held at right angles to the long axis of the camera, so there will be some differences between the individual millimetre elements of the vertical scale.* (Participant #5195)

The presence of other injuries, and specifically the artefacts caused by what is presumably medical equipment were noted by couple of participants:

*There seems to be too much artefact. What looks like a nipple, all of this presumably medical-related equipment, lots of bruising.* (Participant #9037)

*These little blisters here might be caused by some sort of suction thing on the skin, a medical device perhaps.* (Participant #6828)

*[These marks,] they might be caused by some sort of appliance, or something like that.* (Participant #1598)

Again there was considerable inconsistency in whether or not the injuries could have been bitemarks or not:

*I don't know whether they would be bitemarks at all. But using computer imaging to fade in and out for example, I don't know whether you would find them to be bitemarks at all* (Participant #1598)

*Well to me, my initial reaction is that that is not a bitemark.* (Participant #6487)

*I would suggest that those are, and certainly on size, consistent with being human bitemarks* (Participant #5195)



*You are starting to pick up some individual tooth marks here, and you have two arch marks there (Participant #3224)*

*[This is] probably a human bitemark (Participant #7744)*

*I'm not even convinced that they are bitemarks (Participant #1786)*

Others went further still, offering details on orientation, motive, and even the appearance of a suspect dentition:

*There are elements there such as individual bruise marks that suggest that towards the 'B' side there are lower teeth, and a more diffuse pattern opposite to that mark, suggesting the bruising pattern of palatal surfaces of upper teeth. (Participant #5195)*

*Yes, you could say that he has bitten her twice. Was his head turned slightly to the side maybe? I'd say probably. I think the centre line is there and going up like this, because of these two points there, so I think the head was tipped to the side... In all probability these are the two lower canines, and there is canine here and there might be one here. The bloke has obviously had a bit of a suck, but not as much as the other one, and we've got a bit of bruising and there might be something wrong with his central incisors. (Participant #7239)*

*I'd be very suspicious. There's the nipple there, there is often a sexual implication in that. Possibly a suction mark too. (Participant #2737)*

### **Image C**

Image C consisted of a photograph of circular wound pattern on the upper arm of a young child. There is some bruising and a series of individual abrasions that roughly form an elliptical outline. The photograph itself has not been taken perpendicular to the centre of the injury, and the inferior aspect of the child's arm where the injury extends is out of focus.

There are two small forensic rulers, held perpendicular to one another for use as a scale, but unfortunately the vertical rule is not in the same plane as the injury, nor the same plane as the horizontal rule. The two scales are not also at 90 degrees to one another.



Figure 7.3 — Image C

The difficulties associated with scaling this image were only commented on by two participants:

*We have a curved surface and no attempt to take the rulers to follow that injury around the curvature. Secondly the long axis of the lens is not focussed on the centre of the injury, it is focussed on one side of it, which is the only part that I would consider for analysis here. The scales are not at right angles to one another, so scaling is going to be a problem in two dimensions. (Participant #5195)*

*There is problem with the flexibility of the rulers, and also you've got the upper arch and the lower arch, because it's a curved area its difficult to do. There would have to be reservations with the photo and the scale. (Participant #1624)*

The fact that the scale was generally not on the same plane as the bitemark was noted by one other participant:

*They are holding the scale there, but it is on a different plane to the bitemark. (Participant #1786)*

This image caused considerable variation in opinion as to its origin. Some participants did not believe it was a bitemark, or were uncertain:

*I would be hesitant as to whether it is a bitemark. (Participant #6487)*

*It could be a burn, it looks like it has been there for a while. I wouldn't bet on that being anything frankly. (Participant #9037)*

*No, I don't even think that is a bitemark.* (Participant #2159)

Others were convinced that it was definitely a mark caused by teeth:

*That's definitely a human bitemark.* (Participant #7744)

*Now that's a good bitemark.* (Participant #7249)

*The strength of those bites are [sic] very good.* (Participant #1624)

Several participants offered comments on the orientation of the injury, however, they disagreed as to which side represented upper teeth and which side represented lower teeth:

*It looks like a bitemark. I'd say that the right hand side, closest to the thumb might be caused by upper teeth, and the left side caused by lower teeth.* (Participant #6828)

*This here could be the upper teeth, on the right side of the photo.* (Participant #1786)

Yet consider:

*I'd say ... these are the upper anteriors here [indicating on the left].* (Participant #7239)

This last participant offered further information on the biting mechanism:

*He's sort of bitten in then bitten again. He bit once and didn't get good enough grip. And as he is biting he is going into protrusive.* (Participant #7239)

The 'double bite' phenomenon was also commented on by another participant:

*He has had a go twice. It's been pinched in.* (Participant #1786)

### **Image D**

Image D was a close-range photograph of a more severe injury. The mark itself shows drag marks and laceration injuries. There is an ABFO No. 2 scale in the photograph, although the horizontal portion of the scale has been partially cropped. The fact that the upper part of the vertical section of the scale appears larger than the lower portion of the vertical scale suggests that the photograph was taken from a slightly higher angle than the plane of the scale. While the numbers and scale marks are visible, it would not be possible to adjust for the angulation of the camera to the scale with any accuracy due to the fact that only a portion of both lower circles are shown in the photograph. It is difficult to be certain of the location of the injury due to the tight focus, although the contours in the top right hand corner of the photograph suggest that we may be looking at the upper arm. The scale is not parallel to the injury.



Figure 7.4 — Image D

Comments regarding the suitability or otherwise of the scale and photograph were few. The most significant issue, according to one odontologist, was whether the scale was digitally added after the original photograph was taken:

*I have some issues with the scale, the contrast in the rest of the photograph seems to be at odds with the scale. Was it actually in the photograph, or was it added in after the fact? There is a shadow towards the top of the vertical scale that suggests it might have been [there to begin with].* (Participant #5195)

The same odontologist also made the following observations regarding the quality of the photograph:

*I am distressed that central axis of the lens is pointed towards the top of the injury rather than its centre. The specular reflection from the flash also mitigates seeing a lot of detail there.* (Participant #5195)

No other participant commented on the actual quality of the photograph, aside from:

*I wouldn't mind having a bit more of the lower ruler, but that's pretty good.* (Participant #1624)

There were again considerable differences in opinion regarding the origin of this injury. Some odontologists agreed that it represented teeth marks:

*Yes that's probably a human bitemark.* (Participant #7744)

*This does have the characteristics of a likely bite. This is characteristic of someone who bites and then withdraws very rapidly.* (Participant #5195)

*Yes I think that's a very good bitemark.* (Participant #7249)

*Now here's a bitemark with drag marks.* (Participant #7239)

One odontologist commented that they were certain it represented teeth marks due to their own experience with similar markings:

*I know from experience that that's teeth because I did a case at the beginning of the year, that when I first looked at the images I didn't think they were teeth, because the injuries were so severe. But when I saw the models, and scratched them down my arm, they looked just like that.* (Participant #8547)

Yet others were more cautious:

*This is a very interesting mark. Again, its got some elements of what you would consider to have bitemark information in it, but there are other elements that I would think you would have to provide an explanation for.* (Participant #6487)

*Honestly I don't think it's a bitemark. Oh hang on, possibly some scraping there, again there is so much... there could be any number of things that could have caused that. Whether this is individual tooth marks here I doubt. I've never seen anything like that.* (Participant #3224)

*Well I don't think it's a bitemark.* (Participant #1598)

*I would say that this might have multiple aetiologies. I wouldn't confidently say that was a bitemark.* (Participant #1660)

*It looks charred to me, and I'm not certain it's a bitemark.* (Participant #9037)

*It could be a bitemark, but its just pinched in and drawn blood. I mean, it could just have been pliers. Its very hard to tell upper from lower there. Also the shape of the pinch is quite unusual for a bite, I would have expected more of an arc. I wouldn't exclude a bitemark, but it could be other things that have pinched in.* (Participant #1786)

Two participants commented on the potential orientation of the mark, and again, were inconsistent in their answer:

*I'd say that was the lower arch there on the left. (Participant #7239)*

*If I had to have my arm twisted, I could probably say that this is the upper arch there on the left of the photo. (Participant #1786)*

Two comments were made about the circumstances surrounding the injury:

*There was a struggle involved here, because the arm has been moved as the biter has started to bite. Or finishing the bite. Probably starting the bite. (Participant #7239)*

*This is characteristic of someone who bites and then withdraws very rapidly. (Participant #5195)*

### **Image E**

Image E is a medium-range photograph of an ovoid injury on the back of an individual's neck. The injury appears to be older than any of the previous images shown. The ABFO No. 2 scale is well placed in relation to the mark, despite the fact that the scale is being held so that nearly a third of the circle in the left corner of the scale is obscured. The camera itself does not appear to have been placed perpendicular to the scale as the shape appears distorted in both the horizontal and vertical dimensions.



**Figure 7.5 — Image E**

Only two participants commented on this fact:

*The ruler is tilted back. (Participant #1624)*

*The central axis of the lens not central to the injury, and the long axis of the lens isn't perpendicular to the scale. (Participant #5195)*

This error is probably correctable via use of a digital imaging program as the scale appears well-placed relative to the injury. Unfortunately, there is a high degree of specular reflection on the lower portion of the scale, which may indicate that some detail of the injury has been lost in the photograph.

One comment was made regarding the poor contrast of the image:

*Its not a terribly good photo, the contrast isn't very good. (Participant #7239)*

Eight odontologists made specific comment that this bitemark was 'old'. None volunteered any specific mention of how old, apart from one odontologist who suggested:

*It's been months and months since that was made. (Participant #3224)*

It is interesting to note that the description of these marks varied between participants, with some describing them as bruises, and others as scabs, or even puncture marks:

*Here are some bruises. This is an old injury. (Participant #9037)*

*It could be an old bitemark, but it has scabbed up. (Participant #1786)*

*We've got quite a few good little puncture marks there that might [help]. (Participant #7239)*

This suggests two very different types of original injury, as scab formation does not occur unless there is a break in the skin with associated haemorrhage. This suggestion was noted by at least one participant:

*I think there was possibly a break in the skin here originally. (Participant #3224)*

Whether the mark was caused by teeth or not again elicited a wide variety of opinion. Most agreed that it could have been caused by teeth, but noted that there were other objects that could have caused similar injuries:

*It could be a bitemark, but it could be caused by other things too, a chain with a broach or something. (Participant #6828)*

*You could say she possibly had been bitten. On a Likert scale I'd put it at a 6. A few other things could make a mark like that too. (Participant #7239)*

*This again could be a bitemark, but its very old, I wouldn't give an opinion on it. You can't get five individual tooth marks there to call it an arch mark. (Participant #3224)*

One participant found the image more convincing:

*Probably to me, one of the more convincing bitemarks I've seen. (Participant #6487)*

And again others simply assumed it was definitely a bitemark:

*This is an aged bitemark, you are looking at healing here. (Participant #1624)*

*It's an old bitemark, but I'm not sure you would get much out of it. (Participant #8547)*

Only one odontologist offered an orientation on the mark:

*The arc closest to E is clearly from upper teeth, and the arc furthest from E is from lower teeth. (Participant #5195)*

Its relative usefulness for identification purposes was generally agreed to be low. This was mainly deemed to be because of the age of the injury.

*It has features that are consistent with a bitemark, but I think its too old to give a useful analysis on. (Participant #1660)*

*It has some characteristics of a bitemark, but you couldn't go any further than that. (Participant #1598)*

*All you could do with a set of models would be to measure between the canines. If it is an approximate measurement then maybe you have a bitemark. (Participant #7239)*

*It's possibly a human bitemark, but not worth any further analysis. (Participant #7744)*

*This isn't going to get you anywhere other than to say that if this is a bitemark, if you are told a story, that they were bitten on the back of the neck, then maybe. (Participant #6828)*

### **Image F**

The final image was a photograph of a mark on the upper arm of a victim. The relative severity and significance of this injury is low in accordance with Pretty's scale, as it represents only relatively diffuse bruising. There is a flexible rule that has been stuck on the victims arm under the bitemark, and it follows the contour of the upper arm. It is relatively parallel to the injury, however the ends of the scale are out of focus as they bend away from the lens. The absence of a vertical scale makes it difficult to assess whether the photograph



was taken with the lens parallel to the injury or not. Given the curved nature of the arm, and the relative focus of the scale, there is concern that the two arcs of the injury would be out of focus, as only the central portion of the scale is sharp in the image, which corresponds to the *space* between the [supposed] arch marks themselves.



Figure 7.6 — Image F

Comments regarding this image were fairly short and to the point. Most agreed that it was of low forensic significance, consisting only of bruising, but it was likely to be a bitemark:

*I think its probably a bitemark, but old.* (Participant #7744)

*I'd say that was highly likely to be a bitemark.* (Participant #2737)

*It's probably a bitemark.* (Participant #7239)

*I think that it's a bitemark, but I would find it very hard to say what is upper and lower.* (Participant #1786)

Several odontologists preferred not to specify:

*It's really just a blurred kind of bruising, so again I can't really give an opinion.* (Participant #3224)

*Again all you have is bruises. A bit better than the other one perhaps, but it could just be someone's fingers that have caused it.* (Participant #9037)

The lack of a suitable scale and the potential problem with angulation was noted by several participants:

*This stick on scale is the problem. It has to be on a flat surface, and at right angles to the lens. (Participant #5195)*

*There would be issues with the angulation here. (Participant #1660)*

*There isn't much of a scale on it, and you've also got the curvature of the arm, so it's difficult to look at it here (Participant #7239)*

*The scale is not good. I wouldn't think there was any value in analysis. (Participant #7744)*

Despite these concerns, several odontologists felt that pattern analysis would be possible:

*Its got more of a pattern of a bitemark, I think you could get some information out of it. There is a probability that you could get information to relate it to a dentition. (Participant #1598)*

*You could get a pattern out of it, but as to how useful the pattern would be, I'm not sure. (Participant #8547)*

*I would think you would be able to .... Yes, given the patterns, you would look for things in the dentition there that may allow you to 'fail to exclude' and 'find some concordance' possibly. (Participant #1660)*

This bitemark was generally considered to be old by those who offered comment:

*Its too old isn't it. It has been there for a long time. The bruises are beginning to go green, and everything fades. (Participant #7239)*

*Yes it's an old one. (Participant #8547)*

*The bruising is indistinct as it is late. (Participant #1624)*

*It's probably a bitemark, but old. (Participant #7744)*

### **Assessment of the Forensic Value of Bitemarks**

Some odontologists repeatedly offered comments on the relative value of the images presented, however, there was considerable variation in their opinions in this regard. Consider the following contrasting comments, which were made regarding image D:

*Well I don't think it's a bitemark. Possibly some drag lines here, maybe. Once again I don't think you'd get anything out of it. It may be a bite, but even if it was I'm not sure it would lead you to anything. [...] I doubt you'd get anything out of that one. (Participant #1598)*

And:

*Yes that's pretty good. Good for analysis. I wouldn't mind having a bit more of the lower ruler, but that's a pretty good image to work with.* (Participant #1624)

Image C also represented a cause of contention for several participants, who commented on the poor angulation of the photograph and the quality of the scale. One commented:

*[This image is] absolutely not [suitable for analysis] [...] I would have reservations about doing anything with that. If I were asked to analyse this I would probably say no.* (Participant #5195)

However, another participant said

*'No, that's a good bitemark, I could work with that, yes, definitely.'* (Participant #7249)

This inconsistency indicates a fundamental flaw in the methodology of bitemark analysis as practised in this country, and should lead to concerns regarding the reliability of any conclusions reached about matching such a bitemark to a dentition. Interview participants were asked to rate the images, A through F, in accordance with Pretty's severity and significance scale, in order to ascertain the value of this instrument in assessing the forensic value of bitemarks. They were initially shown the text descriptors only and asked to rate the images on this basis. Following this exercise, they were then shown the visual exemplars, and asked to reassess their initial scores in light of the examples shown. Results for the severity and significance (Pretty) scoring exercise are given in Table 7.8 below.

Bitemark Image	Reference	Participant # Assigned Severity and Significance Score														
		7249	1598	9037	6828	7239	5195	8547	7744	2159	2737	1624	1660	1786	8172	3224
A	Text (T)	2	1	1	1	1	2	2	1	1	2	2	1	1	1	1
	Visual (V)	2	2	2	2	1	2	2	1	1	2	2	2	1	2	1
B	T	3	1	1	2	3	3	3	3	2	3	2	2	1	2	2
	V	3	4	3	2	3	2	3	3	1	3	2	2	1	2	2
C	T	4	2	2	3	4	4	4	3	3	4	3	3	4	3	0
	V	4	3	4	4	4	1	4	4	3	4	3	3	4	3	0
D	T	5	1	3	4	6	4	4	4	5	5	5	0	6	5	0
	V	5	5	5	4	5	5	5	4	4	5	5	5	5	5	0
E	T	1	2	2	1	2	3	2	1	1	0	0	2	2	4	0
	V	1	3	2	1	2	3	1	1	1	0	0	2	2	4	0
F	T	1	3	2	0	1	2	1	1	1	1	1	2	2	2	1
	V	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1

Table 7.8 Comparison of Pretty scores assigned using text-only aid and visual aid

### Agreement on Origin of Injury

Interview participants also often made comments on the likely origin of the latent mark, frequently offering opinions on whether or not they suspected the mark was actually a bitemark. After analysis of each participants' comments regarding the potential origin of the injury, it became clear that there were five preferred ways of expressing the likelihood that the injury was a bitemark, ranging from definitively positive comments — *'Well, this is definitely a bitemark'*; moderately positive comments — *'It is probably a bitemark'*; weakly positive comments — *'it might be a bitemark, it is possible'*; comments expressing genuine uncertainty — *'I'm not sure if it is even a bitemark'*; and negative comments — *'I don't think it is a bitemark'*. A number of practitioners did not specifically comment on whether the injury was a bitemark or not, however, their subsequent comments made it obvious that they had proceeded on the assumption that it was — *'Well, it's a relatively poor bitemark for image analysis. There are some features you could use, but I'd like to see some extra photos from different perspectives, maybe a black and white photo. It's not enough information by itself'*.

In the analysis that follows below, these were categorised separately so as to avoid undue weighting towards the positive end of the scale without direct evidence from verbal comments by the participant (as 'no comment, but assumed to be a bitemark') The relative frequency of these comments made by each participant for each bitemark is illustrated in table 7.9 and figure 7.7.

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>No comment</b>	2	2	1	5	3	6
<b>Definitely / Consistent with a bitemark</b>	2	7	7	1	4	3
<b>Probably a bitemark</b>	4	1	1	2	0	3
<b>Possibly / Could be a bitemark</b>	5	1	1	3	7	2
<b>Not sure if it is a bitemark</b>	2	1	3	2	1	1
<b>Don't think it's a bitemark</b>	1	3	2	2	0	0

Table 7.9 N comments regarding origin of injury

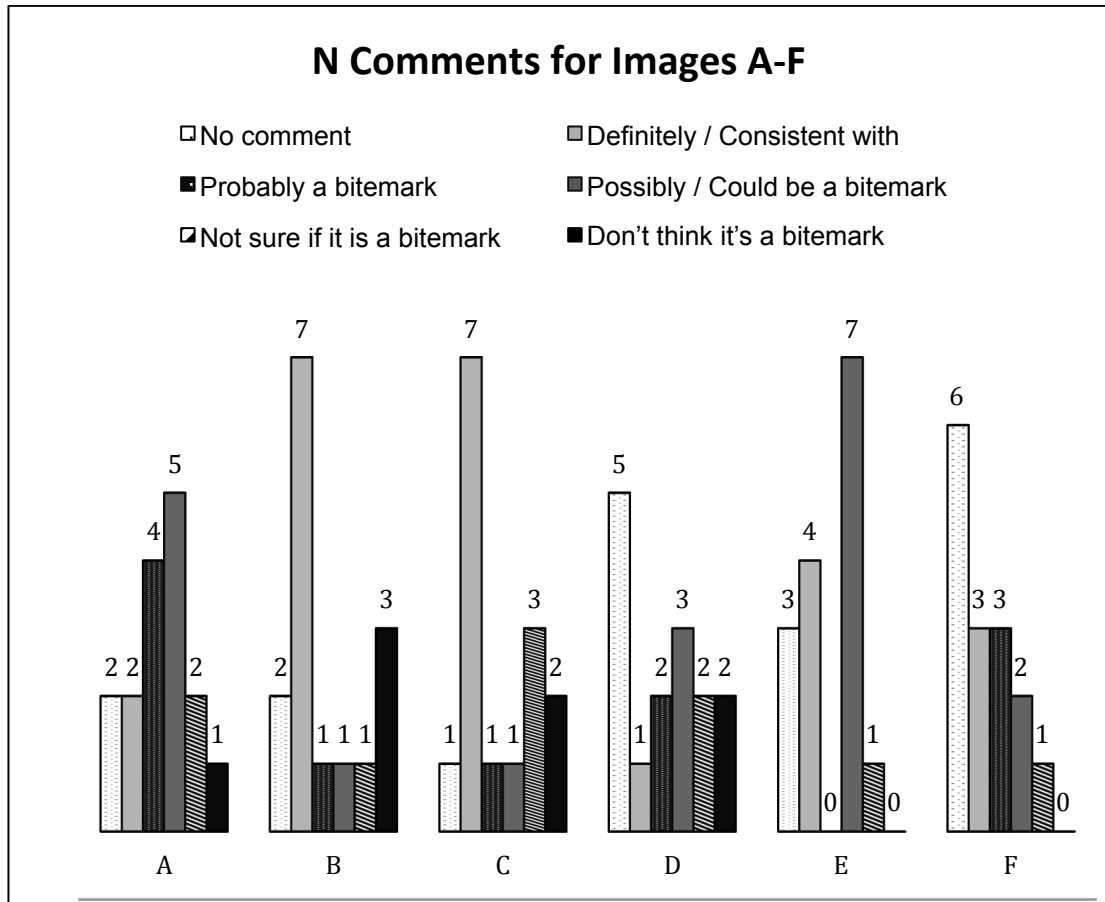


Figure 7.7 N comments regarding origin of injury

Opinions on the origin of the injury ranged from ‘definite’, or ‘consistent with a bitemark’ to ‘I don’t think it is a bitemark’. This represents the entire spectrum of possibilities regarding the origin of a single injury pattern. The only unrepresented category was that of ‘definitely *not* a bitemark’, and this could have been due to the nature of the interview process itself, which was knowingly about bitemarks and bitemark analysis, thus inducing a contextual bias. It is perhaps unsurprising that participants did not offer this definitive conclusion in the face of being offered images described as ‘bitemark photographs’ to view. There tended to be a general clustering of comments towards one end of the spectrum or the other for each bitemark, although the entire range of comments was usually encompassed for each image.

### Reporting – Conclusionary Language and Semantics

In total, 151 distinct conclusionary statements were noted for the 82 cases where the odontologists’ opinion was recorded in the bitemark case files reviewed. Comments were coded and clustered into groups that represented commonality in meaning (table 7.9) in line with standard grounded theory approaches. Comments were generally of two main types:

concerning the origin of the latent mark itself (origin), i.e. was this mark a bitemark?; and concerning the possibility that the mark was caused by a particular perpetrator (comparative).

	<b>Exemplar Comments – Origin</b>	<b>Exemplar Comments – Comparative</b>
<b>Strongly positive</b>	<i>Definite human bite Very likely a human bite</i>	<i>Injuries caused by perpetrator Consistent with the dentition of [the suspect] Strongly suggests [the suspect] Very likely caused by [the suspect]</i>
<b>Weakly positive</b>	<i>Consistent with a human bite Characteristics of a human bite</i>	<i>Concordance with a dentition noted Possibly source of the bite Not possible to exclude [the suspect]</i>
<b>Uncertain</b>	<i>Impossible to determine whether it is a bitemark or not</i>	<i>Not possible to confirm or exclude [the suspect]</i>
<b>Weakly negative</b>	<i>Possibly a human bite mark Probably not a human bitemark</i>	
<b>Strongly Negative</b>	<i>Not consistent with a human bite</i>	<i>Can exclude [the suspect]</i>

**Table 7.10** Categories of conclusionary remarks in odontology reports

The relative proportion of these conclusionary remarks is illustrated by figure 7.8. The most common remark made by odontologists in reports was that of the injury being ‘consistent with a human bitemark’ (n=27). The next most common remarks were that analysis ‘was not possible’ or ‘would be unlikely to yield relevant information’ (n=21), and that ‘it would not be possible to conclusively [or positively] identify’ an individual perpetrator (n=19). ‘Concordance [or similarity] with a dentition’ was explicitly mentioned in 15 cases.

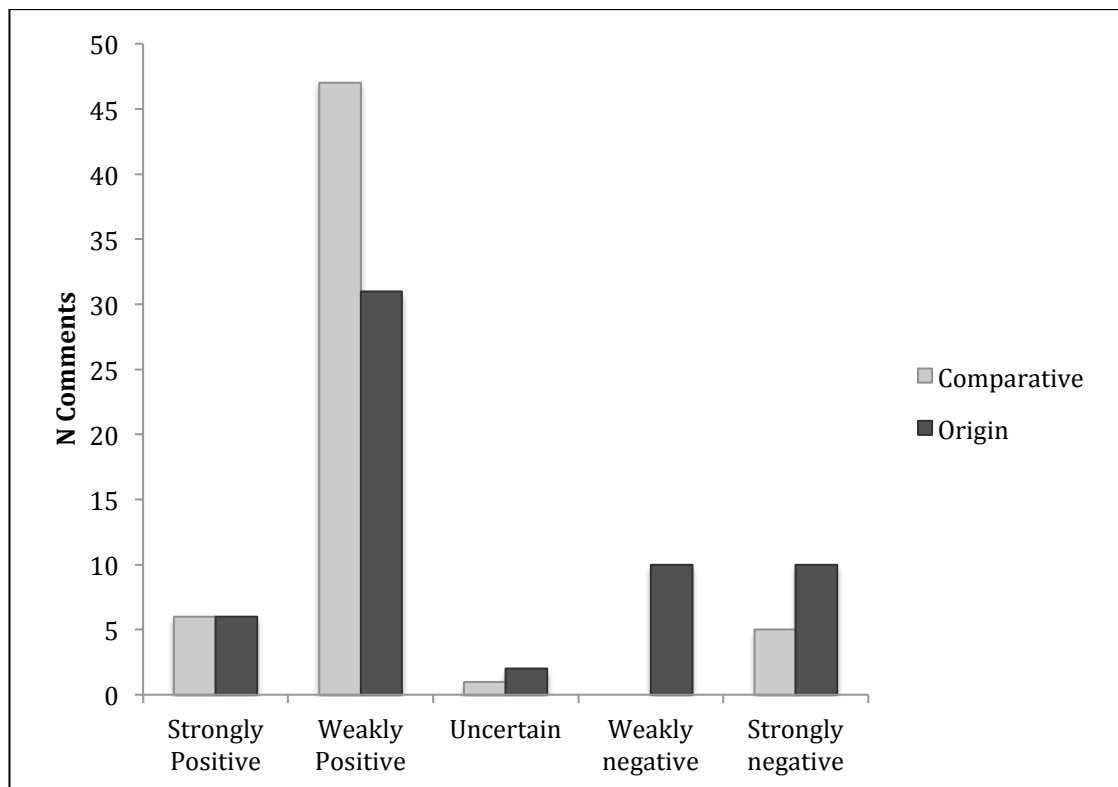


Figure 7.8 Incidence of conclusionary remarks in reports

Of the affirmatory conclusions, the strongest statement made was in an assault case, which concluded that the ‘injuries were caused by the perpetrator’. There was also one instance of the use of ‘strongly suggests the suspect’, and one of ‘consistent with the dentition of [the suspect]’. It was noted in 3 cases that it was ‘highly [or more than] likely’ that the suspect made the bite. More conservatively, the suspect was described as having ‘possibly’ made the bite in only 2 instances. More conservatively still, it was noted that ‘it would not be possible to exclude’ a suspect in 11 cases. Strong exculpatory conclusions occurred in 5 cases, where a ‘definite exclusion’ of a suspect was reported.

During the interview phase, participants were then asked to give definitions for conclusionary terms that they favoured using for bitemark analysis. Table 7.11 indicates selected responses from participants for each of the commonly used terms:

<b>Positive</b>	<ul style="list-style-type: none"> <li>• 100% certainty that it was the guy</li> <li>• greater than 95% chance that it was the assailant</li> <li>• greater than 90%</li> <li>• the mark could only have been left by the assailant</li> <li>• I can confirm that A [e.g. the mark] relates to B [e.g. the dentition]</li> <li>• there is a reasonable degree of certainty, say 1 to 3% chance that it wasn't him</li> </ul>
<b>Probable</b>	<ul style="list-style-type: none"> <li>• about 50%</li> <li>• a greater than 50% likelihood</li> <li>• an 80 to 100% chance that it is the perpetrator.</li> <li>• all the characteristics fit, nothing excludes the suspect</li> <li>• perhaps a 75-80% certainty</li> <li>• greater than 0% chance that you might be wrong</li> <li>• it might be the assailant, but you need more information</li> </ul>
<b>Likely</b>	<ul style="list-style-type: none"> <li>• greater than 50% likelihood</li> <li>• a 75% likelihood</li> <li>• couldn't be excluded</li> <li>• about 60%</li> <li>• more than 60%</li> <li>• about a 20 to 40% chance of being wrong</li> </ul>
<b>Possible</b>	<ul style="list-style-type: none"> <li>• not enough evidence to say</li> <li>• a 50/50 chance</li> <li>• the same as probable</li> <li>• couldn't rule out the suspect</li> <li>• about 50%</li> <li>• anything greater than 0</li> <li>• a 65-75% likelihood</li> </ul>
<b>More likely than not</b>	<ul style="list-style-type: none"> <li>• no obvious discrepancies</li> <li>• about a 90% chance of it being the perpetrator</li> <li>• a greater than 50% likelihood</li> </ul>
<b>Unlikely</b>	<ul style="list-style-type: none"> <li>• no chance of it being the suspect</li> <li>• less than 50% chance</li> <li>• about a 60% certainty that it isn't the suspect</li> <li>• about a 50 to 70% chance that it isn't the perpetrator</li> </ul>
<b>Exclude</b>	<ul style="list-style-type: none"> <li>• complete disparity between the bitemark and the suspect's dentition</li> <li>• a 0% chance that it is the biter</li> <li>• there are no concordant features</li> <li>• 100% certain that it is not the biter</li> <li>• anything less than 10% chance it could be the perpetrator</li> </ul>

Table 7.11 Meaning of conclusionary terms as described by odontologists

## Discussion

### Experience and Qualifications

Most practitioners engaged in odontology casework on at least a permanent part-time basis are of considerable experience and most have undergone a formal post-graduate course in forensic odontology. Those that have not have had formal training via the Graduate Diploma would be unlikely to be challenged in Australian courts given their considerable experience. Generally, the more experienced practitioners (those with 11+ years) expressed more uncertainty regarding the origin of the injury (i.e., appeared to give more 'conservative' opinions), yet the lesser-experienced practitioners (those with ten or less years experience) demonstrated better agreement with one another than those in the other experience ranges



(see table 7.11). This could be a reflection of the fact that they were more likely to rely on their training, rather than their experience.<sup>5</sup>

N Years Experience	Comments on Origin (%) v Years Experience					
	<i>No comment, but assumed to be a bitemark</i>	<i>Definitely or Consistent with a bitemark</i>	<i>Probably a bitemark</i>	<i>Possibly or could be a bitemark</i>	<i>Not sure if it is a bitemark</i>	<i>Don't think it is a bitemark</i>
<b>0-10</b>	29	38	17	17	0	0
<b>11-20</b>	13	29	8	17	25	8
<b>21+</b>	19	14	10	33	10	14

Table 7.12 Experience versus Comments on Likely Origin of Injury

Whittaker's study (Whittaker et al., 1998) suggested that training is more important than experience when it comes to determining the origin of bitemarks as either adult or child, and this premise likely holds true for deciding the more fundamental question of whether a particular injury represents a bitemark or not. Those with the most experience (21 or more years) tended to have the widest range of opinions as to whether the mark was of human dental origin or not, and thus demonstrated the poorest agreement within their experience category. This category also had the greatest variation in initial training (three of the four members without the Graduate Diploma qualification were in this group). These results support Whittaker's general notion that experience does not necessarily correlate with reliability, and also supports the notion that training is vital to ensuring inter-examiner reliability. These results also strengthen the notion that experience, by itself, is a poor substitute for training and research in the discipline, which are likely to yield far more reliable results.

## Quality

The quality of evidence presented to odontologists for analysis is generally poor. This includes material quality, such as poor photographs, as well as the nature of the mark itself. Some odontologists recognise that the onus is potentially on the odontology community to address the issues associated with poor material quality:

<sup>5</sup> Consider one participant's comments regarding the origin of the injury in image D, who relied entirely on their 'experience' to deduce that the injury was a bitemark. This comment was generally at odds with the majority of their experience- group.

*I think it is because of poor education, poor awareness. I think a police photographer is a good photographer, but I don't think they realise the nuances of what we require, even in general dental forensic photography. They don't have the mindset of a forensic dentist. Often they take photographs for the sake of documenting something, which is fine for that purpose, but it isn't a good forensic photograph, if that makes sense. (Participant #6487)*

*They [police photographers] are improving. And I think that is a matter of education through the odontologist. (Participant #1598)*

The corollary of this is that poor quality evidence should lead to greater uncertainty regarding the nature, origin or association of bitemarks. This is discussed further in the section on forensic value analysis below.

### **Methods of Analysis**

It is interesting to note that it was those odontologists with more experience who expressed a preference for acetate and hand-tracing methods of comparison, rather than digital methods. This may be due to older odontologists being less comfortable with newer technologies,<sup>6</sup> however this preference for acetate hand tracing is perhaps unfortunate in light of the research by Sweet and Bowers (1998) and McNamee and colleagues (McNamee et al., 2005). Furthermore, the digital production of overlays by those practitioners who used this method all involved scanning the models of the dentition, rather than of wax impressions of the incisal portions of the teeth. Bowers and Johansen (2003) initially recommend that wax bites be used for production of overlays and expressed that digital scanning of models to create a 'hollow' overlay is perhaps better only when a good quality bitemark (that demonstrates individual features of certain teeth) exists. It makes sense to generate models from wax impressions, bites or scrapings initially. The potential for observer bias is high in bitemark analysis, and the generation of an overlay from the incisal edges of the teeth themselves is fraught with the dangers of 'selection' bias. Even though there are tools available in programs such as Photoshop™ that can assist in minimising this, these are highly reliant on the positioning of the models on the platen, and the contrast then available in the resulting image.

Bowers and Johansen (2003) themselves noted that:

The objective of this selection process is to select the portions of the teeth most likely to have imprinted while biting. As mentioned, this can be a somewhat subjective process introducing the concept of intra-examiner reliability. Be sure that the selections you make are a fair

---

<sup>6</sup> See O'Connor (2010), noting his reference to what he calls the 'digital divide'.

representation of the biting surfaces. There are variables in this procedure that cannot be overcome at this time. These include different scanner hardware and software, cast position upon the platen, scan contrast, and chosen tolerance among others.

Consequently, when scanning a wax impression or bite, they recommend increasing the contrast between the wax and the impressions by filling the resultant marks with a material of a contrasting colour, such as plaster. This clearly defines the outlines of the edges of the teeth. An alternative would be to lightly dust the wax surface with a material such as talcum powder prior to imprinting the dentition, in order to make the non-contact areas matte. The resultant contact areas will have a higher degree of sheen that can be easily distinguished once the wax bite is scanned. Another advantage of using wax 'impressions' to generate an overlay of the dentition is that a variety of modes of injury can be demonstrated. The dentition can be scraped, to simulate drag marks, or placed at an angle if the bite does not appear to be centred.

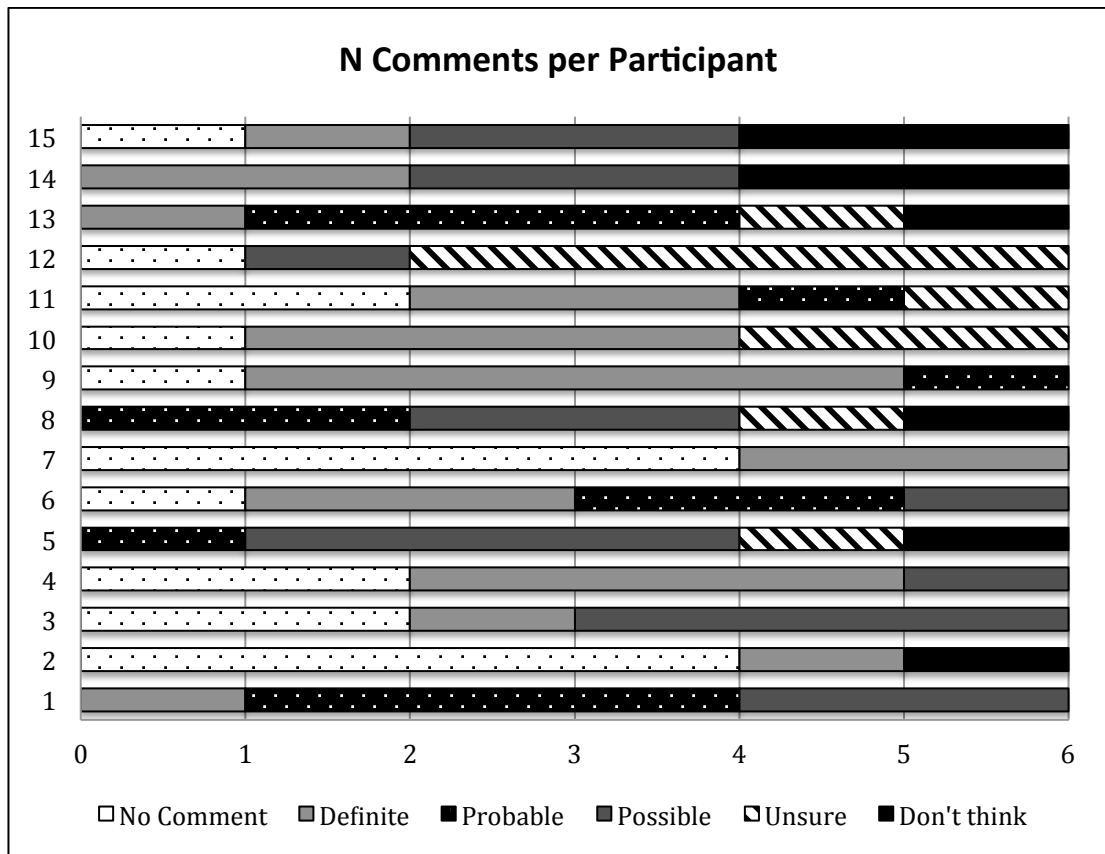
Using wax bites to generate overlays also fits with the philosophy of 'comparing wounds with wounds'. This is followed in other pattern-matching disciplines, particularly in that area which bitemark analysis perhaps has the most similarity with: toolmark analysis. Tool mark analysts rarely compare the surface of the tool to the suspected mark in the first instance. Standard protocol in this discipline is to create known marks from the suspect tool first, and then compare these known marks with the suspected mark (James and Nordby, 2009). This procedure is followed in other areas of pattern comparison too such as footwear impression analysis (comparing test impressions of the suspect shoe with the suspect mark), firearms analysis (using test bullets fired from the suspect gun, and not directly comparing the internal surface of the barrel with the suspect bullet), tyre mark analysis, and friction ridge (finger, hand and footprint) analysis.

While a number of practitioners expressed a preference for their 'own' methods of comparison, including creation of a contour map of the dentition by immersing the models in ink, and creating a clear acrylic model of the incisal edges which can be directly compared, there is currently no scientific evidence supporting the use of any of these techniques. The use of clear acrylic templates has been reported in the literature (McKenna et al., 1999) but only in the form of a case report, and has not yet been the subject of a study relating to its reliability either alone or compared to other methods. The ink-immersion technique rates a mention in a comprehensive textbook on bitemark analysis (Dorion, 2005), yet the author also cautions that there is yet to be any scientific proof of the validity of this method, noting that no inter or intra-examiner reliability study has yet been conducted.

## Odontologist Agreement

This qualitative data reveals that there is considerable variation amongst odontologists in even the most elementary aspects of the forensic diagnostic sieve, that of deciding whether the injury was indeed a bitemark or not. It is interesting to note that the greatest agreement for comments regarding the origin of an individual bitemark was for image E. This was also the image about which there was the most uncertainty as to its origin, with the highest number of practitioners stating that it might only ‘possibly’ or ‘could be’ a bitemark. Given that image E was also that which demonstrated the least degree of agreement between practitioners for assigning of the Pretty score ( $\kappa = 0.17$ ), this seemed to indicate that the more dubious the quality of the injury, the more likely it is that practitioners will agree that uncertainty prevails. The corollary of this is that odontologists seem to be more likely to agree with each other when there is uncertainty about the origin of the mark than when stronger opinions are expressed. This paradoxically suggests that practitioners are more likely to agree when they *don’t* know rather than when they think they do.

The obvious question arises as to whether some odontologists are simply more conservative than others, and are hence reluctant to offer definitive opinions. Could this possibly account for the variation seen in differences of opinion? The answer is, of course, yes. It appears that some participants do have a greater tendency to use stronger terms when characterising the origin of an injury as a bitemark or not. This can be demonstrated by visualising the relative number of comments for each odontologist as plotted in figure 7.9.



**Figure 7.9 — Number of comments per interviewee.** Each vertical bar represents one comment (E.g. Participant 1 made one 'definite' comment, three 'probable' comments, and two 'possible' comments). Total number of comments = 6 for each participant.

It appears that some participants do have a greater tendency to use stronger terms when characterising the origin of an injury as a bitemark or not. Compare Participant #14 (two instances of 'definite', two of 'possible' and two of 'I don't think it is bitemark') with Participant #5 (one instance of 'probable', three of 'possible', one of 'unsure' and one of 'I don't think it is a bitemark'). While this can be identified qualitatively, it is difficult to quantify these differences in any way from one practitioner to the next. That is, it is generally possible to say that one practitioner is more conservative than another, but not in any particularly measurable way – at least from this data. The data merely verifies the fact that there is a wide range of opinion over even the most basic assumptions in bitemark analysis: that of the origin of the mark itself.

Yet it is not as simple a matter as to say that some odontologists are more conservative than others, and this accounts for disagreement in bitemark casework. There was not necessarily a consistent relationship between 'conservative' and 'non-conservative' practitioners in the quality of their comments. For example, two participants were in complete agreement regarding the origin of the patterned injury in image A as being from human teeth, yet in complete disagreement regarding bitemark B, one claiming that they didn't think it was a

bitemark, and the other claiming that it probably was. Similarly, another two participants were in complete opposite agreement regarding images B and C;

Regarding image B:

*That could just be something that has been pressed on their body. I'm not convinced that they are bitemarks.* (Participant #1786)

*Yes these look like human bitemarks* (Participant #2159)

And yet regarding image C, the same two participants stated contrary opinions, this time from the reverse perspective:

*I would probably say that that could be a bitemark, with the abrasions, here, he has had a go twice. Its been pinched in. This here could be the upper teeth, on the right side of the photo. Yes I think so.* (Participant #1786)

*No I don't think that is a bitemark* (Participant #2159)

Consider also the variation in comments made regarding image A, with some participants having noted that it was likely to be a bitemark while others dismissed it as being unlikely. The potential reasons for this, aside from mere differences in opinion are interesting. It can be seen from the comments that there is often inconsistency in the reasoning for odontologist's decisions regarding this fundamental call. In the example above, image A was classified as a bitemark by one participant on the basis of its class characteristics, yet it was dismissed as a bitemark by another participant on the basis of its size. Clearly when odontologists are using different standards in order to determine the origin of an injury, differences in opinion will likely be the result.

Despite standard protocols existing in other pattern identification sciences that describe latent marks in terms of class and individual characteristics, this does not appear to be followed to any great extent by odontologists, and this is potentially a problem that can lead to this sort of inconsistency between odontologists. The basis for including or excluding a mark as having been made by human teeth seems to be variable, as evidenced in the comments made by practitioners for image A regarding the relative size and the presence of the area of central ecchymosis and what this potentially means. Few reports from the casework study referenced 'class characteristics' when assessing the origin of the bitemark, and even fewer participants mentioned the term during the interviews. Yet this is, according to odontology literature, the most fundamental analytical process in bitemark analysis, and thought should be given to standardising this approach.

Few participants commented on the orientation of the bitemark. Most were hesitant to make a call on which were upper and which were lower tooth marks. Several comments regarding the orientation of the mark were made for image C, yet one of the three participants indicated the opposite aspect of the injury as being caused by the upper teeth, compared to the other two. A similar instance occurred during participant comments for image D, where both odontologists who commented on the orientation of the mark gave conflicting statements as to which were the marks caused by upper and lower teeth. What may be of note is this situation is that all of these participants made comments regarding orientation based on their own opinion. None made specific mention as to why they had chosen a particular orientation over another. This practice is similar to stating ‘I think it is x because I said so’. While this may have been circumstance-driven, in that because the participants were talking to a dental-trained colleague, they may have assumed it was obvious as why one particular orientation was clearly favoured over another — but here the point is made: at least one of the three odontologists who commented on image C got the orientation wrong,<sup>7</sup> and one of the two odontologists who commented on image D was also wrong. Of five opinions offered on orientation of bitemark injuries, two were completely opposite to the other three. This situation has been encountered in Australian courts before, in *R v Carroll* (2000), where experts in the perjury trial of 2000 disagreed with the orientation of the bitemark assigned by those from the original trial in 1985. This very lack of consistency in opinion was cited as one of the primary reasons for declaring the odontology evidence unsafe to render a guilty verdict on by the appellate court (see *R v Carroll*, 2001).

Consider also the range of opinions regarding image B. While some practitioners felt very confident in their ability to interpret this injury, offering tremendous detail regarding the position of the centre line, position of the head while biting, and even the status of his central incisors, others in the same peer group (with approximately the same amount of experience and the same level of qualification) remained far more hesitant to even call the mark as being made by teeth. This demonstrates a fundamental issue when the same image is described so differently by practitioners with the same amount of training and experience, suggestive of the fact that the fundamentals of good ‘scientific’ practice regarding analysis of suspected bitemark images has potentially been forgotten.

The concept mentioned by one participant regarding the need to have five individual tooth marks in order to call it an arch mark (for Image E) appears to be grounded in a paper that

---

<sup>7</sup> It is also possible that two of the three participants were wrong and the one aberrant opinion was the correct one — it is impossible to deduce with certainty, as the ground-truth of the case was not known.

was first referenced by Jakobsen and Keiser-Nielsen (1981). The reference paper by Berg and Schaidt (1954), originally in German, proposed that 4 to 5 marks of adjacent teeth must be present before a given mark can be identified as a human arch mark, upper or lower. This was further refined by the referencing authors, who added that ‘a bite mark is not present until one upper and one lower arch mark have been isolated as parts of one and the same bite’ (Jakobsen and Keiser-Nielsen, 1981). While the participant considered this to be a useful guideline to ensure a conservative opinion on suspected bitemarks, it is interesting to note that the 1981 paper continues with ‘one arch mark is never a bite mark, but it may hold adequate detail to allow identification of the originator’ – a statement that appears contrary to the fundamental criteria (i.e. two arches) previously described as being needed to definitively identify the latent image as being a bitemark.

There were several comments regarding the nature of the central bruising or ecchymosis in images A and B. This was variously attributed to ‘suction’, ‘tongue thrusting’ or ‘compression of surface vessels’ during the biting process. Central ecchymosis within bitemarks was originally attributed to suction or tongue thrusting (Barbenel and Evans, 1974; Clark, 1992), yet this has been disputed since the 1980’s (Jakobsen and Keiser-Nielsen, 1981). Despite common acceptance today, the theory regarding compression of capillaries remains to this day a theory based on anecdote (see Dorion, 2005), and although more reasonable, likewise remains empirically untested. There has been no literature aside from the study conducted in 1974 regarding the ability of a sucking action, or tongue-thrusting to cause such an injury.<sup>8</sup> Despite the lack of experimentally derived data, it appears that most authors now support the proposition that these areas are caused by compression and subsequent rupture of sub-surface capillaries in the central area of tissue that is squeezed between the teeth during the biting process. Citation of other causes rests on tenuous and unproven grounds.<sup>9</sup>

Image D drew comments on the force of the bite. Most odontologists suggested that a greater bite force was demonstrated in image D than in the other examples. While this reasoning can be understood, recent research has demonstrated that laceration of skin may be related to complexities of tissue, victim and perpetrator movement, as well as sharpness among other

---

<sup>8</sup> And that study only involved measurement of sucking and thrusting forces, not production of any actual marks on tissue.

<sup>9</sup> Central ecchymosis cannot definitively be caused by suction or tongue thrusting as the phenomenon has been observed with the production of bitemarks on living tissue using only articulated models (see Dorion, 2005).



factors, rather than solely bite force (Bush et al., 2010b), hence definitive conclusions about bite force should be adopted with caution.<sup>10</sup>

Few odontologists offered opinions on the age of the suspected marks. Image E indicated a healing wound that was obvious to any clinician familiar with basic clinical pathology, and this was agreed to by all who commented. Image F showed a mark that consisted entirely of bruising. A systematic review of 167 papers on bruising concluded that ‘we cannot accurately age a bruise from clinical assessment in vivo or from a photograph. Any clinician who offers a definitive estimate of the age of a bruise in a child by assessment with the naked eye is doing so without adequate published evidence’ (Maguire et al., 2005). Despite the publication of ‘estimates’ of the age of bruising using colour as a guide (see Bowers, 2004), the same author published a review paper in 1997 claiming that bitemark ageing using bruises as a guide was highly unreliable, and that odontologists should limit themselves to describing bruises as ‘recent’ or ‘old’. Odontologists in Australia who commented on the bruising in image F did appear to adhere to these limitations, although the exact reasoning (i.e. the presence of green pigmentation for image F, as one participant noted) was perhaps questionable.

### **Severity and Significance Scoring: ‘Forensic Value’ Analysis**

Using only the text descriptions as a reference for assigning scores, only slight agreement among participants was noted, with a kappa score of  $\kappa=0.14$ . Despite the inter-examiner reliability improving significantly after using the pictorial exemplars as a reference, this still only resulted in an overall  $\kappa=0.33$ . This represents only fair agreement among odontologists over all six bitemarks. The main concern with the use of the scale developed by Pretty in aiding assessment of the forensic value of bitemark injuries is the ability of practitioners to reliably apply it to casework. Overall agreement varied according to the nature of the injury itself, and it appears that the scale is designed to be used for relatively recent bitemark injuries. Once older injuries are introduced, practitioners appear to lack certainty as to how to rate them in accordance with this scale.

Wounds that were intermediate in their severity (images B and C) achieved less agreement between participants in assigning a score (table 7.13). Using a kappa calculation after Fleiss (Fleiss, 1981), it can be seen that inter-examiner agreement was greatest for those images at the extremes of the scale; images A, D and F, —representative of the lowest and highest

---

<sup>10</sup> Noting here that there is a semantic difference between commenting on the relative force used to create a bitemark, and the relative violence of an attack. Comments on the latter may sometimes be justified; comments on the former are probably not.

severity injuries — had agreement ranging from ‘moderate’ (0.4–0.6) to ‘substantial’ (0.6–0.8). Image E had poor agreement due to the difficulty in assessing the wound due to its age – and a number of participants refused to assign it a score on this basis

Image	Relative severity	Kappa
A	Low	0.55
B	Moderate	0.28
C	Moderate	0.41
D	High	0.63
E	Uncertain	0.19
F	Low	0.73

**Table 7.13** Practitioner agreement regarding origin of injury

Practitioner agreement in assigning a severity and significance score was greatest when given a visual exemplar in addition to written descriptions of the categories. This suggests that the written descriptions may be ambiguous, or unclear when used by themselves, or that they do not necessarily correlate with the visual images. As one odontologist commented on the difficulty of applying this scale to actual casework:

*This concept that, for example, you can get very mild bruising with no individual tooth marks present, but you can have very obvious bruising with no individual tooth marks present either, I mean, that depends on timing. I also have an issue with the line ‘individual discrete areas are associated with teeth’ – I can’t be sure that they were caused by teeth, so technically I can’t assign this score. (Participant #6828)*

However, somewhat pleasingly, when considering written odontology reports in case files, there was a proportional relationship between the strength of the conclusion and the quality of the bitemark: generally speaking, low quality bitemarks attracted only weakly weighted conclusions regarding identity in written odontology reports (table 7.14).

<b>S&amp;S Score</b>	<b><i>N</i> Comments (Comparative)</b>					
	Strongly Positive	Weakly Positive	Uncertain	Weakly Negative	Strongly Negative	Total
<b>1</b>	1	12	1	0	4	<b>18</b>
<b>2</b>	1	18	0	0	0	<b>19</b>
<b>3</b>	1	8	0	0	1	<b>10</b>
<b>4</b>	2	4	0	0	0	<b>6</b>
<b>5</b>	0	2	0	0	1	<b>3</b>

**Table 7.14 Correlation between conclusions and Severity/Significance Scores**

During the interview phase, however, a number of practitioners offered startling inferences from images that were generally considered of poor quality (or of low significance) by others. In several examples, where more than one odontologist made comments regarding circumstances, or orientation of the bitemark, opinions were completely opposite to one another (consider the different opinions regarding the orientation of imaged D). Despite image F having the lowest severity and significance rating of all of the bitemark images, practitioner comments were generally weighted towards the positivistic end of certainty that the image was a bitemark (see figure 7.7). Pretty's severity and significance scale is a potentially useful tool, in that images that are considered of low significance should have appropriately weighted conclusions and comments. Those that rate low on the scale should potentially have endorsed limitations placed on the strength and type of any remarks made regarding the marks origin, orientation, relationship to a dentition and so on, so as to avoid situations where extreme conclusions are reached about dubious evidence. The fact that inter-examiner reliability in assigning severity and significance scores was relatively low implies that training and education is warranted in the use of this scoring system, and that a visual reference guide should be used when assessing these scores, at least until practitioners become more comfortable with the tool.

### **Bitemark Reporting**

Conclusions Australian odontology reports generally follow conservative lines. Four of the cases analysed during the casework study asked for a direct comparison of two or more dentitions to the suspect mark, and even in these cases, where there was technically a 'closed population' from which to draw inferences about, the odontologists involved were still

unwilling to positively identify suspects on this basis, preferring to definitively exclude rather than include. In only one of these comparative cases was a weakly positivistic conclusion reached, that of ‘the injuries were possibly the result of that from [Suspect A]’. Other conclusive statements in this case were either ‘not possible to exclude’, or ‘not possible to positively identify’. Only in one case was a suspect definitively excluded.

Yet it becomes obvious when reviewing the range of conclusionary remarks between odontologists that there is a fundamental concern with inter-examiner consistency. Despite the use of language that is common to most practising odontologists, the terminology appears to mean different things to different practitioners. It is generally apparent that ‘positive’ represents the most affirmative conclusion, with the terms ‘likely’ and ‘probable’ considered more affirmative than ‘possible’, however, their relative values differ greatly between participants.

Most preferred to express their meaning using a percentage figure, reflecting either the probability that the bitemark belonged to the suspect, or the relative certainty of their own conclusion. Dorion noted that ‘ultimately a comparison of the totality of the evidence will be made between the [suspected biter and the bitemark] and the expert will render an opinion as to the degree of probability that the suspect caused the bitemark in question’ (Dorion, 2005). Yet statements based on an estimated ‘degree of probability’ are nothing but *ipse dixit* until research has demonstrated the relationship between given bitemarks and the possible dentitions that may have caused them. Even descriptive terms such as ‘likely’, or ‘possible’ imply a probabilistic reasoning that is based simply on the opinion of the odontologist, and backed by little, if any, research-derived findings. Use of the term ‘consistent with’ was common by Australian odontologists, however, researchers have warned against use of this phrase, describing it as ‘nebulous’ (Evetts et al., 2000). They suggest that the phrase, by itself, is unbalanced unless the opposite proposition is also addressed, such as ‘the mark is also consistent with that made by any *other number* of objects or dentitions’. This clearly renders such testimony confusing, and does not assist in conveying the weight of the evidence clearly to the trier of fact.

The range of descriptive terms used by Australian odontologists to relate strength of association in this manner was wide, with many overlapping categories. These results suggest that describing a particular match as being ‘probable’ may mean the likelihood of the suspect being the perpetrator is anywhere from 50 to 100%, depending on what definition of ‘probable’ the odontologist considers he or she means. Consider the overlap at the negative

end of the spectrum of conclusions. Some practitioners equated the term ‘unlikely’<sup>11</sup> with ‘exclude’, defining ‘unlikely’ as *‘no chance of it being the suspect’*. Others considered the term ‘exclude’ to mean *‘anything less than a 10% chance it could be the perpetrator’*. Clearly there is an overlap here that is cause for concern when giving evidence – unless there is consistency in these definitions, how are the police, lawyers, judges, and juries supposed to interpret these terms?

Those who tended to express these terms using numerical definitions approached this task in either one of two ways. The first consisted of giving an approximate likelihood of guilt or innocence, as in *‘a 95% chance that it was the assailant’*. The second way these numerical definitions were offered was in terms of the percentage that *they were certain* the bitemark was caused by the perpetrator, for example: *‘I’m 100% certain that this wasn’t the biter’*. There are a number of fundamental issues with these forms of definition. Firstly, it is not the odontologist’s role to comment on the likelihood of guilt or innocence. Secondly, there have been no successful attempts at deducing random match probabilities when comparing bitemarks in human skin to dentitions. Given the nature of bitemarks, where it is obvious to most that faithful reproduction of the dentition on a medium is virtually impossible (unlike the comparison of ante-mortem and post-mortem radiographs), it is impossible to derive accurate measurements of the likelihood of the suspect being the perpetrator when so much information about the dentition is distorted or lost during the biting process. The fact that it has so far also been impossible to quantify this degree of distortion in numerical terms further adds to this difficulty, and thus numerical probabilities regarding the likelihood of a suspect creating a particular bitemark, compared to another random citizen, cannot possibly be regarded as accurate.

The second way of expressing definitions, with percentages of certainty, is also fundamentally flawed. Recent research demonstrates that the relationship between practitioner certainty, performance and ground truth is minimal. Shynkaruk and Thompson affirmed that there is little relationship between practitioner accuracy and confidence, (Shynkaruk and Thompson, 2006), and statements regarding confidence are poor substitutes for those concerning reliability and accuracy. Expression of ‘certainty’ is thus unhelpful for the trier of fact as a benchmark with which to assess the likelihood of guilt or innocence against. The more meaningful way to explain such terminology would be to express such numbers as a random match probability, yet until these numbers are derived from research statistics, offers by odontologists as ‘best guesses’ mean very little.

---

<sup>11</sup> Which the majority considered only a weakly negative term.

However, one must remember that the prime advantage of using descriptors to describe the relationship between a suspect dentition and a bitemark is that doing so specifically *avoids* having to give numerical or probabilistic statements. Given the paucity of research into the frequency of various patterns of dental arrangements, numerical statements could only be misleading at the current time. Yet it is in the acceptance of these descriptors the crux of the problem lies. These descriptors are readily accepted in the field of forensic dental identification (i.e. identification of the deceased) due to the readily accepted ability to identify human remains from dental evidence. Given forensic dental identification's status as one of the 'scientific' methods of identification (along with DNA and fingerprints) it is perhaps more justified that practitioners make the leap from making a statement regarding the degree of similarity between a pre and post-mortem radiograph to one of a statement of identity.<sup>12</sup>

There are potentially a number of reasons why this practice is so accepted in identification of the mass deceased: there is an ability for lay persons to visually assess such radiographs themselves as being in agreement without the need for complex spatial analysis; an ability to explain discrepancies between ante and post-mortem radiographs using accepted theories of dental disease progression and radiographic imaging; there is also a greatly reduced pool of potential individuals that could be a match;<sup>13</sup> and there is subsequently an ability to convincingly eliminate other suspected identifications from this set via the same means. Bitemark analysis enjoys none of these advantages. Statements regarding similarities between the dentition and a bitemark cannot justifiably be extrapolated to statements regarding identity. As expert witnesses, we have a different role in court to that which we are used to in disaster victim identification exercises. Our job is not to identify, it is to add weight to a particular theory of guilt or innocence that we play no part in deciding on ourselves.

One participant actually commented that the actual probability would '*depend on the other evidence presented at the trial*'. This is most certainly *not* how expert evidence should be tendered — experts are only permitted to opine in their designated area, and so should only

---

<sup>12</sup> It is recognized that this statement is contentious, but for the purposes of this thesis, is probably justified as a comparison to the far more contentious issues surrounding bitemark analysis. The Coroner is technically responsible for the final call on an identity, however this does not appear to have stopped the adoption of conclusionary definitions 'usurping' this authority world-wide in the field of fingerprints, DNA and dental evidence, with apparently minimal protest. Additionally, given the finality of the identification process, it is difficult to then assess the accuracy of such calls, as coronial cases are re-examined on a far less frequent basis than criminal or civil, yet despite these philosophical objections, it seems that society as a whole largely accepts these limitations. Such acceptance is a long way off for bitemark analysis.

<sup>13</sup> I.e. only those who are known to be deceased or missing in similar circumstances.

be commenting on the strength of their own particular evidence, which they are considered experts in. It is the judge or jury's responsibility to weigh the collective evidence and make an assessment of the suspect's likely guilt or innocence, not the expert witness's.

The AuSFO terminology used to describe the association between ante-mortem and post-mortem radiographs, in which these definitions are explained in terms of likelihood of identity, are fundamentally inappropriate for bitemark analysis. Similarly, the ABFO definitions of these terms, which relate the relative similarity between the bitemark and the suspect's dentition to the likelihood that the suspect was the 'biter', are also unjustifiable. If standard verbal descriptors are going to be used for reporting on bitemark analysis, a re-think of the use of these terms is necessary in order to avoid potentially misleading the judge or jury.

## Conclusion

Despite there being no concerns with the qualifications or experience of Australian odontologists, this analysis has revealed a number of fundamental inconsistencies regarding the practice of bitemark analysis in Australia that appear to render conclusions between practitioners highly variable. There are clear inconsistencies in method of analysis, reasoning, and terminology all ultimately lead to inconsistencies in opinions, and this is a fundamental failing of bitemark analysis practices in Australia that needs to be addressed.

Other forensic identification disciplines have long upheld the maintenance of standards through published guidelines, and this is one way in which these shortfalls can be addressed with relative immediacy. Guidelines for bitemark analysis will not address the more fundamental concerns regarding the ability of bitemarks to accurately reflect a given dentition, however, it is certainly one method by which we can achieve a rapid increase in inter-examiner reliability by establishing consistent practice methodology. Such guidelines also provide the basis for future research that will address these more fundamental claims, providing a consistent baseline practice from which these studies can then be designed.

Another point emphasised by this research is that there is a preponderance of 'low-grade' bitemark evidence being presented to Australian odontologists, and this presents us with a further concern, in that it is precisely the kind of material that is subject to over-interpretation and misreading by well-meaning practitioners. This issue is addressed in the following chapter, where we consider some of the psychological factors influencing decision making in bitemark analysis, and how we can potentially limit their effects

## CHAPTER 8

## Contextual Effects and Observer Bias In Forensic Odontology

---

The 2009 NAS Report noted that ‘the law’s greatest dilemma in its heavy reliance on forensic evidence concerns the question of whether and to what extent there is any *science* in any given forensic science discipline.’ This observation was further distilled into two independent issues; firstly, the extent to which any forensic discipline is based on reliable science, and secondly, the extent to which practitioners rely on human interpretation, which itself is subject to error, bias and variable performance standards (National Research Council, 2009). Having considered the question of whether bitemark analysis has any basis as a reliable ‘science’, and whether practitioners can reliably reach conclusions regarding bitemark evidence in chapters 5, 6 and 7, we now turn to the second aspect of this statement; the relative influence of human interpretation in bitemark analysis, and how this may affect the overall reliability of decisions made by odontologists.

So-called ‘context effects’ have only recently been the subject of study in forensic science. There is a growing body of literature in other scientific disciplines that suggests observer effects induced by extraneous or emotive information play a significant role in the outcome of decision making, and psychologists have long recognised the effects of bias introduced via the cognitive state of the subject (Risinger et al., 2002). Nordby noted that ‘the [forensic] expert’s role is to refine the context of observation based on expert understanding [...] and any implicit, hidden observational expectations influencing the supplied interpretations must also be examined’ (Nordby, 1992), yet a review of the literature in forensic science reveals that forensic scientists have not realistically progressed with solutions to these issues aside from attempting to willingly ignore these influences (Saks et al., 2003). This is despite the fact that these ‘individualisation’ sciences have already been described as being ‘top heavy with subjectivity and ambiguity, and therefore particularly susceptible to [...] bias’ (Cooley, 2003). Sensitivity to the problem of observer effects is integral to the modern scientific method, where scientists have realised that observer effects can distort findings and produce misleading results in ways that are difficult to correct for (Risinger et al., 2002). A further problem associated with ignoring observer effects in forensic science reporting is that potentially erroneous conclusions reached under the



influence of these effects can be remarkably resistant to *disconfirmation*, even in the face of compelling evidence to the contrary. Sources of these effects include common practices amongst forensic science practitioners, such as communication between the investigators and forensic examiner, cross-communication amongst examiners, and selective re-examination of evidence.

True objectivity has been described as a chimera in forensic analysis, given the nature of interpretation and source of its samples (Whitman and Koppl, 2010), however, the potential for decisions to be influenced by conscious or unconscious practitioner bias potentially robs the trier of fact of necessarily independent information. Forensic evidence is often taken as an independent verification of a guilty (or not guilty) hypothesis — if a forensic examiner reaches a conclusion that includes consideration of factors other than the evidence before them, their conclusions should not carry the independent weight that the trier of fact has assumed is inherent in such testimony. For example, if a fingerprint examiner is aware that a particular latent print belongs to a person whose license plate closely matches the description given by a victim of some crime, but the print is somewhat ambiguous, psychological theory suggests they may unconsciously resolve the ambiguity in favour of calling a match. Despite the actual circumstances, to the jury it appears that the fingerprint examiner has reached his conclusion independent of any other information, and the presence of the fingerprint and existence of a closely matching license plate represents two independent coincidences that together, significantly strengthen the hypothesis of guilt. In fact, these conclusions are anything but independent and the weight of the fingerprint examiner's evidence is far less than it appears. This has been described as the most corrosive aspect of cognitive bias in forensic testimony, as each piece of evidence needs to be considered independent of the other in order for the trier of fact to effectively summate them, and arrive at a realistic probability of guilt or innocence (Risinger et al., 2002).

## **The Bias Blind Spot**

The ignorance of contextual effects is a recognised psychological phenomenon in itself, and has been termed the 'bias blind spot' (Pronin and Kugler, 2007). While forensic science is becoming aware of the potential for practitioner bias, the few articles that have been published by forensic scientists have potentially misunderstood several key concepts regarding the influence of contextual effects. One author, among several suggestions, recommends that the forensic practitioner 'accept bias, remain objective, and limit overconfidence' (Byrd, 2006). These recommendations are admirable, but unachievable in practice and demonstrate a failure to realise that contextual effects occur at a subconscious

level, and therefore cannot be avoided by simply adopting an open mind. Proficiency testing has also been recommended as a method to avoid confirmation bias (Budowle et al., 2009) but this also fails to directly address the issue of cognitive bias present at the time of actual analysis (Krane et al., 2004). Some forensic scientists claim that the notion that subjectivity can affect reliability is refuted via anecdotal evidence (Budowle et al., 2009). Anecdotal evidence is unfortunately a very weak form of proof when such statements are refuted by carefully controlled, blinded studies (Dror and Charlton, 2006; Dror et al., 2005), and Krane and colleagues (Krane et al., 2009) take specific issue with the assumption that bias is not a significant issue in the forensic sciences when there is no experimentally derived data to support these claims. There is, albeit equally weak, anecdotal evidence that suggests the *opposite* of that claimed by Budowle and colleagues, suggesting that expert evidence is susceptible to distortion, whether intentional or not, in order to fit pre-conceived misassumptions.<sup>1</sup>

A more recent example of such evidence is provided by that of the Brandon Mayfield case in the United States, who was arrested as the suspect responsible for the Madrid bombings in 2004 on the basis of an incorrectly matched fingerprint.<sup>2</sup> The US Department of Justice report into the matter noted that one of the main contributing factors to the misidentification of the Madrid Bomber was ‘circular reasoning’, and in their explanation of this incident, neatly describe several incidences of unconscious practitioner bias that were directly attributable to context effects (US Department of Justice, 2006). Other authors have determined that cognitive bias has played a key role in numerous other forensic misidentifications occurring in the twentieth and twenty-first century (Cole, 2006).

From interviews conducted with practising odontologists around Australia, it is apparent that a significant proportion of them refute the concept that bias is influential in any way on their practice. One commented:

*I don't think there really is [any potential for bias] ... I think that you take each case as you see it, as the evidence is given to you, and you have to judge on the quality of the evidence. Other than that, I don't think you are really biased at all. As for unconscious bias, well I don't think the actual analysis phase is really affected. I don't think your opinion would vary considerably. (Participant #1598)*

Another stated:

---

<sup>1</sup> One only needs to consider the *Chamberlain* (1984) and *Splatt* (See Shannon, 1984) cases in Australia by way of illustration.

<sup>2</sup> This fingerprint match was supposedly ‘verified’ by several experienced fingerprint examiners.

*I mean if you are following a particular process, then I don't think cognitive bias should be an influence. I don't think the emotional aspect affects me. Even on an unconscious level, only because I think I have been doing it so long, so I think I am reasonably subjective. I think if you were too objective, you wouldn't be doing it for more than a year or two. (Participant #9037)*

Others admitted the potential for bias existed, but did not believe that it affected them:

*[Regarding 'internal' bias]...I don't think it does [affect me] because I take everything with a grain of salt. (Participant #1660)*

*Yes, I think there is potential for [bias]. I think that female on female bias, for example if the victim is a female and I am female, then you get an empathy... I mean you shouldn't. If there are drugs or alcohol involved you sometimes think ... well.... Maybe they were asking for it you know, you shouldn't, but you do. But it doesn't affect my judgement. (Participant #8547)*

*If anything, I think I do have a bias, and that is to go somewhat the other way – to be sceptical rather than to be open. I think I need to convince myself more than others. (Participant #5195)*

Yet not all odontologists interviewed claimed they were immune to the effects of bias:

*I've always thought there was potential for bias ... I think that people need to be aware of the possibility of bias, that's the first thing to begin with. (Participant #7249)*

*Yes certainly [there is potential for bias], that's why I think you can't really do an identification based on bite marks. I think it's a very imprecise science. There is pressure for wanting to satisfy the 'client', pressure from the solicitors, the feeling that you are capable of getting a result, and ideas of your own that you can solve anything. (Participant #2737)*

These comments also demonstrate another problem with cognitive bias in Australian practicing odontologists, in that a reasonable proportion of them do not genuinely understand the terms 'cognitive bias' or 'observer effects'. Most commonly, it was thought that these terms referred to bias induced by police or prosecution pressure to arrive at a particular conclusion:

*Yes certainly [there is potential for bias]. Police say to us that we have a person who has been 'bitten', and I have had several cases where it hasn't been consistent with a bite mark – so there is some sort of external pressure from police certainly. (Participant #1660)*

*Cognitive bias? Well, yes I think we get too much information from the police. (Participant #7744)*

While this certainly plays a large role in the potential for bias, this is more accurately described as motivational bias. So-called role and conformity effects introduce motivational bias by virtue of one's perceived role and the desire to conform to the belief and perceptions of others. The institutional context of forensic work and its close association with law enforcement agencies has received criticism due to the creation of motivated reasoning and a coalitional alliance towards serving a common goal (Koppl, 2005), but this is a separate (yet inter-related) phenomenon to the issue of cognitive bias. Nonetheless, motivational bias also plays a part in influencing practitioner performance, as recognised by other interviewees:

*There is always pressure for wanting to satisfy the 'client', pressure from the solicitors... Perhaps you can be trying to impress somebody else, such as one of your colleagues. (Participant #2737)*

*I find there is always an unconscious coercion to do the right thing. There is an innate wish to catch the perpetrator. (Participant #6828)*

Cognitive bias specifically refers to the psychological sway towards one opinion versus another as a result of having information extraneous to the task at hand.<sup>3</sup> Context effects — psychological influences on decision making induced by knowledge of circumstantial information extraneous to the immediate task — most obviously give rise to motivational bias, which may be conscious or unconscious, however, they may also give rise to cognitive bias, particularly when there is ambiguity in the choice between two alternative hypotheses (Giannelli, 2007b). This latter form of bias is easily over-ridden when the evidence presents an obvious choice between two hypotheses, but becomes problematic when evidence is 'borderline', of poor quality, or ambiguous. Bitemark analysis is particularly susceptible to both manifestations of bias due to the context in which it is collected and analysed, which is rich in subliminal information that renders the practitioner susceptible to motivational bias, and; the general nature of the evidence itself, which is often of poor quality, ambiguous and has potential for misinterpretation.

---

<sup>3</sup> In other words, bias induced by 'knowing' something.

## Context Effects

### Emotional effects

The contextual details surrounding analysis of bitemarks are often highly emotive. Typically, the process of bitemark analysis involves interaction with a human victim, usually one of a violent crime such as rape, assault, or homicide; and collection of the evidence is usually performed, or at least assisted, by the odontologist, in order to ensure accurate documentation of the physical evidence. In the process, the practitioner meets or deals with the victim (either living or deceased), which potentially induces a flood of emotional cognitive input, particularly when the case involves other significant trauma or injury. Dror and colleagues (Dror et al., 2005) have already demonstrated that these emotive influences have a significant effect on forensic decision-making. To suggest that forensic odontologists are somehow immune to emotive influences sits at odds with other evidence, such as the higher incidence of post-traumatic stress disorder in forensic workers exposed to death and the dead (McCarroll et al., 1996, 2002; Ursano et al., 1999). It has also been suggested that the presentation of evidence in a suggestive way, such as the labelling of evidence as ‘defendant’ or ‘victim’, feed the examiner unnecessary and potentially biasing information (Jonakait, 1991).

### Confirmation bias

The close relationship that forensic practitioners engender with law enforcement agencies renders them susceptible to cognitive bias through the wider problem of information sharing. Practitioners may selectively use this external information, either consciously or unconsciously garnered from their associates, to assist them in reaching their conclusions.<sup>4</sup> This form of bias, known as confirmation bias, is well studied phenomenon in eyewitness line-ups, where witnesses who are initially tentative with their identifications become more definitive after learning that the person they identified is the prime suspect according to the police (Giannelli, 2007b). In recognition of this particular aspect of contextual bias, one odontologist commented:

*...I mean, there is immediately an inference [as to the likelihood of guilt] in that the suspect has been arrested for starters.* (Participant #2159)

Confirmation bias has played a role in numerous forensic scandals, and was recently acknowledged as one of the leading causes of the misidentification of the 2004 Madrid bomber (US Department of Justice, 2006). It has been claimed that experts, particularly

---

<sup>4</sup> This of course, usually occurs on an unconscious level.

those with experience, are less vulnerable to confirmation bias, however Dror and colleagues (Dror et al., 2006) have provided evidence to refute this claim. In their study, five fingerprint experts were given two prints each from casework archives that they had individually identified approximately five years earlier as a definitive match. These pairs were also blindly verified by two independent latent print examiners as being ‘matches’ prior to the study commencing. After being told by one of their colleagues that they were the same pair of prints that were used to erroneously identify the Madrid bomber (but subsequently told to then ignore this information and concentrate only on the print before them) the five subjects were asked to decide whether the two fingerprints matched. This time, only one of the five participants identified the prints as a match. Three changed their opinions to ‘no match’ and the other decided that there was insufficient information to make a definite decision. Other studies using less emotive contextual influences verified that even low-level extraneous information affects ‘match’ versus ‘non-match’ decisions (Dror and Charlton, 2006; Lagenburg et al., 2009).

Confirmation bias is not only of concern to forensic practice, it may also manifest in research where the testing of a hypothesis is carried out by searching for confirmatory instances, rather than potentially falsifying ones (Ask and Granhag, 2005), in a return to inductive methods that have been long superseded.

### **Target shifting**

Despite its reputation as the ‘gold standard’ forensic science, DNA analysis has also been the subject of criticism regarding subjective interpretation and confirmation bias. The use of low copy number analysis, partial samples and mixtures in order to obtain a DNA profile suggests that the incidence of ambiguity and subsequent interpretation in DNA casework occurs in more than a trivial fraction of cases (Whitman and Koppl, 2010). The existence of ambiguity regarding which peaks belong to which donor, in addition to the problems of allelic drop-out (and drop-in) often require the analyst to make a judgement call on the significance of electropherogram peaks. If the analyst has prior knowledge of a suspect’s profile, as commonly occurs in many laboratories, then they may be more inclined to include some ambiguous readings and dismiss others by claiming them as artefacts. This ‘target shifting’ naturally occurs in favour of supporting the prosecution theory, as the profile used for comparison is usually that of the defendant (Thompson, 2009). This phenomenon can also be seen in forensic odontology. The examining odontologist is usually presented with a bitemark that is rarely analysed independently from knowledge of the suspect’s dentition. Such procedures can lead to selective, confirmatory hypotheses akin to painting the target

around an arrow. In *R v Carroll* (1984), the prosecuting odontologist noted that this was an essential requirement for an unbiased assessment of the bitemark.

[Counsel] Now when you said, just a moment ago, [‘]remember that we are looking at this point at the injury[’], the point is, is it, by this stage you haven’t looked at the casts of any teeth of the suspect? — Oh, that’s very important that you do not do that at this stage. (Trial transcript)

Bush and colleagues have stated that examination of the latent mark independently of the dentition is not a suitable approach for bitemarks (Bush et al., 2010a), noting that a feature that was noted during examination of the dentition may not be accurately represented in skin or vice versa. Stipulating in advance what may or may not be present can lead to possible bias, inaccurate scenario account, or dental perpetrator misdirection.’ However, this approach fails to account for the nature of bias, specifically the concepts of target shifting and interpretation. While they noted that features of the dentition such as a missing tooth sometimes ‘appeared’ in the bites they created, Bush and colleagues’ study was carried out on cadaver models and involved the interpretation of indentations in the skin, rather than bruising. It is not certain that such distortion would occur if the teeth caused bruising or lacerations on live skin, which has an entirely different consistency to that of cadaveric skin. The NAS report, discussing this issue in relation to toolmarks, notes that the *a priori* stipulation of what features may or may not be considered suitable for analysis might not be possible, and hence examination of the tool itself in question might be warranted. However, they note that this is in contrast to DNA techniques where the examiner is initially ‘blinded’ to the reference sample, and that this represents a more ‘scientifically justified series of steps that lead to results with well-characterized confidence limits, [which] is the goal for all the methods of forensic science.’

## Other Sources of Cognitive Bias

### Observer Effects

Experimenter or observer effects involve the unintentional transfer of behaviour to subjects of the experiment via the researcher’s expectancy. The Hawthorne effect describes a related phenomenon that results in subjects performing better, or more deliberately when they know they are being studied (Holden, 2001; Wickstrom and Bendix, 2000). Of interest in many inter-observer odontology studies is that dental students often out-perform general dental practitioners, and in some instances, forensic odontologists themselves (McNamee and Sweet, 2003; Whittaker et al., 1998). This finding is often reported in studies, but little discussion of its significance follows, and is perhaps a manifestation of the Hawthorne effect

in a population ‘conditioned’ to perform well. While obviously being applicable to research, experimenter effects can exert influence in the work environment of the odontologist, where the police, managers and colleagues could be considered the ‘researchers’, and the odontologist the ‘subject’. This effect was recognised by at least one interviewee:

*The police are often [...] convinced that they have got the right person, but it is transmitted, almost by osmosis to you. (Participant #6828)*

In conjunction with role and conformity effects, the Hawthorne effect further increases the likelihood of confirmation bias. Acknowledgement of the existence of observer effects in research has given rise to research methods that have attempted to minimise them. While it is acknowledged these ‘ideal’ experimental conditions are difficult to achieve in forensic science, a few such studies have been attempted, and none of their results have suggested that forensic practitioners are immune to observer effects. A study conducted recently on the effects of emotional information on fingerprint analysis (Hall and Player, 2008) concluded that fingerprint examiners were not *particularly* susceptible to emotional bias; however, all of the participants in this study knew that the information they were given was part of a mock case. The actual emotional effect experienced by each examiner is thus difficult to assess. The design of this study has been criticised on a number of points, and it is suggested that the conclusion reached by the authors is not supported by the data they obtained (Dror, 2009). Perhaps of more concern in this study was that conclusions from all 70 practitioners who examined the same print varied across the entire allowable spectrum: ‘positive identification’, ‘some detail in agreement but not sufficient to identify’, ‘not suitable for comparison’, and ‘exclusion’. This reveals more important information about the overall reliability of fingerprint examination than it does about the effect of emotional information on practitioner performance.

### **The Contrast Effect**

It is rare that forensic examiners are presented with perfectly preserved evidence. Similarly, forensic odontologists rarely have the opportunity to analyse the perfectly created and preserved, ‘ideal’, bitemark. More often than not, the evidence is of poor quality, and therefore is open to interpretation via a number of alternative hypotheses. One effect that becomes particularly significant in this circumstance is the contrast effect. This phenomenon describes the tendency to shift the judgement standard after repeated exposure to stimuli of a certain threshold, and is particularly inherent in subjective comparison work, such as that performed by forensic odontologists (Saks et al., 2003). The susceptibility to contrast effects is demonstrated when the odontologist gradually begins to ‘see’ the



association between the mark and the dentition after lengthy analysis. The fact that such analysis is done in conjunction with a reference such as the suspect dentition also introduces bias via a ‘target-shifting’ mechanism, and these two forms of bias then act synergistically.

The contrast effect is a classic example of one of the most dangerous types of context effect in bitemark analysis. It has been suggested that mitigation of this effect can be achieved by the use of ‘dental line-ups’, where more than one model is offered for analysis by the odontologist. This helps avoid trying to ‘make things fit’ by offering different thresholds of association to compare with one another. While this is rarely done in practice today, it has been the subject of recommendation for some time in bitemark analysis (See Sweet, 2001, in Williams and Hahn, 2002). A very similar issue was raised in the 1985 *Carroll* trial:

[Counsel] It is not the case then that you were asked to pick out a cast of teeth from a cast of hundreds, as it were? — [Witness] That is correct.

You were given two casts of teeth, both apparently of the same person [...] So the choice presented to you was severely narrowed, wasn’t it? — In these circumstances it was just the one, in that I had to examine the dentition against the photographs of the bitemarks.

[Counsel] [...] but in making your comparisons, you only had one individual’s set of dentition? — Yes.

[...] So it is not the case that you were presented with the photographs, asked to examine the photographs, make a description of the likely features of the dentition which may have caused the marks, and having done that exercise, then to be presented with a model of teeth and to check what you had postulated from examination of the photographs against the features apparent on the set of teeth presented; that did not occur, did it? — The postulation I made was to look at the photographs and decide that it was a human bitemark, but I usually work the other way around. I usually examine the models and then in relation to the photographs [...]

One can readily make mistakes also by assuming the likelihood of the teeth fitting the photograph? — That is way one takes a great deal of care and trouble in making comparison, why you do not proceed to the next point until you have satisfied yourself on the one. (Trial transcript, p538)

Another odontologist was also questioned in this regard:

[Counsel] [...] I will put it to you this way: you had in fact given to you a set of teeth, a number of photographs, then it was a matter of theorising whether the teeth shown in the cast could have made the bruising appearing in the photograph? — [Witness] It was a procedure to consider whether the teeth did make those marks.

So it was a matter of not trying to identify what type or range of teeth may have made the marks, but it was a question of trying to find a theory to explain how those teeth as shown in the cast could have made the marks in the photographs? — It was a process to determine whether the teeth that I was – the cast that I was provided with had made those marks. (Trial transcript, p754)

It appears that perhaps neither witness fully understood the implications of what defence counsel had put to them. The first witness fails to directly answer counsel's lengthy question regarding the fact that an unbiased approach to analysing the suspect's dentition did *not* occur. The second witness also fails to address the issue that defence counsel raises. Defence counsel later broaches the issue with the first witness from a slightly different perspective:

[Counsel] As part of the process of forming an opinion, would you agree with the procedure should be that the investigator should collect sufficient information about bite marks and the teeth of a suspect, he must thoroughly describe both before beginning the actual comparison? — [Witness] Yes one can do that from the models of the suspect.

But before the comparison is done, there should be done by the expert a detailed examination and a thorough description of points concerned in the photograph and points concerned with the model of the suspect's dentition before the comparison is made? — As we set out as a routine, yes.

Here, the witness agrees with the proposition that a thorough analysis of the bitemark and the models should proceed separately, and prior to the actual comparison. Whether this actually occurred is a matter that the record does not make clear.

Several interviewees commented on the bias nature of certain analytical processes that can be seen to induce the contrast effect in people other than the forensic examiners, namely juries. Notably, criticism was drawn to one of the techniques used in *Carroll* (2001) whereby digital images of a bitemark and a dentition were superimposed on one another, and faded in and out in order to highlight the similarities between the two.

*I think people can be influenced by the process of 'fading in and out'. For example what was done with the Carroll case. That was completely suggestive in my opinion.* (Participant #1598)

In the 2001 *Carroll* trial, the experts used a computer simulation to fade images of Carroll's teeth in and out, over the photograph of the bitemark, in order to demonstrate its similarity. This process, where images of the teeth were superimposed over the bitemark was shown to the jury on at least forty separate occasions during the odontology evidence. Many of these occasions did not call for the superimposition to be shown, however, when counsel or the witness requested to bring up an image of the bitemark for demonstration of a particular point, it was usually presented with the superimposed cast of the dentition, or the wax 'scraping', over the mark itself. For example, during the evidence of one of the prosecution experts, the following occurred:

[Counsel] Can we go back to the photograph of bitemark 1, please?

[Snr Sgt Garner] Do you want the upper teeth in this one?

[Counsel] Yes please. And then just take away the teeth please. (*Trial transcript p754*)

Another such example during Dr Bamford's evidence:

[Counsel] Now could we now go, please, to [photograph] 7E, senior sergeant. And could we loose [sic] the teeth, and also the numbers, please? (*Trial transcript p704*)

And again during Dr Whittaker's evidence:

[Counsel] Can we go to bite mark number 1, please senior sergeant?

[Snr Sgt Garner] The upper?

[Counsel] Yes please. Take away the teeth please. (*Trial transcript p923*)

Similar instances abound during the evidence of all four expert witnesses at the 2001 trial. This may not have been intentional on the part of the prosecution, however, given that as few as eight or perhaps ten occasions actually called for the superimposition to be shown, the induction of the contrast effect on the jury cannot be easily dismissed. The danger with the contrast effect is that both odontologists, witnesses and juries become susceptible to seeing associations that simply are not there, on repeated exposure to evidence where associations may be borderline similar but not necessarily representative of 'matches'.

### **The Overconfidence Effect**

Practitioners have been generally found to be overconfident in their ability to perform, particularly when performing routine or often-repeated tasks. As one odontologist admitted:

*It isn't uncommon to get the feeling that you are capable of getting a result, and then ideas of your own that you can solve anything.* (Participant #2737)

It has been well established in the literature that there is only a very weak link between confidence and accuracy (Krug, 2007; Risinger et al., 2002) and that one is no substitute for the other. Despite this, it has been suggested that the overconfidence effect is related more to tasks involving vocabulary and general knowledge, and that tasks involving perception and sensory information (such as those the forensic odontologist would be involved in) are subject to an *under*-confidence effect, however, recent studies have refuted this notion (Pallier et al., 2002). Overconfidence of the expert carries with it the sequelae of an unconscious biasing effect on juries and judges, who despite claims of impartiality, are still encouraged to include assessment of witness demeanour as part of the process of assessment of the expert's evidence (Porter and Parker, 2001). It is human nature that people are more easily convinced by confident witnesses, however, experts should be careful not to overstate

their claims, and in particular, not to confuse the concepts of confidence and accuracy. Consider the following exchange that occurred between the chief odontology witness for the prosecution and prosecuting counsel in the 1985 *Carroll* trial:

[Counsel] In your opinion, is there a possibility that someone else's dentition could have made that bitemark that we see depicted on the thigh of Diedre Maree Kennedy? — No. I have done so much work on this and I have convinced myself quite clearly and it is my opinion that the only person who could have made the bitemark was Raymond John Carroll. (Trial transcript, p412)

And in the 2001 trial a similar comment was elicited from the chief witness:

[Counsel] ... the conclusion that you are expressing is that Mr Carroll is the only person on the planet who could have left those marks? — [Witness] I believe the conclusion that I am expressing is that the likelihood of Mr Carroll not having left those marks is so huge that it is inconceivable that it was not him. (Trial transcript, p641)

Also:

[Counsel] Your evidence is, in effect, that Mr Carroll is the only person in the world who could have left bitemark number 1? — [Witness] No, I didn't say that. I said there was a very, very, very high probability indeed that he was the perpetrator, but I would not exclude the possibility that somewhere in the world there is one or maybe more other people who could conceivably have done this.

I see? — I must say the chances are extremely remote. (Trial transcript, p658)

And:

It's a human bitemark, in my opinion, and I think the technique that the officers have used and demonstrated can establish beyond doubt that the dentition of using, or having in possession, they can superimpose onto those marks. (Trial transcript p696)

And again later:

[Counsel] Now, the bitemark that we see up there, could the teeth shown in that [other] person have caused that mark? — [Witness] No, its not possible.

[...]

That's your opinion, is it, it is not possible? — Absolutely not possible (Trial transcript p703)

It is easy to see how juries can be persuaded by such oracular pronouncements, yet Shynkaruk and Thompson affirmed that there is a striking dissociation of practitioner accuracy and confidence, particularly in deductive reasoning (Shynkaruk and Thompson, 2006). Statements regarding confidence are consequently poor substitutes for those concerning reliability and accuracy. Unfortunately, this principle appears to be poorly understood in forensic science: a recent US Department of Justice report noted that FBI fingerprint

examiners were routinely over-confident in their ability to declare matches based on latent evidence, yet goes on to explain that examiners are required to have 100% certainty in their identification conclusions (US Department of Justice, 2006). The two statements are clearly at odds with one another, and serve as an example of the recognition yet simultaneous dismissal of well-founded criticism.

## Minimising Cognitive Bias

Forensic odontology practice is littered with opportunities for the induction of motivational and cognitive bias. Odontologists need to be aware of the potential for contextual effects in order to develop systems that minimise their influence. Cognitive bias cannot be ‘willed away’, as many forensic practitioners would insist is possible, because by its very nature it is not under the conscious control of the individual (Loftus and Cole, 2004). It was recognised by most odontologists that the phenomenon itself referred to unconscious mechanisms:

*I’m sure there is bias. You try not to .... We are all humans. I don’t think it is on a conscious level, with someone who is a reasonably experienced practitioner.* (Participant #7239)

And yet, most agreed that simple self-awareness wasn’t sufficient to counter these effects:

*I don’t think the practitioner can really attempt to minimise [bias] him or herself at all. There should be protocols across the industry and the law that attempt to minimise bias.* (Participant #2159)

Several organisations and academics have called for the introduction of procedures and protocols in order to minimise observer effects in forensic identification science. The Justice Project issued a report that called for the introduction of ‘protocols to minimise and regulate the flow of information as to reduce inadvertent bias’ (The Justice Project, 2008). Budowle et al (2009) suggest ‘keeping a log of inclusions, exclusions and inconclusive results to support the proposition that forensic scientists are not overly biased and do provide substantial testing that can benefit either accused individuals or the government’, however this fails to address the issue of observer effects. This rather defensive approach fails to appreciate that firstly, judicial casework cannot validate the error rate of any particular technique as the ground truth is never known with certainty. Secondly, it is assumed that establishment of a ratio of inclusions versus exclusions would provide some proof that forensic science is not biased towards the prosecution or the government, given the large number of expected ‘exclusions’ or ‘inconclusives’ compared to ‘inclusions’. The ratio of

exclusions to inclusions is not what concerns critics. This data would say nothing regarding the more relevant issue of how many of these conclusions are actually correct.

Budowle and colleagues have recommended that the best way to overcome bias is by ‘peer review’, despite a lack of any literature verifying this claim. Peer review was often cited as a means of minimising observer bias and contextual effects during the interviews with Australian odontologists:

*Its hard to quantify.... I think being aware of [contextual effects and bias] is not really enough to mitigate it, its human nature, that's the problem. We do peer review. When we do it in pairs, the other person usually peer reviews it, so I think that helps.* (Participant #1624)

*To make that happen [to minimise bias], we invariably work as a team of two, and then have a third person peer review the case, who has been independent from our analysis.* (Participant #5195)

Yet this position fails to appreciate that observer effects are induced by environmental influences, and thus will affect *all* who work in a similar environment to some degree. The fact that Budowle and colleagues describe this process as ‘blind verification’ demonstrates that they do not fully grasp the nature of observer effects, despite referencing the work of Saks and colleagues (Saks et al., 2003). His use of the term ‘verification’ implies an affirmation of a result already achieved, when what is required is an independent assessment of the evidence in a context-free environment.

Pretty and Sweet (2001b) found that adherence to the ABFO guidelines on evidence collection was generally good. The majority of ABFO diplomats did not collect the bitemark evidence personally on all occasions, and the authors note this is a concern and recommend that the individual who will ultimately analyse the bitemark should collect the evidence from the suspect. This recommendation runs contrary to that offered by the studies conducted on observer effects and bias, which suggest that analysis and evidence collection should be carried out by independent individuals.

Role and conformity effects can be minimised by engaging as little as possible with the victim, law enforcement agencies and lawyers. The analysis should be conducted independent of these influences. This is most practically achieved by separating the phases of collection and analysis of odontological evidence. Where possible, the odontologist who is responsible for collecting the evidence should not be actively engaged in any subsequent analysis. This is already routine practice in other areas of forensic pattern analysis; for example, the crime scene examiners who lift latent prints, collect hair and fibres, and gather

bullet cartridges are rarely involved in the subsequent analysis or conclusionary phases of the same evidence. Such a protocol would go part way in ensuring the odontologist is not exposed to the highly emotive contextual information that naturally accompanies such procedures. This was seen as a good idea by most odontologists, several of whom raised the concept without any prompting:

*I quite like the idea of someone who takes the impression of the suspect should then pass the models onto a colleague who is uncontaminated with this emotional stuff that surrounds the crime. (Participant #6828)*

*Its probably better if the person analysing the case doesn't take the impressions. Because I find that if I have to go and take the impressions down at the watch-house, I know I am sussing him out, he's raped the girl or bashed her up or something like that. I think on some kind of subliminal level you are sussing out the likelihood of the person having done the crime, which isn't really your job. (Participant #7239)*

*[One] way around it would be to get the information and then give it to someone else, who hasn't seen anybody, and doesn't know anything about the case. (Participant #8547)*

However, this appears to be rarely, if ever, done in practice:

*You should be given the circs [circumstances], certainly, but not to take the impression or visit the suspect. But they always ask you to, because they usually have him down in the cells. (Participant #7239)*

*But you are usually collecting the information, doing the analysis and then seeing the suspect too, because there are so few of us. (Participant #8547)*

Several authors have recommended a separation of test and interpretation in order to guard against the effects of observer bias (Cole, 2004b; Koppl, 2005; Risinger et al., 2002; Whitman and Koppl, 2010). This is easily applied in forensic odontology by separating production of the overlay from the suspect's dentition and the subsequent analysis of the degree of match between the overlay and the bitemark. This practice is likely to lessen the likelihood of post-hoc reasoning in the explanation of discrepancies, and the concept was also raised by several of the interviewees:

*I don't think it's a good idea for people to be analysing the bitemark at the same time they have the dentition of the accused. I think that can bias you as to what marks you see. I think that they should be analysed separately. (Participant #1598)*

Emotive influences could further be avoided by limiting the amount of extraneous information available to the odontologist responsible for analysis of the bitemark. This includes analysis of the bitemark independent of knowledge surrounding the case in order to minimize emotional influences, before any viewing of the suspect's dentition in order to minimize target shifting, and before any other circumstantial evidence is revealed, such as the presence of the suspect's fingerprint, or DNA, to minimize confirmation bias.

*I think that when a case arises, its really important to disengage yourself from the crime, talking to the police officer right up front and saying "I don't want to know anything about the victim, crime or the suspect. What I need is this...". I'm usually at some pains not to get the details. The less I know the better. All I want from them are photographs. I don't want to do anything that might prejudice me one way or the other. (Participant #5195)*

Not all odontologists agreed with this, and many expressed a preference for more information regarding a particular case, rather than less:

*Well, we'd sometimes like more information from the police. Not case history, but on how the person was attacked, purely to align the bite. I think that sort of thing would be great. (Participant #1624)*

Forrest (2004) recently argued that context-free forensic science is impossible in most cases, and that context is indeed *required* for the majority of forensic scientific work such as toxicology and forensic pathology in order to accurately interpret the sample. In recognition of the fact that some extraneous information may be necessary in order for the forensic examiner to appropriately interpret their results, Krane et al (2008) proposed a protocol known as 'sequential unmasking', where information regarding the known sample is withheld until an initial interpretation of the results has been documented. Such a protocol is already in place in several DNA laboratories in the United States and proponents of this method claim it allows an unbiased analysis of the evidence by sequencing the laboratory work-flow such that evidentiary samples are interpreted, and the interpretation is fully documented, before reference samples are compared. This still provides the practitioner with the necessary information in order to draw conclusions about the evidence, but does so in a way that minimizes observer effects.

One potential issue with this approach in forensic odontology is related to another problem exemplified by the literature on the inter-reliability of bitemark comparison: that of deciding which features of a bitemark are valuable in determining the degree of match between the mark and the suspect dentition in the first place. Validity studies on the accuracy of various



analytical methods have differed in the selection of critical indicators of ‘accuracy’: Sweet and Bowers (1998) compared the area and degree of rotation of the outline of the tooth in their study of the accuracy of five methods of overlay production. McNamee and colleagues (McNamee et al., 2005) used the area and the centroid position on an *x* and *y* axis to compare the accuracy of computer generated overlays. Blackwell and colleagues (Blackwell et al., 2007) used the peak points (i.e. area of each tooth with the greatest *z*-dimensional value) and mesial and distal corners of the occlusal edges. Lasser and colleagues (Lasser et al., 2009) also used a third-dimension variable in their study. Other studies have used modifications of the now abandoned ABFO scoring system.<sup>5</sup> Hence the problem begins with the simple fact that there is yet to be agreement on what aspects of a bitemark are potentially probative, and which are not. All too often the odontologist is able to ‘explain away’ discrepancy via post-hoc reasoning, a criticism levelled at fingerprint analysis, firearms and toolmark analysts, and even DNA analysts (Thompson, 2009).

A sequential unmasking approach, if not necessarily an exact replica of that followed by DNA analysts, *could* be adopted in bitemark analysis. Ideally, the decision on what features are relevant and the identification of specific points of comparison should ideally be made *prior* to examining the suspect dentitions. Minimal information should be initially revealed to the analyzing odontologist at this stage. The first question to be answered is ‘could this be a bite mark?’ The answer to this question does not depend on the type or nature of the case, and thus the odontologist has minimal, if any, need for this information at this stage. Assuming the answer is ‘yes’, the odontologist should then attempt to identify the potentially relevant features of the mark that indicate the class and individual characteristics of the dentition, for example; marks from the upper versus lower teeth; marks that indicate the relative position of the incisal edges and position of the canine cusp tips; unusual or distinctive spatial arrangements, or the potential presence or absence of certain teeth. Analysis of the dentition should proceed separately from that of the bitemark. Ideally, more than one dentition, from persons who are unknown as to their involvement in the case to the odontologist, should be presented for analysis in order to avoid the generation of a purely confirmatory hypothesis. Again, class and individual characteristics should be noted at this stage.

Only having analysed both the mark and the dentition separately should the odontologist then attempt to combine this information in a single analytical technique. Following the combined analysis, other information, such as the reported position of the biter relative to the

---

<sup>5</sup> Introduced in 1981, it was withdrawn in 1984, yet published as late as 1995 in the Journal of the American Dental Association as a ‘credible method of staged evaluation’ (Rothwell, 1995).

victim in an assault case, can be revealed so that the odontologist can assess the relevance of this new information to the former conclusion. This process still ensures that the odontologist receives information that affords them the greatest opportunity to generate meaningful conclusions while removing unnecessary and potentially biasing detail.

Evidence line-ups have also been suggested as a means of minimising cognitive bias (Risinger et al., 2002; Whitman and Koppl, 2010). This is not a new suggestion; Miller (1987) demonstrated that less false associations occurred in the comparison of human hairs when the evidence was presented via a line-up method involving the comparison of the unknown sample to five known (but non-matching) exemplars, rather than the more usual method of comparing the sample to a single (also non-matching) exemplar. However, this situation is rarely encountered in practice:

*[The police] are always keen for me to match a bitemark to only one particular dentition, so it's never an open match. So there is an unconscious bias from the start. (Participant #6828)*

Despite this, the possibility of its usefulness was raised by several interviewee participants:

*Often I am given one bitemark and only one set of teeth to match. I think a better way would be to have some sort of 'witness line up' of dentitions. (Participant #7744)*

Most importantly, odontologists should consider avoiding analysis of bitemark evidence that is of poor or dubious quality, where the risk for contrast effects is greatest. The bitemark severity scale developed by Pretty (2007) may prove a useful starting point for assessment of bitemark evidence in this regard. Recent research has suggested that bitemark evidence of poor quality leads to greater disagreement between odontologists conclusions (Bowers and Pretty, 2009), which in turn suggests that contextual effects and the types of bias discussed herein play a much greater role than appreciated in analysis of bitemarks, especially those considered to be low quality. Further research is certainly warranted in order to better quantify the relationship between the quality of bitemark evidence and the accuracy of practitioner performance.

Lastly, it is suggested that odontologists should avoid giving statements of 'certainty' in their reports. Research has long demonstrated that there is no link between certainty and ground-truth, and any statement regarding certainty is potentially misleading and irrelevant to proper assessment of the evidence. It is outside the role of the forensic expert to offer an opinion on the value of their evidence. This remains a privileged task for the trier of fact alone.

## Conclusion

It is well established that contextual effects are universal phenomena, which may not be eradicable, but nonetheless can be minimised, through carefully designed collective and analytical processes. The growing body of literature in other disciplines suggests that observer effects induced by extraneous or emotive information play a significant role in the outcome of forensic decision making. It is difficult to assess the extent to which cognitive bias influences forensic odontologists opinions as there have been no studies to date on this subject.

This is less than ideal for two reason; firstly because of the more obvious concerns it raises about the validity of conclusions reached about potential bitemark matches; and secondly, because we then fail in our responsibility to provide the trier of fact with reasonable information regarding these influences in order to allow them to undertake an accurate assessment of the weight of such evidence. Furthermore, quantification of these influences is critical in order to develop and evaluate training and procedures within the odontology community so as to improve overall practitioner performance.

Dror and Rosenthal were careful in their interpretation of results from studies on cognitive bias in fingerprint examiners: ‘The fact that fingerprint experts can be unreliable and biasable does not mean that they are not ordinarily reliable and unbiased. It is not our place to determine what is the acceptable norm for expert performance’ (Dror and Rosenthal, 2008). Similarly, there is no claim that odontologists are routinely biased or over-confident in their assessment of bitemarks, however, the results of chapter 7 clearly demonstrate that there is a fundamental concern with the reliability of bitemark analysis as practised today, and it is a well established proposition that context effects exist and can be at least partly responsible for such inconsistency in performance. Additionally, until studies quantifying the degree to which odontologists are subject to these effects, they remain an unknown and potentially significant factor to consider when evaluating the overall reliability of bitemark analysis. Given the current criticisms of bitemark analysis, and the potential future direction it will take, it would seem logical to at least take steps to minimise potential biasing effects until there is experimental data available to qualify under what circumstances, and to quantify to what extent, they influence our analysis and interpretation of bitemark evidence.

## CHAPTER 9

# Conclusion

---

## Is *Daubert* Relevant?

Despite the wealth of legal and scientific commentary, it appears that the *Daubert* precedent has had minimal impact on the relative admission of forensic science in courts of law. Although challenges to forensic identification evidence are now far more commonplace since the *Daubert* and *Kumho* opinions were released, the pattern of exclusion of forensic science evidence has changed minimally at best. A few judges have taken *Daubert* and *Kumho*'s requirements very seriously and have applied them without mercy to forensic identification science in light of the amended Federal Rule 702, but the vast majority of them have not yet dared to exclude such mainstream practices such as fingerprint analysis on grounds of reliability. Even more controversial practices such as bitemark analysis have been excluded or limited only in a minority of cases where such evidence was challenged.

Chapters two and three demonstrated that these *Daubert* 'factors' of testing, peer review, error rate and general acceptance have been misinterpreted and misconstrued by many in the legal and forensic science arenas, and they do not yet represent the gatekeeping 'pillars' of expert evidence in court that many hoped they would. Where forensic identification evidence has been challenged successfully, detailed analysis of judicial reasoning has revealed a far more subtle pattern of exclusion or limitation, based on reasons such as unfounded statistics, failure to follow protocols and procedures, unreasonable conclusions by the witness, improperly carried out tests and experiments, and unbelievable claims of infallibility. These shortfalls still amount to a failure to demonstrate the 'reliability' of any given science, but this notion of reliability has rarely followed the *Daubert*-specific criteria.

Chapter four demonstrated that *Daubert* and the trends of judicial exclusion of evidence in the United States has had little bearing on the admissibility of expert evidence in Australia. Attempts to introduce *Daubert*-like precedents have been repeatedly over-ruled by higher courts at both state and federal level. Perhaps one reason for this reluctance to introduce reliability requirements in Australia is because there is no perceived need to introduce more stringent limitations on what experts can and cannot say in court due to the relative insignificance of a 'junk science' epidemic in this country. The Australian judiciary seem far

more pre-occupied with more fundamental concerns, such as ensuring experts have specialised knowledge outside the ken of the ordinary trier of fact. The matter of whether such knowledge is reliable, or has been reliably applied to facts of the case is a matter that the judge or jury is entitled to assess on their own, and merely goes to the weight of such evidence. Australia has a far more liberal approach to the admission of expert evidence, and this is jealously guarded. Consequently, approaches to improving the reliability of forensic odontology, in particular bitemark analysis as it is practised in this country, would be even more misguided if one were to use the *Daubert* guidelines as a framework.

*Daubert* was a decision made by judges in order to solve jurisprudential issues. It was never intended to act as guidance for conducting the day-to-day business of science. *Daubert*'s application to furthering research and practice is still peripheral to the well-established notions inherent in the hypothetico-deductive model of 21<sup>st</sup> century science. Yet what *Daubert* has done is to draw attention to the concerns about forensic science already noted by a small number of others to a much wider audience. The close relationship that forensic science has with the law means that it cannot afford to completely sideline the concerns raised by this judicial precedent – but here the crux lies: *Daubert* has raised the problem, but not solved the issue.

There are clearly concerns with the overall reliability of certain disciplines, but that is not to say that certain forensic identification sciences are inherently unreliable. The fact that these disciplines cannot yet quantify this degree of reliability is what makes these concerns a serious issue. This issue is not going to be solved by the legal system, and this has been evident in the way that courts in the US are still mishandling forensic science evidence, the *Daubert* precedent and the most recent iteration of Rule 702. The broad concept of requiring 'reliable science' in court is an admirable one, yet it is one that the courts have limited powers to ensure. While the legal framework has a relatively recently-generated notion of what it considers reliable science, science itself has, for many years been adhering to, if not exceeding, these standards through the self-regulatory mechanisms that define the very word *science*. Legal definitions, commentary and rhetoric should not take precedence over the existing standards, intrinsic to the hypothetico-deductive paradigm, that have been several hundred years in the making.

The problem that is faced by forensic science, in particular the forensic identification sciences, is that we have enjoyed the title of 'science' for too long without seriously considering its implications. Now that *Daubert* has forced many to reflect on what, exactly, these disciplines are actually offering in terms of scientific content, the realisation falls upon

us that there might not be all that much there. Despite Australia's reluctance to require reliability as a cornerstone for admission of scientific evidence in its courts, its judicial philosophy continues to place the moral burden of ensuring that expert testimony is based on solid scientific grounds on both the expert and the discipline. In this respect, Australian experts have no less liability than their US counterparts to ensure the reliability of their testimony, despite the differences in written doctrine.

## The Scientific Status of Bitemark Analysis

Turning to a critical exploration of the reliability of bitemark analysis, Chapter 5 explored the first aspect of its three-pronged central theorem, that of the dentition being considered unique. Numerous philosophers have told us that uniqueness is not an establishable proposition, and belief in such represents nothing more than assumption. Studies attempting to prove the uniqueness of any trait have yet to provide convincing evidence that any feature or combination of features is unique, as they face insurmountable logical and mathematical issues, most of which cannot be overcome using modern-day experimental techniques. Yet regardless of the technical difficulties associated with these studies, 'uniqueness' is neither relevant to the theory or practice of forensic identification, nor to the courts. The forensic expert's fundamental role is not to individualise or even to identify, based solely on the forensic evidence presented to them — it is to provide the trier of fact with additional probative information, within the bounds of their expertise, that may strengthen or weaken the likelihood of guilt. Experts need not usurp the trier of fact's role by drawing prejudicial and unsupportable conclusions regarding identity.

Directing forensic research resources towards proving uniqueness is therefore fruitless, and diverts resources away from more useful projects. To attempt to prove uniqueness also puts the cart before the horse when more basic issues such as the meaning of the term *match*, the quantification of level of certainty and the standard of practitioner performance — all of which will influence how we actually define the term 'unique' — are yet to be agreed upon. Areas such as fingerprints, odontology, and firearms and toolmark analysis appear to defend the uniqueness proposition largely due to a misunderstanding of what the criticisms of their discipline are. Mistakes and misidentifications are not made because someone has an identical fingerprint to someone else in the world. They are made because of guesswork, poor performance, lack of standards, bias and observer error. The concept of uniqueness is unrelated to that of reliability, and proof of its existence or otherwise does little to address these concerns.

The second tenet of bitemark identification relies on the notion that a faithful reproduction of the dentition is left in a medium. It has already been well established that objects do not leave identical marks in media even when the same object is applied more than once. Fingerprint analysts readily agree that no two prints, even from the same finger, will ever be the same. There is no reason to suggest that bitemarks may somehow be exempt from this phenomenon. There is already good evidence to suggest that bitemarks from the same dentition will always differ from one another in appearance due to the nature of the medium in which they are recorded. Bush and colleagues have established not only this proposition to be true (Bush et al., 2009, 2010a), but also that the variation between the bitemarks from the same dentition can be greater than that seen between bitemarks from different dentitions (Bush et al., 2011). Similar conclusions can be drawn from studies in fingerprints (see Kaye, 2003a; Srihari, 2009), and studies of multiple images of the frontal sinuses in humans (Christensen, 2005). The concept that there is tremendous and sometimes unpredictable distortion in bitemarks has long been recognised since the earliest days of research into this area (Jakobsen and Keiser-Nielsen, 1981), and few odontologists would dispute that this is the case. Thus we can argue that the second tenet of bitemark identification theory is also irrelevant as far as improving the overall reliability of the discipline as it has already been largely proven and indeed accepted as false.

It is the third proposition in the theory of bitemark analysis that holds the key to improving the overall reliability of this discipline, yet there is a paucity of data regarding the validity of this principle, that of the ability of odontologists to match bitemarks to dentitions. The studies that have attempted to characterise this have either been so poorly designed that their results cannot be relied upon to any great extent, or have called the ability of most odontologists to correctly match dentitions to bitemarks into grave question.

## **Bitemark Analysis – Contemporary Australian Practice**

One of the major difficulties in establishing data relevant to the reliability of odontology practice in this regard is the wide variety of available methods by which bitemark analysis can proceed. Use of these methods depends on several factors such as the nature of the case; the particular forensic question being asked; the training and experience of the odontologist; and the perceived accuracy of any given method by the practitioners themselves. Chapter 7 explored the current status of bitemark analysis in Australia with these issues in mind. While it has demonstrated that the practice of bitemark analysis itself is not necessarily aligned with those sensationalised cases that have caught the public eye, there remains more fundamental concerns with bitemark analysis techniques and the practise of odontologists in

this country that need to be addressed prior to any *meaningful* analysis of the overall reliability being undertaken. Examination of the practice of bitemark analysis in Australia has led to the distillation of some key areas of concern that inextricably linked to the reliability of the technique as a whole.

### **Nature of casework**

Bitemark casework in Australia is sporadic, and represents only a small proportion of forensic odontology practice, on average probably much less than 10%. The circumstances surrounding bitemark analysis in Australia today does not usually correlate with those of former, very public, high profile murder cases where bitemark evidence was deemed to be an essential element of the material evidence against a defendant. The majority of these sensationalised trials, both in Australia and overseas, date from the 1980's where evidence from at least one odontology centre suggests that bitemark analysis originally formed a much greater proportion of casework.<sup>1</sup> Court appearances by odontologists to testify about bitemark cases are exceedingly rare, with most practitioners having never testified in court regarding bitemarks. Of those that have been called as expert witnesses on the subject of bitemarks, very few have had experience beyond one or two cases. The lack of suitable opportunity to engage with bitemark analysis and act as an expert witness in this area is concerning for the purposes of remaining current within the profession.

### **Quality of Evidence**

There is generally poor standard of detailing and recording bitemark evidence. This suggests that continued efforts need to be made to educate police and medical professionals regarding the proper procedures for recording forensic dental evidence via photographic means. Bitemarks that contain three-dimensional information, such as indentations in skin, appear to be exceedingly rare. Most casework is presented as photographs, with little opportunity afforded to an odontologist to physically examine the suspected injury themselves. Additionally, 'forensically significant' bitemarks that enable meaningful analysis and conclusions to be drawn are in the minority as far as evidence offered for expert opinion. This, combined with the fact that the detailing and recording of such evidence is often sub-standard, further reduces the likelihood of obtaining meaningful results from the majority of bitemark evidence offered for expert opinion in Australia.

---

<sup>1</sup> Consider the figures from this centre in 1986, where 60% of the work undertaken by this unit involved bitemark analysis, compared to the year 2009, where bitemark analysis contributed to less than 8% of the total odontology case load. Possible reasons for this is the advance of more scientifically reliable techniques in Australia, such as DNA profiling, for identification of perpetrators, in addition to increasing awareness of the inherent subjectivity in bitemark analysis.



## **Interpretation of Bitemark Injuries**

Odontologists practicing bitemark analysis in this country should be considered well-qualified and highly experienced in the field of forensic odontology, however, there is still a noticeable tendency to over-interpret injuries and draw conclusions that cannot be realistically supported by any scientific literature. Some basic tenets of bitemark injury interpretation appear to have been forgotten, or misconstrued. Specific areas of opinion given by Australian odontologists that are cause for concern are:

### ***Is the injury a bitemark?***

The most fundamental question regarding the nature of a suspected mark has been shown to be poorly considered amongst odontologists. Differences in opinion as to whether the suspect mark was or was not made by teeth are difficult to reconcile in the absence of clear reasoning for any particular decision. Given the imperfect nature of skin as a medium, distortions of the size and shape of bitemarks are highly likely, and as such should not be used to definitively rule out the possibility that a mark was made by teeth. Similarly, the fact that a similar mark has never been encountered in a practitioners career forms no basis for an opinion that the mark therefore could not have been made by teeth. It is suggested a more fundamental approach to analysis of a suspect mark is warranted, based on the presence or absence of so called ‘class’ and ‘individual’ characteristics. These are well described in the literature, and form a more solid and defensible proposition that a mark may have been made by teeth when compared to experience alone as they provide objective criteria upon which to distinguish a bitemark from any other patterned injury. This again follows standard practices in other pattern-matching disciplines.

### ***Ageing of bitemarks***

The ability to age a bitemark is not supported by the current literature, and comments regarding the age of a bitemark that consists of bruising only should be considered highly contestable. The ability to distinguish age of bruises has been demonstrated to be reliable only in context of the wound being described as ‘recent’ or ‘older’.

### ***Nature of circumstances in which the bitemark occurred***

While the presence of bitemarks themselves generally indicate that a reasonable degree of force was involved in the biting process in order to generate such a mark, the presence of lacerations as opposed to bruises does not necessarily imply a greater force was used in the biting process. While retrospective reviews have demonstrated a link between the nature of the crime and the position of the bitemark, statements regarding the motive or circumstances

surrounding the bitemark represent conjecture on the part of the odontologist, and are likely to face strong criticism in court.

### ***Central ecchymosis***

There has been debate on the origin of central areas of ecchymosis within bitemarks. While current theories favour the presence of central areas of ecchymosis being due to compression of vessels as the tissue is squeezed between teeth, this too remains largely untested experimentally. The presence of a central area of ecchymosis adds nothing except weight to the argument that the injury is indeed a bitemark. It is unreasonable to conclude that the presence of central ecchymosis is related to suction or tongue-thrusting, or that this means there is then a sexual implication in the nature of the biting attack.

### ***Adult or child as the perpetrator***

The assessment of whether a bitemark was caused by an adult or child should be approached with caution. Due to distortion and population variance, measurements of arch width can be misleading. The literature is thin regarding ranges for child inter-canine distances, and it has been demonstrated that arch length is can be one of the most distorted features of a bitemark. Levine proposed that a maxillary intercanine distance of greater than 30mm is likely to represent an adult and less than 25mm is likely to indicate a child, and this standard appears to remain extant today. Yet there is considerable research that has demonstrated that intercanine width is highly variable both within and between populations. Measurements for adults have been recorded as little as 21mm. Studies on accuracy rates in this area suggest that there is a significant chance that the odontologist will make an incorrect call when deciding whether a bitemark had been produced by an adult or a child, and these factors need to be considered when expressing certainty in this area.

### **Method of Analysis**

The overlay method of comparative analysis is still the most common, and most readily accessible method used by odontologists in Australia. The preponderance of hand-traced overlays is concerning in light of the literature supporting digitally-generated methods as the more accurate, especially when supported by quantitative measurements. While a significant proportion of practising odontologists use digital methods of overlay production, the generation of the outlines of teeth from scanned models is still subject to criticism. Ideally, overlays should be generated from digitisations of wax bites rather than scanned models of the dentition, and this is supported by the current literature as the more accurate and less subjective method. This also fits neatly with comparison techniques in other pattern-matching disciplines. Construction of overlays involving 'wax scrapings' rather than mere

imprints could also be considered as a supplementary measure to standard ‘bitemark’ overlays, in keeping with the philosophy of comparing marks known to be made by a particular tool to suspect marks, however, this is only likely to prove useful if the bitemark itself demonstrates features suggestive of drag marks.

Direct scans of the models of the dentition should only be used when the bitemark is considered ‘good quality’ and where it is the intent of the odontologist to demonstrate particular features of the dentition, yet this represents the most commonly used method in Australian practice. This sits at odds with the relatively low incidence of highly forensically significant injuries. Given the nature of bitemark casework in Australia, where most evidence is of low quality, it is suggested that this should not be the first-line approach in fabrication of digital overlays, although it remains a technique that could be considered supplementary to the production of an overlay from a wax bite. Other techniques for comparison, such as the acrylic template method, need to be scientifically validated prior to adopting it as a primary means of comparison.

3-dimensional techniques of analysis have limited application at the current time due to the fact that very little bitemark casework, at least in Australia, ever contains 3-dimensional information by which to compare. Distortion of the mark in human skin renders it unlikely that a manual ‘fit’ of a suspect casts into the bitemark itself will ever yield a satisfactory comparison. It is recognised that this may be of some use where the bitemark is on an inanimate object, however, it is not recommended as a stand-alone means of comparison and should be supplemented by another, validated technique.

No one method of analysis provides enough information to support the theory of relationship between a suspect dentition and a bitemark. An ‘integrated’ technique involving both a pattern associated analysis and a metric analysis was suggested by Bernitz and colleagues (Bernitz et al., 2008), and this seems reasonable given the difficulties associated with both of these analysis in isolation.

### **Standard of Reporting**

Written reports on bitemark evidence in Australia tend to follow conservative lines. However, there still remains a proportion of reports that draw questionable conclusions given the quality of the material evidence, the positivity of the conclusion as reached by the odontologists and the supposed significance of the injury itself (as described by the Severity and Significance score). These mismatches between conclusions and material strength of the evidence may suggest that Australian odontologists may need further training in ascertaining what may or may not be suitable for analysis, or that odontologists should perhaps be more

mindful of the relationship between the strength of the conclusion they reach and the quality of the evidence before them, so as to avoid these rhetoric mismatches. There is the possibility that the severity and significance scale developed by Pretty requires refining in order to capture subtleties in bitemark appearance that would otherwise deem them as having a high (or low) significance rating. Further research into the validation of this scale is certainly warranted in order to determine its ultimate usefulness in forensic practice.

There is a need to temper the tone of written reports with the quality of the evidence in question. For example, photographs of suspected bitemarks without an adequate scale are essentially useless for further analysis other than to provide a qualitative opinion on the nature of the injury. An ABFO No. 2 scale should be considered the gold standard in this regard. That is not to say that analysis of images that do not contain an ABFO scale cannot be pursued, but any conclusion reached should be considered of low weight, and this needs to be clearly articulated in the opinion.

### **Language**

At the current time, where there is no reasonable data on which to base a numerical probability of the likelihood that a suspect was the originator of a bitemark, probabilistic statements are entirely inappropriate both from a legal and scientific perspective for summarising the results of bitemark analysis. The difficulty remains in defining what terminology is appropriate to use when describing the results of analysis. Given the wide variation in meaning between individual practitioners regarding terminology such as ‘possible’, ‘probable’ and so forth, use of these terms still renders them subject to criticism. Yet there is no legal or scientific requirement to summarise the results of any given bitemark analysis in one or two word phrases, using such highly suggestive words. These terms have evolved from the DVI aspect of forensic odontology with regard to opinions on identity, and given the true limitations of bitemark analysis not only is the positive identification of a suspect from a bitemark an indefensible proposition given the current status of the science, but any *other* comment on the possible identity of a suspect from tooth marks should also be considered inappropriate. It is not the odontologist’s role to ‘identify’; that is a task for the trier of fact alone; and so any language that implies an attempt at identity is actually misguided, no matter what other disciplines consider acceptable.

Given the criticism that other areas of forensic science have faced regarding their own particular scales of conclusionary terms, including criticism by the courts, it seems reasonable that odontologists should consider restricting themselves to describing the association between the dentition in question and the latent mark, until research is able to

justify a nominal scale of conclusions. Far from being considered of little use to the legal process, courts have been receptive to this suggestion in a number of jurisdictions in the US and have specifically limited forensic identification experts to this role.

Such a scale might be along the lines of:

- a. *Strong association* between number and position of the teeth of the suspect and the latent mark
- b. *Weak association* between the number and position of the teeth of the suspect and the latent mark
- c. *Minimal association* between the number and position of teeth of the suspect and the latent mark
- d. *Significant disparity* between the number and position of teeth of the suspect and the latent mark

Such a scale leaves determination of identity up to the trier of fact, who can use these conclusions, along with other evidence, in order to determine the overall likelihood of guilt or innocence. Granted, the degree of association that qualifies as ‘strong’, or ‘weak’ is likely to be contestable (and this scale is nothing more than a suggestion for the purposes of illustration), but this points towards a need for further research into the quantification of bitemarks, and the range of possible alternative marks that could be made by a single dentition.

As an exception to this principle, it would perhaps be reasonable to conclude that one dentition was *more likely* to have made the mark than another when given two (or more) suspected dentitions to compare, as a comparison between a greater and a lesser match for a given mark can then be demonstrated. However, use of the term ‘match’ should certainly be avoided in conjunction with any conclusions regarding the association between bitemarks and a given set of teeth:

*Use of the word ‘match’ to denote the outcome of comparative analysis in odontology and elsewhere in forensic science should be banned as perniciously misleading. To the understanding of a lay person, a match means an identification has been made to a particular person or object to the exclusion of all other persons or objects. (Starrs, 1994)*

Odontologists who testify in court may be concerned that unless they have something concrete to offer the court, their evidence may be dismissed as not worthwhile, yet there is no legal precedent for this. Case law indicates that Australian courts are relatively liberal ‘admitters’ of scientific evidence in both criminal and civil cases. This liberal approach is

reflected in the Uniform Evidence Acts that speak only for qualification of the witness as an ‘expert’ in order for them to be allowed to express their opinion in court. Odontologists must accept that their evidence, considered alone may not be strong enough to convict or exonerate, but understand that this does not make it inadmissible or somehow unusable by the legal system. By contrast, comments regarding associations of identity are potentially more prejudicial than probative given the lack of well-founded evidence to suggest that such associations are possible, and this *does* represent legal grounds for exclusion

### **Potential for Bias**

A significant proportion of the odontology community do not agree that contextual bias is a significant issue in their own odontology practice, despite the literature having well established that contextual effects are universal phenomena. Chapter 8 explored this thinking, and demonstrated that odontology is littered with potential sources of bias, from both a historical review of practices via casework, and consideration of current practices and procedures. The preponderance of low-grade bitemark injuries as seen in chapter 7 also represents the format of evidence most susceptible to some of the more pervasive types of cognitive bias and contextual effects, and forensic odontologists cannot ignore the potential for these to affect their decision making. Such biases may not be eradicable, but nonetheless can be minimised, through carefully designed collective and analytical processes.

Despite the longstanding recognition of these effects, there have been no studies to date on the influence of cognitive bias in forensic odontology yet such quantification should be considered critical for developing and evaluating standardised procedures within the odontology community in order to improve overall practitioner performance. It is suggested that until this occurs, conservative protocols for evidence collection and subsequent analysis should be articulated with the potential for cognitive bias in mind in order to minimise these effects. Separation of the collection and analytical phases is one relatively simple way in which this can be achieved with minimal changes to existing procedures. The ability to compare more than one dentition to a bitemark should be considered another advantage in minimising target-shifting, and where possible, a range of dentitions from various suspects should be compared, rather than basing conclusions on the result of just one comparison exercise.

## The Future

There is little justification for reasoning along the lines of ‘because I am the expert’ in 21<sup>st</sup> century jurisprudence, particularly when such crucial decisions regarding life and liberty are made on the basis of identification testimony.

*It does strike me that forensic scientists ought to ‘own up’ to the humility of the process, which is what most scientists do that are good scientists... consider a doctor who is giving cancer treatment and you ask him what the prognosis is, he might say ‘somewhere between 5 and 70%, we don’t have any data better than that.’ (Faigman, 2009)*

In the 1980’s, it was stated that ‘the aim of [bitemark] analysis is to identify the owner of the guilty teeth’ (Jakobsen and Keiser-Nielsen, 1981). These authors also declared that ‘one arch mark is never a bite mark, but it may hold adequate detail to allow identification of the originator’. This philosophy of ‘identifying’ perpetrators from bitemarks has only recently come under scrutiny, despite the absence of any robust data having been a problem since its inception, and there have been several articles that have individually called for an end to this concept of ‘identification’ from bitemarks (Clement and Blackwell, 2010a; Pretty and Sweet, 2010) due to the obvious lack of scientific literature supporting the concept that individuals can be identified from their tooth marks on human skin.

What chapter 7 demonstrates, above all, is that there is a need for consistency in the practice of bitemark analysis between practitioners. This alone will help improve the reliability of the discipline and can be achieved with relative ease. A draft series of guidelines, based on the concerns noted with current Australian practice is offered as an Appendix to this thesis. Yet there is also a dire need to reinforce the practice of bitemark analysis with some robust *scientific* evidence to support its position in the forensic identification armamentarium. Offering results from studies that attempt to characterise rates of accuracy, based on standardised analysis techniques ascertained under controlled conditions (as opposed to using historical casework samples) using larger populations to select from than only three or four dentitions, is one relatively straightforward way in which odontologists can assist the trier of fact in assigning weight to their testimony.

Other research avenues still to be potentially explored are; studies into the relative forensic significance of varying features of a bitemark, in order to establish which of the many geomorphic aspects of a bitemark injury make association with a dentition more likely; studies that demonstrate the limits of usefulness of certain types of bitemark injury as forensic evidence, as an extension of the work suggested by Pretty; studies that demonstrate

the possible ranges of dentitions, characterised by the geomorphic arrangements of individual teeth, that could cause a particular injury, and; studies that then relate this data to population data in order to provide a likelihood ratio that a given mark was made by a given dentition.

These represent only a small number of suggestions that alone represent many years of carefully directed research that will assist only to steady the shaky foundations bitemark analysis currently rests on. There are many other avenues to be explored, such as how human factors influence decision making in forensic casework; this alone represents a whole series of studies that are yet to even be considered in forensic odontology. Given the paucity of fundamental research supporting the practice of bitemark analysis, the profession urgently needs to place reasonable limits on its claims so as to avoid potential ruin in the future. Deciding what these limits are depends on the current status of the discipline as practised in Australia today, and not on sensationalised reports of isolated cases in other countries that have sadly been the focus of the public, the legal system, and the profession in the last few years. Bitemark evidence is, by itself, likely to be a poor and ultimately inappropriate foundation on which to base identification of individuals, but that does not mean that it has no role as a forensic tool, or that improvements in its current, albeit limited, capabilities should not be attempted.

While it appears to be a widely accepted legal principle in Australia that experts are allowed to offer their opinion, whatever that may be, one must also bear in mind the words of a late 19<sup>th</sup> century medico-legalist, Dr Paul Brouardel:

*If the law has made you a witness, remain a man of science; you have no victim to avenge or guilty or innocent person to ruin or save. You must bear testimony within the limits of science.*

As practitioners of bitemark analysis, it ultimately falls to us to define and defend our own limits, and this can only be done as far as the scientific basis allows us. There is an urgent need to adapt and adhere to some basic guidelines that recognise these limits, until more robust scientific evidence is added to this foundation that will allow us to make greater claims. Failure to do so leaves us subject to ridicule, exclusion, and even excommunication from the legal domain; that province of today's fair and just society we can potentially offer so much to.



# References

---

- Abbott, N. & Magnusson, L. (2008) An Enigmatic Degree of Medical Certainty. *Utah Bar J*, 21(4), 20-24.
- Acharya, A.B. & Mainali, S. (2009) Limitations of the mandibular canine index in sex assessment. *J Forensic Leg Med*, 16(2), 67-69.
- Advisory Committee (2000) Notes on the Federal Rules of Evidence (2000 Amendment).
- Allen, R.J. (1994) Expertise and the *Daubert* Decision. *J Crim Law and Criminology*, 84:4, 1157.
- American Board of Forensic Odontology, Diplomates Reference Manual. Available at <http://www.abfo.org>, accessed on 02 June 2011
- Anderson v Liberty Lobby, Inc.* 477 US 242, 253 (D.C. Cir. 1986)
- Annells, M. (1996) Grounded theory method: Philosophical perspectives, paradigm of inquiry, and postmodernism. *Qualitative Health Research*, 6(3), 379.
- Anonymous (1) Interview with author. (27 Oct 2009)
- Antoine-Tubbs v Local 513 Air Trans Div* 50 F.Supp. 2d.601 (N.D.Tex)
- Arheart, K.L. & Pretty, I.A. (2001) Results of the 4th ABFO Bitemark Workshop--1999. *Forensic Sci Int*, 124(2-3), 104-111.
- Asherson, N. (1965) Identification by Frontal Sinus Prints: A Forensic Medical Pilot Survey. Lewis and Co, London.
- Ask, K. & Granhag, P.A. (2005) Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, 2(1), 43-63.
- Attorney Accountability Act of 1995 H.R. 988, 104th Cong., 1st Sess (1995) (amending Fed. R. Evid. 702).
- Australian Law Reform Commission. (1985) *Discussion Paper 23 - Evidenc*. Commonwealth of Australia.
- Australian Law Reform Commission. (2005a) *Discussion Paper 69 - Review of the Uniform Evidence Acts*. Commonwealth of Australia.
- Australian Law Reform Commission. (2005b) *Report 102 - Review of the Evidence Act 1995*. Commonwealth of Australia.
- Avon, S.L., Victor, C., Mayhall, J.T. & Wood, R.E. (2010) Error rates in bite mark analysis in an in vivo animal model. *Forensic science international*, 201(1-3), 45-55.
- Babchuk, W.A. (2009) Grounded theory for practice-based application: Closing the embarrassing gap between theory and empirical research. *Proceedings of the 28th Annual Midwest Research Research-to-Practice Conference in Adult, Continuing, and Community Education*, 2009, Chicago.
- Baker Valley Lumber Co. v Ingersoll-Rand Co.* 148 N.H. 609, 614 (2002)
- Balding, D.J. (1999) When can a DNA profile be regarded as unique? *Sci Justice*, 39(4), 257-260.
- Balding, D.J. (2005) *Weight-of-evidence for DNA profiles*. John Wiley & Sons, Chichester.
- Barbenel, J.C. & Evans, J.H. (1974) Bite marks in skin - mechanical factors. *J Forensic Sci Soc*, 14(3), 235-238.
- Bariciak, E.D., Plint, A.C., Gaboury, I. & Bennett, S. (2003) Dating of bruises in children: an assessment of physician accuracy. *Pediatrics*, 112(4), 804.
- Barrett v International Armaments, Inc.* 1999 US Dist. LEXIS 4708

- Barsley, R.E. & Lancaster, D.M. (1987) Measurement of arch widths in a human population: relation of anticipated bite marks. *J Forensic Sci*, 32(4), 975-982.
- Baum, C.L. & Arpey, C.J. (2005) Normal cutaneous wound healing: clinical correlation with cellular and molecular events. *Dermatol Surg*, 31(6), 674-86; discussion 686.
- Beecher-Monas, E. (2008) Reality Bites: The Illusion of Science in Bite-Mark Evidence. *Cardozo L. Rev.*, 30, 1369-1410.
- Beecher-Monas, E. (1998) Blinded By Science: How Judges Avoid the Science in Scientific Evidence. *Temple Law Review*, 71, 55.
- Beecher-Monas, E. (2007) *Evaluating scientific evidence : an interdisciplinary framework for intellectual due process*. Cambridge University Press, Cambridge ; New York.
- Begley, S. (2004) Despite its reputation, fingerprint evidence isn't really infallible. *Wall Street Journal*, 04 Jun 2004.
- Bell, D. (2000) From *Frye* to *Daubert*. *Aust J Forensic Sci*, 32, 61-64.
- Bell, D.S. (2006) Controversy Corner. *Aust J Forensic Sci*, 38, 107-110.
- Benedict, N. (2004) Fingerprints and the Daubert Standard for Admission of Scientific Evidence: Why Fingerprints Fail and a Proposed Remedy Note. *Ariz. L. Rev.*, 46, 519.
- Benner, P. (1982) From novice to expert. *AJN The American Journal of Nursing*, 82(3), 402.
- Benson, B.W., Cottone, J.A., Bomberg, T.J. & Sperber, N.D. (1988) Bite mark impressions: a review of techniques and materials. *J Forensic Sci*, 33(5), 1238-1243.
- Berg, S. & Schaidt, G. (1954) Methodik unter Beweiswert des Bisspurenvergleiches. *Kriminalwiss. Beil. Z. Kriminalistik*, 8, 128.
- Berger, M.A. (1993) Quoted in "Supreme Court to judges: Start thinking like scientists. *Science*, 261, 22.
- Berger, M.A. (2005) What has a decade of Daubert wrought? *Am J Public Health*, 95 Suppl 1, S59-65.
- Bernitz, H., Owen, J.H., van Heerden, W.F. & Solheim, T. (2008) An integrated technique for the analysis of skin bite marks. *J Forensic Sci*, 53(1), 194-198.
- Bernstein, D.E. (1996) Junk Science in the United States and the Commonwealth. *Yale Journal of International Law*, 21, 123.
- Bernstein, D.E. (2007) Expert Witnesses, Adversarial Bias, and the (Partial) Failure of the Daubert Revolution. *Iowa L. Rev.*, 93, 451.
- Bernstein, D.E. (1994) The Admissibility of Scientific Evidence after *Daubert v Merrell Dow Pharmaceuticals Inc.* *Cardozo L rev*, 15, 2139.
- Bernstein, D.E. & Jackson, J.D. (2003) The Daubert Trilogy in the States. *Jurimetrics*, 44, 351.
- Betz, P. (1994) Histological and enzyme histochemical parameters for the age estimation of human skin wounds. *Int J Legal Med*, 107(2), 60-68.
- Bishara, S.E., Jakobsen, J.R., Treder, J. & Nowak, A. (1997) Arch width changes from 6 weeks to 45 years of age. *Am J Orthod Dentofacial Orthop*, 111(4), 401-409.
- Black, B. (1994) The Supreme Court's View of Science: Has *Daubert* Exorcised the Certainty Demon? *Cardozo L Rev*, 15, 2129.
- Black, B., Ayala, F.J. & Saffran-Brinks, C. (1994) Science and the Law in the Wake of Daubert - A New Search for Scientific Knowledge. *Texas L Rev*, 72, 715-784.
- Blackwell, S.A. et al. (2007) 3-D imaging and quantitative comparison of human dentitions and simulated bite marks. *International Journal of Legal Medicine*, 121(1), 9-17.

- Bogner, A. & Menz, W. (2001) *Deutungswissen und Interaktion Zu Methodologie und Methodik des theoriegenerierenden Experteninterviews: The theory-generating expert interview as a research instrument. Soziale Welt*, 52(4), 477-500.
- Bogner, A., Littig, B. & Menz, W. (Eds.) (2009) *Interviewing experts* Palgrave Macmillan, Basingstoke [England] ; New York.
- Bonfanti, M.S. & De Kinder, J. (1999) The influence of manufacturing processes on the identification of bullets and cartridge cases-a review of the literature. *Science & Justice*, 39(1), 3-10.
- Bonifant, N. (2010) *Blackwell v Wyeth: It's our courtroom and we'll Frye (only) if we want to* – the Maryland Court of Appeal's unstated adoption of *Daubert*. *Maryland L Rev*, 69, 719.
- Bourke, J. (1993a) Misapplied Science: Unreliability in Scientific Test Evidence - Part I. *Australian Bar Review*, 10(2), 124.
- Bourke, J. (1993b) Misapplied Science: Unreliability in Scientific Test Evidence - Part II. *Australian Bar Review*, 10(3), 183.
- Bourne v Town of Madison* 2007 US Dist. LEXIS 34537
- Bowers, C.M. (2004) *Forensic dental evidence : an investigator's handbook*. Academic Press, Boston.
- Faigman, D.L., Saks, M.J., Sanders, J. & Cheng, E.K. (Eds.) (2007a) *Forensic Odontology* Thomson/West, Minnesota.
- Bowers, C.M. (2001) Jurisprudence issues in forensic odontology. *Dent Clin North Am*, 45(2), 399-415, ix.
- Bowers, C.M. (2007b) Commentary on: Keiser JA, Bernal V, Wadell JN, Raju S. The uniqueness of the human anterior dentition: a geometric and morphometric analysis. *J Forensic Sci* 2007;52:3. *J Forensic Sci*, 52(6), 1417; Discussion 1418.
- Bowers, C.M. & Johansen, R.J. (2002) Digital imaging methods as an aid in dental identification of human remains. *J Forensic Sci*, 47(2), 354-359.
- Bowers, C.M. & Johansen, R.J. (2003) *Digital Analysis of Bite Mark Evidence using Adobe Photoshop*. Forensic Imaging Services, Santa Barbara:.
- Bowers, C.M. & Pretty, I.A. (2009) Expert disagreement in bitemark casework. *J Forensic Sci*, 54(4), 915-918.
- Boyce, R.N. (1963) Judicial Recognition of Scientific Evidence in Criminal Cases. *Utah Law Review*, 8, 313, 314.
- Boychuk Duchscher, J.E. & Morgan, D. (2004) Grounded theory: reflections on the emergence vs. forcing debate. *J Adv Nurs*, 48(6), 605-612.
- Brim v State* 695 So.2d 268 (Fla. 1997)
- Brown, E.D., Snider, B. & Svilik, V., Ruling on Reliability and Relevance in a *Daubert* Hearing: The Methodology-Conclusion Debate and Other Issues. <http://cyber.law.harvard.edu/daubert/ch4.htm>, accessed on 16 June 2009,
- Brown, T.T. & Goryakin, Y. (2009) The oral health status of adults 65 and older in California:1995-2006. *J Calif Dent Assoc*, 37(8), 571-578.
- Bryant, A. & Charmaz, K. (Eds.) (2007) *The sage handbook of grounded theory* Sage Publications Ltd, Thousand Oaks, CA.
- Budowle, B. et al. (2009) A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *Journal of Forensic Sciences*, 54(4), 12.
- Budowle, B., Buscaglia, J. & Perlman, R.S., Review of the scientific basis for friction ridge comparisons as a means of identification: Committee findings and recommendations. [http://www.fbi.gov/hq/lab/fsc/backissu/jan2006/research/2006\\_01\\_research02.htm](http://www.fbi.gov/hq/lab/fsc/backissu/jan2006/research/2006_01_research02.htm), accessed on 21 Aug 2011.

- Budowle, B., Chakraborty, R., Carmody, G. & Monson, K. (2001) Reply to Weir. *For Sci Communications*, 3(1).
- Burbidge, R.J. (2001) Junk Science: Medicine and Law. *Australian Law Journal*, 75, 761.
- Bush, M.A., Bush, P.J. & Sheets, H.D. (2011) A study of multiple bitemarks inflicted in human skin by a single dentition using geometric morphometric analysis. *Forensic Sci Int*, In press.
- Bush, M.A., Cooper, H.I. & Dorion, R.B. (2010a) Inquiry into the scientific basis for bitemark profiling and arbitrary distortion compensation. *J Forensic Sci*, 55(4), 976-983.
- Bush, M.A., Miller, R.G., Bush, P.J. & Dorion, R.B. (2009) Biomechanical factors in human dermal bitemarks in a cadaver model. *J Forensic Sci*, 54(1), 167-176.
- Bush, M.A., Thorsrud, K., Miller, R.G., Dorion, R.B. & Bush, P.J. (2010b) The response of skin to applied stress: investigation of bitemark distortion in a cadaver model. *J Forensic Sci*, 55(1), 71-76.
- Bush, P. Personal Communication. 2011
- Butera v Department of Public Prosecutions (Vic)* (1987) 164 CLR 180
- Byrd, J.S. (2006) Confirmation bias, ethics, and mistakes in forensics. *J Forensic Ident*, 56(4), 511.
- Cadbury Schweppes Pty Ltd v Darrell Lea Chocolate Shops Pty Ltd* 228 ALR 719
- Cameron, J.M., Grant, J.H. & Ruddick, R. (1973) Ultraviolet Photography in Forensic Medicine. *J Focus Photography*, 1.
- Capra, D.J. (1997) The Daubert Puzzle, *Ga. L. Rev.*, 32, 699.
- Casley-Smith v FS Evans and Sons Pty Ltd (No.1)* 49 SASR 314
- Caudill, D.S. & Redding, R.E. (2000) Junk Philosophy of Science: The Paradox of Expertise and Interdisciplinarity in Federal Courts Symposium: The Law's Scientific Revolution. *Wash. & Lee L. Rev.*, 57, 685.
- Caudill, D.S. & LaRue, L.H. (2006) *No Magic Wand: the idealization of science in law*. Rowman & Littlefield Publishers, Lanham, Md.
- Cecchi, R. (2010) Estimating wound age: looking into the future. *Int J Legal Med*, 124(6), 523-536.
- Cecil, J.S. (2005) Ten years of judicial gatekeeping under Daubert. *Am J Public Health*, 95 Suppl 1, S74-80.
- Cecil, J.S. & Willging, T.E. (1995) Accepting Daubert's Invitation: Defining a Role for Court-Appointed Experts in Assessing Scientific Validity. *Emory L J*, 43, 995.
- Chamberlain v The Queen (No. 2)* (1984) 153 CLR 521
- Champod, C. (2000) Identification and Individualisation - Overview and Meaning of ID. In *Encyclopaedia of Forensic Sciences (Vol III)*, (Eds, Siegal, J.A., Saukko, P.J. & Knupfer, G.C.) Academic Press, London, pp. 1077-1084.
- Champod, C. & Evett, I.W. (2001) A probabilistic approach to fingerprint evidence. *J Forensic Ident*, 101-122.
- Champod, C. & Margot, P. (1996) Computer assisted analysis of minutiae occurrences on fingerprints. In *Proceedings of the International Symposium for Fingerprint Detection and Identification*, (Eds, Almog, J. & Springer, E.) p. 305.
- Charles, C.M. (1995). *Introduction to educational research* (2nd ed.). Longman, San Diego, CA.
- Charmaz, K. (2000) Grounded Theory: objectivist and constructivist methods. In *Handbook of Qualitative Research*, (Eds, Denzin, N.K. & Lincoln, Y.S.) Sage, Thousand Oaks, CA, pp. 509-535.
- Charmaz, K. (2003) Qualitative interviewing and grounded theory analysis. *Inside interviewing: New lenses, new concerns*, 311-330.
- Charmaz, K. (2007) Constructionism and the grounded theory method. *Handbook of constructionist research*, 397-412.

- Chen, J. & Moon, Y.S. (2008) The statistical modelling of fingerprint minutiae distribution with implications for fingerprint individuality studies. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008
- Cheng, E.K. & Yoon, A.H. (2005) Does Frye or Daubert Matter? A Study of Scientific Admissibility Standards. *Va L Rev*, 91, 471.
- Cheng, E.K.Y., Albert H. (2005) Does Frye or Daubert Matter - A Study of Scientific Admissibility Standards Essay. *Va. L. Rev.*, 91, 471.
- Chesebro, K.J. (1994) Taking *Daubert's* "Focus" Seriously: the Methodology/Conclusion Distinction. *Cardozo L rev*, 15, 1745.
- Christensen, A.M. (2005) Assessing the variation in individual frontal sinus outlines. *Am J Phys Anthropol*, 127(3), 291-295.
- Cicchetti (1981) Testing the Normal Approximation and Minimal Sample Size Requirements of Weighted Kappa When the Number of Categories is Large. *Appl Psychological Measurement*, 5(1), 101-104.
- Claar v Burlington NRR* 29 F.3d 499 (9th Cir. 1994)
- Clark, D.H. (1992) *Practical forensic odontology*. Wright, Oxford ; Boston.
- Clark v Ryan* (1960) 103 CLR 486
- Clarke, A. (2005) *Situational analysis: Grounded theory after the postmodern turn*. Sage Publications, Inc, San Francisco.
- Clement, J.G. (2011) Bite Marks. In *Current Practice in Forensic Medicine*, (Eds, Gall, J.A.M. & Payne-James, J.) Wiley-Blackwell, Chichester, pp. 291-308.
- Clement, J.G. & Blackwell, S.A. (2010), Is current bite mark analysis a misnomer? *Forensic Sci Int*, 201, 33-37.
- Close, F.E. (2009) *Nothing: a very short introduction*. Oxford University Press, Oxford ; New York.
- Cole, S.A. (2001) *Suspect Identities: A history of fingerprinting and criminal identification*. Harvard University Press, Cambridge, MA.
- Cole, S.A. (2004a) Grandfathering Evidence: Fingerprint Admissibility Rulings from Jennings to Llera Plaza and Back Again. *Am. Crim. L. Rev.*, 41, 1189.
- Cole, S.A. (2004b) More than Zero: Accounting for Error in Latent Fingerprint Identification Criminology. *J. Crim. L. & Criminology*, 95, 985.
- Cole, S.A. (2005) "Implicit Testing": Can Casework Validate Forensic Techniques? *Jurimetrics*, 46, 117.
- Cole, S.A. (2006) Prevalence and Potential Causes of Wrongful Conviction by Fingerprint Evidence, *The. Golden Gate University Law Review*, 37, 39.
- Cole, S.A. (2007) Toward Evidence-Based Evidence: Supporting Forensic Knowledge in the Post-Daubert Era. *Tulsa L Rev*, 43(2), 101.
- Cole, S.A. (2009) Forensics without uniqueness, conclusions without individualisation: the new epistemology of forensic identification. *Law Probability and Risk*, 8(3), 233-255.
- Collins, E.R. & Stone, R.S. (2005) How Unique Are Impressed Toolmarks? An empirical study of 20 worn hammer faces. *AFTE J*, 37(4), 252-295.
- Collins, J.M. & Jarvis, J. (2009) The Wrongful Conviction of Forensic Science. *For Sci Pol & Mgmt*, 1(1), 17-31.
- Commissioner for Government Transport v Adamcik* (1961) 106 CLR 292
- Committee on the Evaluation of the Reference Manual on Scientific Evidence (2009) Evaluation of the Reference Manual on Scientific Evidence: Letter Report.
- Commonwealth v Patterson* 445 Mass 626; 840 N.E. 2d.12; (Mass 2005)

- Conley, J.M.G., Scott W. (2005) Scientific Evidence in the State Courts: Daubert and the Problem of Outcomes. *Judges J.*, 44, 6.
- Cooley, C.M. (2003) Psychological Influences and the State Employed Forensic Examiner: How to Elicit Evidence Concerning Observer Effect Errors through Cross-Examination and Discovery. *Illinois Association of Criminal Defense Lawyers Newsletter*, 2003.
- Corbin, J.M. & Strauss, A.L. (2008) *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Inc., Thousand Oaks, CA.
- Cornell University Law School, US Constitution Annotated. Available at: <http://www.law.cornell.edu/anncon/>, accessed on 17 July, 2009.
- Crime Lab Report* (2009) Forensic Pattern Identification: A history lesson, and some advice, for Saks and Faigman. 21 Jan 2009. Available at: <http://crimelabreport.com>
- Cross on Evidence. Available at: <https://www.lexisnexis.com/au>, accessed on 23 April, 2009.
- Dailey, J.C. & Bowers, C.M. (1997) Aging of bitemarks: a literature review. *J Forensic Sci*, 42(5), 792-795.
- Daubert Tracker. Available at: <http://www.dauberttracker.com>, last accessed on 04 Jan, 2010.
- Daubert v Merrell Dow Pharmaceuticals Inc.* (1993) 509 US 579
- Daubert v Merrell Dow Pharmaceuticals Inc. II* 43 F.3d 1311 (9th Cir. 1995)
- Davies, G.L. (2004) Recent Australian Developments: A Response to Peter Heerey. *Civil Justice Quarterly*, 23, 396-399.
- Dawkins, R. (1976) *The Selfish Gene*. Oxford University Press, New York.
- Dean-Willcocks v Commonwealth Bank of Australia* 45 ACSR 564
- Deason, E.E. (1998) Court-Appointed Expert Witnesses: Scientific Positivism Meets Bias and Deference. *Or. L. Rev.*, 77, 59.
- DeCoux, E.L. (2007) Admission of Unreliable Expert Testimony Offered by the Prosecution: What's Wrong with Daubert and How to Make It Right, The. *Utah L. Rev.*, 2007, 131.
- Deitch, A. (2009) An Inconvenient Tooth: Forensic odontology is an inadmissible junk science when it is used to "match" teeth to bitemarks in skin. *Wisconsin L Rev*, 2009(5), 1205-1236.
- Denzin, N.K. & Lincoln, Y.S. (2011) *The Sage handbook of qualitative research*. Sage, Thousand Oaks, CA.
- Deputy v Lehman Bros., Inc.* 345 F.3d 494 (7th Cir. 2003)
- DeVore, D.T. (1971) Bite marks for identification? A preliminary report. *Med Sci Law*, 11(3), 144-145.
- Diamond, S.S. (2008) How Jurors Deal With Expert Testimony and How Judges Can Help. *J L & Pol'y*, 16, 47.
- Dixon, L.S. & Gill, B.P. (2001) *Changes in the standards for admitting expert evidence in federal civil cases since the Daubert decision*. Rand Institute for Civil Justice, Santa Monica, CA.
- Dobbin, S.A., Gatowski, S.I., Eyre, R.J., Dahir, V.B., Merlino, M.L. & Richardson, J.T. (2007) Federal and State Trial Judges on the Proffer and Presentation of Expert Evidence. *Justice System Journal*, 28(1), 1-15.
- Dolinak, D., Lew, E. & Matshes, E. (2005) Forensic Photography. In *Forensic Pathology: principles and practices*, Elsevier Inc, San Diego, pp. 631-636.
- Dorion, R.B. (1987) Transillumination in bite mark evidence. *J Forensic Sci*, 32(3), 690-697.
- Dorion, R.B.J. (2005) *Bitemark evidence*. Marcel Dekker, New York.
- Dressler, J., Bachmann, L., Koch, R. & Muller, E. (1999) Estimation of wound age and VCAM-1 in human skin. *Int J Legal Med*, 112(3), 159-162.

- Drinnan, A.J. & Melton, M.J. (1985) Court presentation of bite mark evidence. *International dental journal*, 35(4), 316.
- Dror, I.E. (2009) On proper research and understanding of the interplay between bias and decision outcomes. *Forensic Sci Int*, 191, e-17-e18.
- Dror, I.E. & Charlton, D. (2006) Why experts make errors. *J Forensic Ident*, 56(4), 600.
- Dror, I.E., Charlton, D. & Peron, A.E. (2006) Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci Int*, 156(1), 74-78.
- Dror, I.E., Péron, A.E., Hind, S. & Charlton, D. (2005) When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, 19(6), 799-810.
- Dror, I.E. & Rosenthal, R. (2008) Meta-analytically Quantifying the Reliability and Biasibility of Forensic Experts. *J Forensic Sci*, 53(4), 900-903.
- Edmond, G. Interview with author. 2009
- Edmond, G. (2000a) Deflating Daubert: Kumho Tire Co. v Carmichael and the Inevitability of General Acceptance (FRYE). *U.N.S.W.L.J.*, 23, 38.
- Edmond, G. (2000b) Judicial Representations of Scientific Evidence. *The Modern Law Review*, 63, 216.
- Edmond, G. (2003) After Objectivity: Expert Evidence and Procedural Reform. *Sydney L. Rev.*, 25, 131.
- Edmond, G. (2005) Judging Surveys: Experts, Empirical Evidence and Law Reform. *Fed. L. Rev.*, 33, 95.
- Edmond, G. (2007) Supersizing Daubert Science for Litigation and Its Implications for Legal Practice and Scientific Research Symposium: Expertise in the Courtroom: Scientists and Wizards - Panel Three: Science, Scientists and Ethics. *Vill. L. Rev.*, 52, 857.
- Edmond, G. (2008a) Specialised Knowledge, the Exclusionary Discretions and Reliability: Reassessing Incriminating Expert Opinion Evidence. *U.N.S.W.L.J.*, 31, 1.
- Edmond, G. (2008b) Pathological Science? Demonstrable Reliability and Expert Forensic Pathology Expertise. *UNSW Law Research Paper No.2008-6*,
- Edmond, G. (2009) Merton and the Hot-Tub. *Law & Contemp. Probs.*, 72, 159.
- Edmond, G. & Mercer, D. (1997) Keeping 'Junk' history, philosophy and sociology of science out of the courtroom: Problems with the reception of *Daubert v Merrell Dow Pharmaceuticals Inc.* *University of NSW Law Journal*, 20(1), 48.
- Edmund, G. & Mercer, D. (2004) Daubert and the Exclusionary Ethos: The Convergence of Corporate and judicial Attitudes towards the Admissibility of Expert Evidence in Tort Litigation Special Issue on Toxic Torts and Environment Justice. *Law & Pol'y*, 26, 231.
- Edwards, H.T. (2009) Statement before the United States Senate Committee on the Judiciary. *Strengthening Forensic Science in the United States: A Path Forward*, 2009
- Ege v Yukins* 2005 US Dist. LEXIS 15953
- Einstein, C. (2000) Understanding the Evidence Act 1995 - A Daunting Task! *Civil Issues in Evidence*, Judicial Commission of New South Wales, 2000
- Eisenhardt, K.M. & Graebner, M.E. (2007) Theory Building from Cases: Opportunities and Challenges. *Academy of Management J*, 50(1), 25-32.
- Ellram, L.M. (1996) The Use of the Case Study Method in Logistics Research. *J Business Logistics*, 17(2), 93-138.
- English, W.R., Robison, S.F., Summitt, J.B., Oesterle, L.J., Brannon, R.B. & Morlang, W.M. (1988) Individuality of human palatal rugae. *Journal of forensic sciences*, 33(3), 718.

- Epstein, R. (2001) Fingerprints Meet Daubert: The Myth of Fingerprint Science Is Revealed. *S. Cal. L. Rev.*, 75, 605.
- Erdmann, M.V., Caldwell, R.L. & Moosa, M.K. (1998) Indonesian 'king of the sea' discovered. *Nature*, 395, 335.
- Estate of Kenneth Griffin v Hickson* 2002 US Dist LEXIS 8568
- Eubanks v Hale* (Ala. 1999) 752 So. 2d 113
- Evetts, I.W., Jackson, G., Lambert, J.A. & McCrossan, S. (2000) The impact of the principles of evidence interpretation on the structure and content of statements. *Sci Justice*, 40(4), 233-239.
- Evidence Act 1995 (Cth).
- Faigman, D.L. (1995) Making the Law Safe for Science: A proposed rule for the admission of expert testimony. *Washburn L.J.*, 35, 401.
- Faigman, D.L. (2000) Law's Scientific Revolution: Reflections and Ruminations on the Law's Use of Experts in Year Seven of the Revolution, The Symposium: The Law's Scientific Revolution. *Wash. & Lee L. Rev.*, 57, 661.
- Faigman, D.L. (2007) Anecdotal Forensics, Phrenology and Other Abject Lessons from the History of Science. *Hastings L.J.*, 59, 979-1000.
- Faigman, D.L. (2010) Evidentiary Incommensurability: A Preliminary Exploration of the Problem of Reasoning From General Scientific Data to Individualized Legal Decision-Making. *Brooklyn L. Rev.*, 75, 1115-1489.
- Faigman, D.L. Interview with author. (28 Oct 2009)
- Faigman, D.L., Saks, M.J., Sanders, J. & Cheng, E.K. (2007) *Modern Scientific Evidence: The Law and Science of Expert Testimony*. Thomson/West, Minnesota.
- Farrell, M.G. (1994) Daubert v Merrell Dow Pharmaceuticals, Inc.: Epistemology and Legal Process. *Cardozo L Rev*, 15, 2183.
- Federal Judicial Center. (1994) *Reference Manual on Scientific Evidence*. Federal Judicial Center.
- Federal Rules of Evidence* 2008. Federal Evidence Review, Arlington VA.
- Fenner, G.M. (1996) The Daubert Handbook: The Case, its Essential Dilemma, and its Progeny. *Creighton L Rev*, 29, 939.
- Ferriola, T.J., Scientific Principles of Friction Ridge Analysis & Applying Daubert to Latent Fingerprint Identification. Available at: <http://www.clpex.com/Articles/ScientificPrinciplesbyTomFerriola.htm>, accessed on 12 Nov, 2009.
- Fisher, J. (2008) *Forensics under fire : are bad science and dueling experts corrupting criminal justice?* Rutgers University Press, New Brunswick, N.J.
- Fleiss, J.L. (1981) The Measurement of Interrater Agreement. In *Statistical Methods for Rates and Proportions*, John Wiley & Sons Inc, New York,
- Flora, G., Tuceryan, M. & Blitzer, H. (2009) Forensic bite mark identification using image processing methods. *Proceedings of the 2009 ACM symposium on Applied Computing*, 2009
- Forrest, R. (2004) Context-free forensic science. *Science & Justice*,
- Foster, K.R. & Huber, P.W. (1999) *Judging Science : Scientific Knowledge and the Federal Courts*. MIT Press, Cambridge, Mass.
- Fradella, H.F., O'Neill, L. & Fogarty, A. (2003) Impact of Daubert on Forensic Science, The. *Pepp. L. Rev.*, 31, 323.
- Freckelton, I. Interview with Author. 2009
- Freckelton, I. (1997) Judicial Attitudes Towards Scientific Evidence: The Antipodean Experience. *U.C. Davis L. Rev.*, 30, 1137.



- Freckelton, I., Reddy, P. & Selby, H. (1999a) *Australian Judicial Perspectives on Expert Evidence: An Empirical Study*. Australasian Institute of Judicial Administration Inc., Melbourne.
- Freckelton, I., Reddy, P. & Selby, H. (1999b) *Australian Judicial Perspectives on Expert Evidence: An Empirical Study - Summary of Key Findings and Outcomes*. Australasian Institute of Judicial Administration Inc., Broome, Western Australia.
- Freckelton, I., Reddy, P. & Selby, H. (2000) *Australian Magistrates' Perspectives on Expert Evidence: A Comparative Study - Summary of Key Findings and Outcomes*. Australian Institute of Judicial Administration, Melbourne.
- Freckelton, I. & Selby, H. (2009) *Expert Evidence: Law, Practice, Procedure and Advocacy*. Thomson Reuters, NSW, Australia.
- Freeman, R.A. (1986) Consecutively rifled polygon barrels. *AFTE J*, 18(3), 15-52.
- Fricke, G. (1997) Trial by Jury. *Research Paper 11 1996-97*, Parliament of Australia, ACT.
- Frye v United States* 293 F. 1013 (D.C.Cir., 1923)
- Gable, J.D. & Wilkinson, M.D. (2007) Good Science Gone Bad: How the Criminal Justice System Can Redress the Impact of Flawed Forensics Symposium. *Hastings L.J.*, 59, 1001-1030.
- Gall, J.A.M. & Payne-James, J. (2011) *Current practice in forensic medicine*. Wiley-Blackwell, Chichester, West Sussex, UK.
- Galton, F. (1892) *Finger prints*. Macmillan and Co., London, New York.
- Garrett, B.L. (2008) Judging Innocence. *Columbia Law Review*, 108, 55.
- Garrett, B.L. & Neufeld, P.J. (2009) Invalid Forensic Science and Wrongful Convictions. *Va L Rev*, 95, 1-97.
- Gasson, S. (2003) Rigor In Grounded Theory Research: An Interpretive Perspective on Generating Theory From Qualitative Field Studies. In Whitman, M., Woszczynski, A (2003) *Handbook of Information Research*. IGI Global, PA.
- Gatowski, S.I., Dobbin, S.A., Richardson, J.T., Ginsburg, G.P., Merlino, M.L. & Dahir, v (2001) Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-Daubert world. *Law Hum Behav*, 25(5), 433-458.
- Gauch, H.G. (2003) *Scientific method in practice*. Cambridge University Press, New York.
- General Electric Co. v Joiner* (1997) 522 US 136
- German, E., Latent Print Examination. Available at: <http://onin.com/fp/>, accessed on 01 Oct, 2009.
- Gershman, B.L. (2003) Misuse of Scientific Evidence by Prosecutors. *Okla. City U. L. Rev.*, 28, 17.
- Giannelli, P.C. (1980) The Admissibility of Novel Scientific Evidence: *Frye v United States* a Half-Century Later. *Columbia Law Review*, 80, 1197.
- Giannelli, P.C. (1983) Background Paper Prepared for the National Conference of Lawyers and Scientists. *F.R.D.*, 99.
- Giannelli, P.C. (2002) Fingerprints Challenged Scientific Evidence. *Crim. Just.*, 17, 33.
- Giannelli, P.C. (2003) The "Science" of Wrongful Convictions. *Crim. Just.*, 18, 55-59.
- Giannelli, P.C. (2007a) Bite Mark Evidence. *Litigation*, 33(3), 19.
- Giannelli, P.C. (2007b) Confirmation bias. *Crim. Just.*, 22(3), 60-62.
- Giannelli, P.C. (2009) Daubert "Factors". *Crim. Just.*, 23, 42.
- Gimbel, S. (2011) *Exploring the Scientific Method*. University of Chicago Press, Chicago.
- Glaser, B.G. (1992) *Basics of Grounded Theory Analysis: Emergence Vs Forcing*. Sociology Press, Mill Park.
- Glaser, B.G. (2002) Constructivist grounded theory? *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 3(3), 2002

- Glaser, B.G., Strauss, A.L. & Strutzel, E. (1968) The discovery of grounded theory; strategies for qualitative research. *Nursing Research*, 17(4), 364.
- Glenn, J.C. & Gordon, T.J. (2003) *Futures research methodology*. American Council for the United Nations University, the Millennium Project,
- Gless, A.G. (1995) Some Post-Daubert Trial Tribulations of a Simple Country Judge: Behavioural Science Evidence in Trial Courts. *Behav Sci Law*, 13, 261-291.
- Golafshani, N. (2003) Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, 8(4), 597-607.
- Gold, M.H., Roenigk, H.H., Smith, E.S. & Pierce, L.J. (1989) Human bite marks. Differential diagnosis. *Clin Pediatr (Phila)*, 28(7), 329-331.
- Goodwin, R.J. (2000) Hidden Significance of Kumho Tire Co. v Carmichael: A Compass for Problems of Definition and Procedure Created by Daubert v Merrell Dow Pharmaceuticals, Inc., The. *Baylor L. Rev.*, 52, 603.
- Gottesman, M.H. (1998) From Barefoot to Daubert to Joiner: Triple Play or Double Error Symposium--Courts on Trial. *Ariz. L. Rev.*, 40, 753.
- Government of the Virgin Islands v Austin Jacobs* Crim. App. No. 2001-65, Memorandum Opinion (Unpublished)
- Government of the Virgin Islands v Austin Jacobs* 53 Fed.Appx. 651, 2002 WL 31887857 (C.A.3 (Virgin Islands))
- Graham, D. (1973) *The use of X-ray techniques in forensic investigations*. Churchill Livingstone, Edinburgh.
- Grann, D. (2009) Trial by Fire - Did Texas execute an innocent man? *The New Yorker*, 07 Sep 2009, Reporter at Large, pp 42-63.
- Green, M.D. (1992) Expert witnesses and sufficiency of evidence in toxic tort substances litigation: The legacy of Agent Orange and Bendectin litigation. *Nw U L Rev*, 86, 643.
- Grellner, W. & Madea, B. (2007) Demands on scientific studies: vitality of wounds and wound age estimation. *Forensic Sci Int*, 165(2-3), 150-154.
- Grieve, D. (1996) Possession of Truth. *J Forensic Identification*, 46(5), 521-528.
- Groscup, J.L. (2004) Judicial Decision Making About Expert Testimony in the Aftermath of Daubert and Kumho. *Journal of Forensic Psychology Practice*, 4(2), 57 - 66.
- Groscup, J.L., Penrod, S.D., Studebaker, C.A., Huss, M.T. & O'Neil, K.M. (2002) The Effects of Daubert on the Admissibility of Expert Testimony in State and Federal Criminal Cases. *Psych Pub Pol L*, 8(4), 339-372.
- Gustafson, G. (1966) *Forensic odontology*. American Elsevier Pub. Co., New York.
- Haber, L. & Haber, R.N. (2003) Error Rates for Human Latent Fingerprint Examination. In *Advances in Automatic Fingerprint Recognition*, (Ed, Ratha, N.K.) Springer-Verlag, New York,
- Haber, L. & Haber, R.N. (2008) Scientific validation of fingerprint evidence under Daubert. *Law Probability and Risk*, 7(2), 127-141.
- Hall, L.J. & Player, E. (2008) Will the introduction of an emotional context affect fingerprint analysis and decision making? *Forensic Science International*, 181(1), 36-39.
- Hamby, J., Brundage, D.J. & Thorpe, J.W. (2009) The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries. *AFTE J*, 41(2), 99-111.
- Harmon, R., Budowle, B., Langenburg, G., Houck, M. & Kelly, J. (2006) Questions About Forensic Science. *Science*, 311(5761), 607-610.
- Harris, A.M., Wood, R.E., Nortje, C.J. & Thomas, C.J. (1987) The frontal sinus: forensic fingerprint? A pilot study. *The Journal of Forensic Odonto-Stomatology*, 5(1), 9.

- Harris, R.C. (2008) *Black robes, white coats : the puzzle of judicial policymaking and scientific evidence*. Rutgers University Press, New Brunswick, N.J.
- Harvard Law Review (1994) Confronting the New Challenges of Scientific Evidence Developments in the Law. *Harv. L. Rev*, 108, 1481.
- Hatcher, J.S., Jury, F.J. & Weller, J. (1957) Firearms Investigation. *Identification and Evidence, The Stackpole Company, Harrisburg (Pennsylvania)*,
- Heinzerling, L. (2006) Doubting Daubert. *J.L. & Pol'y*, 14, 65.
- Henson, B. & Politovich, M., Let it Snow, Let it Snow, Let it Snow.  
<http://www.proquestk12.com/curr/snow/snow395/snow395.htm> Participant #snow, accessed on 20 Sep 2010.
- Hernandez-Cueto, C., Girela, E. & Sweet, D.J. (2000) Advances in the diagnosis of wound vitality: a review. *Am J Forensic Med Pathol*, 21(1), 21-31.
- Heydon, J.D., Commentary on Justice Einstein's Paper. Available at:  
[http://www.lawlink.nsw.gov.au/lawlink/supreme\\_court/ll\\_sc.nsf/pages/SCO\\_speech\\_heydon\\_201100](http://www.lawlink.nsw.gov.au/lawlink/supreme_court/ll_sc.nsf/pages/SCO_speech_heydon_201100), accessed on 05 July, 2009.
- HG v The Queen* Transcript of Proceedings S67/1998 HCAT (8 September 1998)
- HG v The Queen* 197 CLR 414 (1999)
- Hocking v Bell* 1945 71 CLR 430 at 440
- Hoepfl, M.C. (1997) Choosing Qualitative Research: A Primer for Technology Education Researchers. *J Technology Education*, 9(1), 47-63.
- Holden, J.D. (2001) Hawthorne effects and research into professional practice. *J Eval Clin Pract*, 7(1), 65-70.
- House of Commons Science and Technology Committee (2005) Forensic Science on Trial. *Seventh Report of Session 2004-05*,
- Hrabosky, K.M. (1999) *Kumho Tire v Carmichael*: Stretching Daubert beyond Recognition Casenote. *Geo. Mason L. Rev.*, 8, 203.
- Hunt, A.C. (2007) Ring-resolution of bruises-a little recognised phenomenon. *Journal of Forensic and Legal Medicine*, 14(2), 85-86.
- Hurrelbrink v State* 2001 Tex. App. LEXIS 2195 (Unreported)
- Hutchison, C.T. & Ashby, D.S. (1994) *Daubert v Merrell Dow Pharmaceuticals, Inc.*: Redefining the Bases for Admissibility of Expert Scientific Testimony. *Cardozo L Rev*, 15, 1875.
- Idoport Pty Ltd v National Australia Bank Limited* [1] NSWSC 828 (19 August 1999)
- Imwinkelreid, E.J. (1993) The Next Step After Daubert: Developing a Similarly Epistemological Approach to Ensuring the Reliability of Nonscientific Expert Testimony. *Cardozo L Rev*, 15, 15.
- Imwinkelreid, E.J. (1999) The Taxonomy of Testimony Post-*Kumho*: Refocussing on the Bottomlines of Reliability and Necessity. *Cumb. L. Rev.*, 30, 185.
- Imwinkelried, E.J. (1992) Attempts to Limit the Scope of the Frye Standard for the Admission of Scientific Evidence: Confronting the Real Cost of the General Acceptance Test. *Behav Sci Law*, 10, 441.
- Imwinkelried, E.J. (1995) Evidence Law Visits Jurassic Park: The Far-Reaching Implication of the Daubert Court's Recognition of the Uncertainty of the Scientific Enterprise. *Iowa L. Rev*, 81, 55.
- Inman, K. & Rudin, K. (1997) *An Introduction to Forensic DNA Analysis*. CRC Press, Boca Raton, Florida.
- International Association of Arson Investigators Amicus Brief. *Michigan Millers Mutual Insurance Corporation v Benfield*, (1998) 140 F.3d 915 (11th Cir.)
- Interpol. (2011) *Disaster Victim Identification Guide*. Interpol.

- Jain, A.K., Prabhakar, S. & Pankanti, S. (2002) On the similarity of identical twin fingerprints. *Pattern Recognition*, 35(11), 2653-2663.
- Jakobsen, J.R. & Keiser-Nielsen, S. (1981) Bite mark lesions in human skin. *Forensic Sci Int*, 18(1), 41-55.
- James, H. & Cirillo, G.N. (2004) Bite mark or bottle top? *J Forensic Sci*, 49(1), 119-121.
- James, S.H. & Nordby, J.J. (2009) *Forensic science : an introduction to scientific and investigative techniques*. CRC Press/Taylor & Francis Group, Boca Raton, FL.
- Jango v Northern Territory (No 4)* (2004) 214 ALR 608
- John Philip Reynolds v Raymond John Carroll*. Magistrates Court, Ipswich, 11th July 1984, Nos C/F 996-1014/84,
- Jonakait, R.N. (1990) The Supreme Court, Plain Meaning, and the Changed Rules of Evidence. *Tex L Rev*, 68, 745.
- Jonakait, R.N. (1991) Forensic Science: The Need for Regulation. *Harv J Law Tech*, 4, 109-191.
- Jonakait, R.N. (1993) The Meaning of *Daubert* and What That Means for Forensic Science. *Cardozo L Rev*, 15, 2103.
- Journal Citation Reports. Available from: [www.isiknowledge.com](http://www.isiknowledge.com), accessed on 30 Sep, 2009.
- Kaye, D.H. (2003a) Questioning a courtroom proof of the uniqueness of fingerprints. *Int Stat Rev*, 71(3), 521-533.
- Kaye, D.H. (2003b) The Non-Science of Fingerprinting: *United States v Llera-Plaza*. *QLR*, 21, 1073.
- Kaye, D.H. (2004) On Falsification and Falsifiability: The First Daubert Factor and the Philosophy of Science Reflections. *Jurimetrics*, 45, 473.
- Kaye, D.H. (2009) Identification, individualization and uniqueness: What's the difference? *Law Probability and Risk*, 8(1), 55-66.
- Kaye, D.H. (2010) Probability, Individualisation and Uniqueness in Forensic Science Evidence: Listening to the Academies. *Brooklyn L Rev*, In Press,
- Keele, D. (2007) Ideological Effects in Published Versus Unpublished Judicial Opinions. *Annual Meeting of the Midwest Political Science Association*, 2007
- Keiser-Nielsen, S. (1975) Certainty v Probability in Dental Identification. *Second Continuing Education Course in Forensic Odontology*, November 1-2 1975
- Keiser-Nielsen, S. (1977) Dental Identification: Certainty V Probability. *Forensic Science*, 9, 87-97.
- Keiser-Nielsen, S. (1980) *Personal Identification by Means of the Teeth*. John Wright and Sons, Bristol.
- Keko v Hingle*. WL 508406 (E.D. La. July 8, 1999) (Unreported)
- Kelle, U. (2005) " Emergence" vs." Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(2), 2005
- Kemp, A., Maguire, S.A., Sibert, J., Frost, R., Adams, C. & Mann, M. (2006) Can we identify abusive bites on children? *Arch Dis Child*, 91(11), 951.
- Kent, C.S. (1999) Daubert Readiness of Texas Judiciary: A Study of the Qualifications, Experience, and Capacity of the Member of the Texas Judiciary to Determine the Admissibility of Expert Testimony under the Daubert, Kelly, Robinson, and Havner Tests. *Tex. Wesleyan L. Rev.*, 6, 1.
- Kieser, J.A. & Bernal, V. (2007) Author's Response to Dr Bowers. *J Forensic Sci*, 52(6), 1418.
- Kieser, J.A., Bernal, V., Neil Waddell, J. & Raju, S. (2007) The uniqueness of the human anterior dentition: a geometric morphometric analysis. *J Forensic Sci*, 52(3), 671-677.
- Kinley, D. & Rose, A. (1999) The Quest for Truth: A Comparative Analysis of the Role of Experts in Litigation. *Aust J Forensic Sci*, 31, 5-18.

- Kirby, M.J. (2003) Expert Evidence: Causation, Proof and Presentation. *The Judicial Review*, 6, 131.
- Kirby, M.J. Interview with Author. 2010
- Klinkner, M. (2009) Forensic science expertise for international proceedings: an old problem, a new context and a pragmatic resolution. *Int'l J Evidence and Proof*, 13, 102.
- Koehler, J.J. (1996) On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *U. Colo. L. Rev.*, 67, 859.
- Koehler, J.J. (1997) Why DNA Likelihood ratios should account for error (even when a National Reserach Council report says they shouldn't). *Jurimetrics J*, 37, 425-437.
- Koehler, J.J. (2007) Fingerprint Error Rates and Proficiency Tests: What They are and Why They Matter Symposium. *Hastings L.J.*, 59, 1077.
- Kondo, T. (2007) Timing of skin wounds. *Leg Med (Tokyo)*, 9(2), 109-114.
- Kondo, T., Ohshima, T. & Eisenmenger, W. (1999) Immunohistochemical and morphometrical study on the temporal expression of interleukin-1alpha (IL-1alpha) in human skin wounds for forensic wound age determination. *Int J Legal Med*, 112(4), 249-252.
- Koppl, R. (2005) How to Improve Forensic Science. *Eur J Law and Economics*, 20, 255-286.
- Kouble, R.F. & Craig, G.T. (2004) A comparison between direct and indirect methods available for human bite mark analysis. *J Forensic Sci*, 49(1), 111-118.
- Koukoutchos, B.S. (1994) Solomon Meets Galileo (And Isn't Quite Sure What to do With Him). *Cardozo L Rev*, 15, 2237.
- Kovera, M.B. & McAuliff, B.D. (2000) The Effects of Peer Review and Evidence Quality on Judge Evaluations of Psychological Science: Are Judges Effective Gatekeepers? *J. App. Psychology*, 85(4), 574-586.
- Krafka, C., Dunn, M.A., Johnson, M.T., Cecil, J.S. & Miletich, D. (2002) Judge and Attorney Experiences, Practices, and Concerns Regarding Expert Testimony in Federal Civil Trials. *Psych Pub Pol L*, 8(3), 309-332.
- Krane, D.E., Doom, T.E., Mueller, L., Raymer, M.L., Shields, W.M. & Thompson, W.C. (2004) Commentary on: Budowle B, Shea B, Niezgoda S, Chakraborty r. CODIS STR loci data from 41 sample populations. *J Forensic Sci* 2001;46:453-489. *J Forensic Sci*, 49(6), 1388-9; author reply 1390-3.
- Krane, D.E. et al. (2008) Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *Journal of Forensic Sciences*, 53(4), 1006-07.
- Krane, D.E. et al. (2009) Commentary on: Budowle B, Bottrell MC, Bunch SG, Fram R, Harrison D, Meagher S, Oien CT, Peterson PE, Seiger DP, Smith MB, Smrz MA, Soltis GL, Stacey RB. A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. *J Forensic Sci*, 55(1), 273-274.
- Krauss, D.A. & Sales, B.D. (2001) The Effects of Clinical and Scientific Expert Testimony on Juror Decision Making in Capital Sentencing. *Psych Pub Pol L*, 7,
- Krug, K. (2007) The Relationship Between Confidence and Accuracy: Current thoughts of the literature and a new area of research. *Applied Psychology in Criminal Justice*, 3(1), 7-41.
- Kruszelnicki, K.S., Snowflakes Identical.  
<http://www.abc.net.au/science/articles/2006/11/13/1784760.htm>, accessed on 14 January, 2010.
- Kumho Tire Co Ltd v Carmichael* (1999) 526 US 137
- Kvale, S. (1983) The qualitative research interview: A phenomenological and a hermeneutical mode of understanding. *Journal of phenomenological psychology*,
- Kvale, S. (1994) Ten standard objections to qualitative research interviews. *Journal of Phenomenological Psychology*, 25(2), 147-173.
- Kwan, Q.Y. (1977) *Inference of Identity of Source*. Ph.D. thesis, University of California.

- La Morte, T.M. (2003) Sleeping Gatekeepers: *United States v Llera Plaza* and the Unreliability of Forensic Fingerprinting Evidence under Daubert Comment. *Alb. L.J. Sci. & Tech.*, 14, 171.
- Lagenburg, G., Champod, C. & Wertheim, B.A. (2009) Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons. *J Forensic Sci*, 54(3), 571-582.
- Langlois, N.E. & Gresham, G.A. (1991) The ageing of bruises: a review and study of the colour changes with time. *Forensic science international*, 50(2), 227.
- Lardizabal, P.P. (1995) Cartridge case study of the Heckler & Koch USP. *AFTE J*, 27(1), 49-51.
- Lasser, A.J., Warnick, A.J. & Berman, G.M. (2009) Three-dimensional comparative analysis of bite marks. *J Forensic Sci*, 54(3), 658-661.
- Law Report on ABC Radio National: Forensic Dentistry Bites Back.  
<http://www.abc.net.au/rn/lawreport/stories/2000/122913.htm>, accessed on 29 Dec, 2009.
- Laws, K. & McLeod, R. (2004) Case study and grounded theory: Sharing some alternative qualitative research methodologies with systems professionals. *Proceedings of the 22nd International Conference of the Systems Dynamics Society*, 2004
- Lederer, F.I. (1986) Resolving the Frye Dilemma - A Reliability Approach. *Jurimetrics*, 26, 240.
- Lee, H.C. & Gaensslen, R.E. (2001) *Advances in fingerprint technology*. CRC Press, Boca Raton, Fla.
- Leibniz, G.W. & Rescher, N. (1991) *G.W. Leibniz's Monadology (1714): an edition for students*. University of Pittsburgh Press, Pittsburgh, Pa.
- Leonard, J.D. (1999) *Removal: What Every Litigator Should Know*. Wichita Bar Association, Wichita, Kansas.
- Lesolang, R.R., Motloba, D.P. & Lalloo, R. (2009) Patterns and reasons for tooth extraction at the Winterveldt Clinic: 1998-2002. *SADJ*, 64(5), 214-5, 218.
- Letters Forum (1991) Forensic Labs. *Harv. J. L. & Tech.*, 5, 1.
- Levine, L.J., (1984) Bite marks in child abuse. In *Clinical Management of Child Abuse and Neglect*, (Eds, RG, S. & DC, B.) Quintessence, Chicago, pp. 53-59.
- Lincoln, Y.S. & Guba, E.G. (1985) *Naturalistic inquiry*. Sage Publications, Inc,
- Loftus, E.F. & Cole, S.A. (2004) Contaminated Evidence. *Science*, 304, 959.
- Louisell, D.W. & Mueller, C.B. (1977) *Federal Evidence*.
- MacDonald, D.G. & Laird, W.R.E. (1977) Bite marks in a murder case. *Int J Forensic Dent*, 3, 24-27.
- MacDonald, M. & Schreiber, R.S. (2001) Constructing and deconstructing: Grounded theory in a postmodern world. *Using grounded theory in nursing*, 35-54.
- MacFarlane, T.W., MacDonald, D.G. & Sutherland, D.A. (1974) Statistical Problems in Dental Identification. *J Forensic Sci Soc*, 14, 247-252.
- Maguire, S., Mann, M.K., Sibert, J. & Kemp, A. (2005) Can you age bruises accurately in children? A systematic review. *Arch Dis Child*, 90(2), 187-189.
- Makita (Australia) Pty Ltd v Sprowles* (2001) 52 NSWLR 705
- Mansfield, J.H. (1996) Scientific Evidence under Daubert. *St. Mary's L.J.*, 28, 1.
- Marriot, A., Harding, W., Devlin, N. & Benwell, G. (2001) The delivery of orthodontic care in New Zealand. Part 1: Analysis of a census of orthodontists. *NZDJ*, 97, 87-92.
- Martin-de las Heras, S., Valenzuela, A., Javier Valverde, A., Torres, J.C. & Luna-del-Castillo, J.D. (2007) Effectiveness of comparison overlays generated with DentalPrint software in bite mark analysis. *J Forensic Sci*, 52(1), 151-156.
- Martin-de-las-Heras, S. & Tafur, D. (2009) Comparison of simulated human dermal bite marks possessing three-dimensional attributes to suspected biters using a proprietary three-dimensional comparison. *Forensic Sci Int*, 190(1-3), 33-37.

- Martin, P.W., Finding and Citing the "Unimportant" Decisions of the US Courts of Appeals. <http://topics.law.cornell.edu/wex/papers/lir2007-1>, accessed on 22 June, 2009.
- Maskin, A. (1994) The Impact of Daubert on the Admissibility of Scientific Evidence: The Supreme Court Catches Up With a Decade of Jurisprudence. *Cardozo L Rev*, 15, 1929.
- Mayer, J. (1935) Identification by Sinus Prints. *Virginia Med. Monthly*, Dec, 517-519.
- McCarroll, J.E., Fullerton, C.S., Ursano, R.J. & Hermesen, J.M. (1996) Posttraumatic stress symptoms following forensic dental identification: Mt Carmel, Waco, Texas. *Am J Psychiatry*, 153, 778-782.
- McCarroll, J.E., Ursano, R.J., Fullerton, C.S., Liu, X. & Lundy, A. (2002) Somatic Symptoms in Gulf War Mortuary Workers. *Psychosomatic Medicine*, 64, 29-33.
- McClellan, P. (2007) Concurrent Expert Evidence. *Medicine and Law Conference: Keynote Address*, 2007
- McCormick, C.T. (1954) *Handbook of the Law and Evidence*.
- McCormick, M. (1982) Scientific Evidence: Defining a New Approach to Admissibility. *Iowa L Rev*, 67, 879.
- McFarland, H., Case law regarding Questioned Document Examination. [http://writeexam.com/case\\_law.php](http://writeexam.com/case_law.php), accessed on 12 Nov, 2009.
- McInnis, M. (1995) Expert Evidence and the Federal Courts: Current Developments. *International Institute of Forensic Studies "Experts and Lawyers: Surviving in the Brave New World"*, Broome, 1995
- McKenna, C.J., Haron, M.I. & Taylor, J.A. (1999) Evaluation of a bitemark using clear acrylic replicas of the suspect's dentition--a case report. *J Forensic Odontostomatol*, 17(2), 40-43.
- McLachlan, H. (1995) No two sets the same? Applying philosophy to the theory of fingerprints. *Philosopher*, 83(2), 12-18.
- McNamee, A.H. & Sweet, D. (2003) Adherence of forensic odontologists to the ABFO guidelines for victim evidence collection. *J Forensic Sci*, 48(2), 382-385.
- McNamee, A.H., Sweet, D. & Pretty, I. (2005) A comparative reliability analysis of computer-generated bitemark overlays. *J Forensic Sci*, 50(2), 400-405.
- McQuiston-Surrett, D. & Saks, M.J. (2008) Communicating opinion evidence in the Forensic Identification Sciences: Accuracy and Impact. *Hastings L.J.*, 59, 1159.
- Mead, R.A. (2001) Unpublished Opinions as the Bulk of the Iceberg: Publication Patterns in the Eighth and Tenth Circuits of the United States Court of Appeals. *Law Libr. J.*, 93, 589.
- Melnick, R.L. (2005) A Daubert motion: a legal strategy to exclude essential scientific evidence in toxic tort litigation. *Am J Public Health*, 95 Suppl 1, S30-4.
- Merlino, M.L., Murray, C.I. & Richardson, J.T. (2008) Judicial gatekeeping and the social construction of the admissibility of expert testimony. *Behav Sci Law*, 26(2), 187-206.
- Merlino, M.L., Springer, V.S., Kelly, J.S., Hammond, D., Sahota, E. & Haines, L. (2007) Meeting the Challenges of the *Daubert* Trilogy: Refining and Redefining the Reliability of Forensic Evidence. *Tulsa L Rev*, 43, 417.
- Merritt, D.J.B., James J. (2001) Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals. *Vand. L. Rev.*, 54, 69.
- Mervis, J. (1993) Supreme Court to judges: Start thinking like scientists. *Science*, 261, 22.
- Meuwly, D. (2006) Forensic individualisation from biometric data. *Science & Justice*, 46(4), 205-213.
- Miles, M.B., Huberman, A.M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). Sage Publications, Newbury Park, CA.
- Miller, L.S. (1987) Procedural bias in forensic science examinations of human hair. *Law and Human Behavior*, 11(2), 157-163.

- Miller, R.G., Bush, P.J., Dorion, R.B. & Bush, M.A. (2009) Uniqueness of the dentition as impressed in human skin: a cadaver model. *J Forensic Sci*, 54(4), 909-914.
- Mills, J., Bonner, A. & Francis, K. (2006) The development of constructivist grounded theory. *International Journal*, 5(March), 1-10.
- Mnookin, J.L. (2001) Fingerprint Evidence in an Age of DNA Profiling Symposium: DNA Lessons from the Past--Problems for the Future. *Brook. L. Rev.*, 67, 13-70.
- Mnookin, J.L. (2008) The validity of latent fingerprint identification: confessions of a fingerprinting moderate. *Law Probability and Risk*, 7(2), 111-118.
- Moenssens, A.A. (1999a) Panel III. Junk Science, Pre-Science and Developing Science. *National Conference on Science and Law*, 1999.
- Moenssens, A.A. (1999b) Meeting the Daubert Challenge to Handwriting Evidence. *Forensic Science Communications*, Oct 1999.
- Moenssens, A.A., Henderson, C.E. & Portwood, S.G. (2007) *Scientific evidence in civil and criminal cases*. Foundation Press Thomson/West, New York, NY.
- Moles, R.N., The Role and Function of the Expert Witness. <http://netk.net.au/Reports/ExpertWitness.asp>, accessed on 09 July, 2009.
- Moriarty, J.C. & Saks, M.J. (2005) Forensic Science: Grand Goals, Tragic Flaws and Judicial Gatekeeping. *Judges J.*, 44, 16.
- Munang, L.A., Leonard, P.A. & Mok, J.Y. (2002) Lack of agreement on colour description between clinicians examining childhood bruising. *J Clin Forensic Med*, 9(4), 171-174.
- Murphy v The Queen* (1989) 167 CLR 94, 111
- Nambiar, P., Bridges, T.E. & Brown, K.A. (1995) Quantitative forensic evaluation of bite marks with the aid of a shape analysis computer program: Part 2; "SCIP" and bite marks in skin and foodstuffs. *J Forensic Odontostomatol*, 13(2), 26-32.
- National Research Council (1996) *The Evaluation of Forensic DNA evidence* National Academy Press
- National Research Council (2004) FBI Should Revise Approach to Bullet Lead Comparison And Carefully Limit Its Use in the Courtroom. *Media Release*. 10 Feb 2004, available at <http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=10924>
- National Research Council (2009) *Strengthening Forensic Science in the United States : A Path Forward* National Academies Press
- Nell, G. (2006) The Admissibility of Expert Evidence. *Bar News*, Summer 2006/2007, 63-70.
- Neufeld, P.J. (2005) The (near) irrelevance of Daubert to criminal justice and some suggestions for reform. *Am J Public Health*, 95 Suppl 1, S107-13.
- Nichols, R.G. (2007) Defending the scientific foundations of the firearms and tool mark identification discipline: responding to recent challenges. *J Forensic Sci*, 52(3), 586-594.
- Nordby, J.J. (1992) Can we believe what we see, if we see what we believe? Expert disagreement. *J Forensic Sci*, 37(4), 1115-1124.
- NRMA Ltd v Morgan & Ors* (Unreported, Supreme Court of NSW, 1 September 1998)
- O'Connor, S. (1995) The Supreme Court's Philosophy of Science: Will the Real Karl Popper Please Stand Up? *Jurimetrics J*, 35, 263.
- O'Connor, T., Forensic Odontology. <http://www.drtoconnor.com/3210/3210lect02b.htm>, accessed on 31 Oct, 2011.
- Ocean Marine Mutual Insurance Association (Europe) OV v Jetopay Pty Ltd* FCA 1463
- Ogders, S.J. & Richardson, J.T. (1995) Keeping bad science out of the courtroom - changes in American and Australian expert evidence law. *University of NSW Law Journal*, 18(1), 108.



- Ohshima, T. (2000) Forensic wound examination. *Forensic Sci Int*, 113(1-3), 153-164.
- Olson, S.M. (1991) Studying Federal District Courts through Published Cases: A Research Note. *Just. Sys. J.*, 15, 782.
- Osland v The Queen* (1998) HCA 75
- Oxford Centre for Evidence-based Medicine, Levels of Evidence.  
<http://www.cebm.net/index.aspx?o=1025> Participant #levels, accessed on 1 Oct, 2009.
- Pallier, G. et al. (2002) The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, 129(3), 257-299.
- Pankanti, S., Prabhakar, S. & Jain, A.K. (2002) On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1010-1025.
- Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods*. Sage Publications, Thousand Oaks, Calif.
- Pearson, K. (1930) *The Life and Letters of Sir Francis Galton*. University Press, Cambridge.
- People v Kelley* 549 P.2d 1240, 1244 (Cal. 1976)
- People v Ballard* No. 225560 Macomb Cir. Court LC No. 98-001651-FC (Mich. Feb 28, 2003)
- People v Milone* 356 N.E.2d 1350, 1358 (Ill. App. Ct. 1976)
- People v Partee* 2008 WL 616249 (Cal.App. 4 Dist., Mar 07 2008)(Unreported)
- People v Smith*, 443 N.Y.S.2d 551, 556B57 (Cty. Ct. 1981)
- People v Wright* 1998 Mich. LEXIS 2580
- People v Wynne* 1999 Mich. App. LEXIS 2855
- Nordberg, P., Daubert on the Web. [www.daubertontheweb.com](http://www.daubertontheweb.com), accessed on 05 Oct, 2009.
- Peterson, J.L. & Markham, P.N. (1991) Crime Lab Proficiency Testing Results, 1978-1991, 1:Identification and Classification of Physical Evidence. *J Forensic Sci*, 40(6), 994-1008.
- Peterson, J.L. & Murdock, J.E. (1989) Forensic Science Ethics: Developing an Integrated System of Support and Enforcement. *J Forensic Sci*, 34, 753.
- Pfadenhauer, M. (2009) At Eye Level - the expert interview - A talk between expert and quasi-expert. In *Interviewing Experts*, (Eds, Bogner, A., Littig, B. & Menz, W.) Palgrave Macmillan,
- Phillips, V.L., Saks, M.J. & Peterson, J.L. (2001) The application of signal detection theory to decision-making in forensic science. *J Forensic Sci*, 46(2), 294-308.
- Polentz, M.C. (1996) Post Daubert Confusion with Expert Testimony. *Santa Clara L Rev*, 36, 1187.
- Popper, K. (2002) *The Logic of Scientific Discovery*. Routledge, London.
- Porter, C. & Parker, R.W.R. (2001) The Demeanour of Expert Witnesses. *Aust J Forensic Sci*, 33(2), 45-50.
- Powers v State* 2006 Miss. LEXIS 585
- Pretty, I. & Sweet, D. (2010) A paradigm shift in the analysis of bite marks. *Forensic Science International*, 201, 38-44.
- Pretty, I.A. (2003) A web-based survey of odontologist's opinions concerning bite mark analyses. *J Forensic Sci*, 48(5), 1117-1120.
- Pretty, I.A. (2007) Development and validation of a human bite mark severity and significance scale. *J Forensic Sci*, 52(3), 687-691.
- Pretty, I.A. & Sweet, D. (2001a) Digital bite mark overlays--an analysis of effectiveness. *J Forensic Sci*, 46(6), 1385-1391.
- Pretty, I.A. & Sweet, D. (2001b) Adherence of forensic odontologists to the ABFO bite mark guidelines for suspect evidence collection. *J Forensic Sci*, 46(5), 1152-1158.

- Pretty, I.A. & Sweet, D. (2001c) The scientific basis for human bitemark analyses - a critical review. *Sci Justice*, 41(2), 85-92.
- Pronin, E. & Kugler, M.B. (2007) Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *J Exper Soc Psych*, 43(4), 565-578.
- Pyrek, K. (2007) *Forensic science under siege : the challenges of forensic laboratories and the medico-legal death investigation system*. Elsevier Academic Press, Amsterdam ; Boston.
- R v Johnson* (1994) 75 A Crim R 522, 535
- R v Bonython* (1984) 38 SASR 45
- R v Carroll* (1985) 19 A Crim R 410
- R v Carroll* (2000) QSC 308
- R v Carroll* (2001) QCA 394
- R v Carroll* (2001) Trial transcript.
- R v Carroll* (1985) Trial transcript
- R v Gallagher* (2001) NSWSC 462
- R v Gilmore* (1977) 2 NSWLR 935
- R v Karger* (2001) 83 SASR 1
- R v Lewis* (1987) 29 A Crim R 267
- R v Pantoja* (1996) 88 A Crim R 554
- R v Parker* (1912) VLR 152
- R v Tang* (2006) 65 NSWLR 681
- Radmer, T. & Thomas Johnson, L. (2009) The Correlation of Dental Arch Width and Ethnicity. *J Forensic Identification*, 59(3), 268-274.
- Raekallio, J. (1972) Determination of the age of wounds by histochemical and biochemical methods. *Forensic Sci*, 1(1), 3-16.
- Ramirez v State* 819 So.2d 836 (Fla., 2001)
- Rao, V.J. & Souviron, R.R. (1984) Dusting and lifting the bite print: a new technique. *J Forensic Sci*, 29(1), 326-330.
- Rawson, R.D. (1976) Solarisation as an aid to bite mark analysis. *Int J Forensic Dent*, 3(10), 31-33.
- Rawson, R.D., Ommen, R.K., Kinard, G., Johnson, J. & Yfantis, A. (1984) Statistical evidence for the individuality of the human dentition. *J Forensic Sci*, 29(1), 245-253.
- Rawson, R.D., Vale, G.L., Sperber, N.D., Herschaft, E.E. & Yfantis, A. (1986) Reliability of the scoring system of the American Board of Forensic Odontology for human bite marks. *J. Forensic Sci.*, 31, 1235, 1259.
- Raymond, T. (2006) The Future for Forensic Scientists. *Aust J Forensic Sci*, 38(3), 3-21.
- Real, M.L. (2000) Daubert - A Judge's View. *The Practical Litigator*, 11(3), 7.
- Reed v State* 391 A.2d 364, 371-72 (MD, 1978)
- Reichs, K.J. (1993) Quantified comparison of frontal sinus patterns by means of computed tomography. *Forensic Sci Int*, 61(2-3), 141-168.
- Reporter's Notes. (1999) *Proposed Amendments to the Uniform Rules of Evidence*. National Conference of Commissioners of Uniform State Laws,
- Restatement of the law third, torts, liability for physical and emotional harm*. American Law Institute Publishers, St. Paul, MN 2010.

- Rice, P.R. (2000) Truth in Test Tubes: The Standard for Screening Scientific Evidence Is Still Muddled Years After Daubert. *Legal Times*, 2000, pp
- Rice, P.R. & Delker, N.-E.W. (2000) Federal Rules of Evidence Advisory Committee : A Short History of Too Little Consequence. *F.R.D.*, 191, 678.
- Richards, D. (2007) Elite interviewing: approaches and pitfalls. *Politics*, 16(3), 199-204.
- Risinger, D.M. (2000) Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock? *Alb. L. Rev.*, 64, 99.
- Risinger, D.M. (2007a) Appendix: Cases Involving the Reliability of Handwriting Identification Expertise Since the Decision in *Daubert*. *Tulsa L Rev*, 43, 477.
- Risinger, D.M. (2007b) Goodbye To All That, or, A Fool's Errand, by one of the Fools: How I stopped worrying about court responses to handwriting identification (and "forensic science" in general) and learned to love misinterpretations of *Kumho Tire v Carmichael*. *Tulsa L Rev*, 43, 447.
- Risinger, D.M., Saks, M.J., Thompson, W.C. & Rosenthal, R. (2002) The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion. *Cal. L. Rev.*, 90(1), 1-55.
- Riviello, R.J. (2010) *Manual of forensic emergency medicine : a guide for clinicians*. Jones and Bartlett Publishers, Sudbury, Mass.
- Roach, J., No Two Snowflakes the Same" Likely True, Research Reveals". <http://news.nationalgeographic.com/news/2007/02/070213-snowflake.html>., accessed on 14 January, 2010.
- Robrecht, L. (1995) Grounded Theory: Evolving Methods. *Qualitative Health Research*, 5(2), 169-177.
- Roisman, A.Z. (2009) Taming the *Daubert* Tiger. *The Practical Litigator*, 20(3), 49-64.
- Rothwell, B.R. (1995) Bite marks in forensic dentistry: a review of legal, scientific issues. *J Am Dent Assoc*, 126(2), 223-232.
- Roxburgh, T. (1933) Galton's work on the Evidential Value of Finger Prints. *Sankhya: Indian Journal of Statistics*, 1, 50-62.
- Rudin, N. & Inman, K. (2006) The Shifty Paradigm, Part II. *CAC News*, First Quarter, 16.
- Rudin, N. & Inman, K. (2008) Who Speaks for Forensic Science? The Conviction and Exoneration of a Straw Man. *CAC News*, Fourth Quarter, 10.
- Ruiz-troche v Pepsi Cola Bottling Co.* 161 F.3d 77, 85 (1st Cir. 1998)
- Russell, B. (2004) *The problems of philosophy (1912)*. Barnes & Noble Books, New York.
- Russell, S. (2006) Down in Flames - 'Rock-solid' evidence for arson turns out to be anything but. *New Scientist*, 14 Nov 2006, pp 42-45.
- Saks, M.J. Interview with author. (02 Nov 2009a)
- Saks, M.J. (1995) The Phantom of the Courthouse. *Jurimetrics*, 35, 233.
- Saks, M.J. (2000) The Aftermath of *Daubert*: An Evolving Jurisprudence of Expert Evidence. *Jurimetrics J*, 40, 229-241.
- Saks, M.J. (2002) The Legal and Scientific Evaluation of Forensic Science. *Seton Hall L. Rev.*, 33, 1167.
- Saks, M.J. (2009b) *Forensic Identification Science: A Summary of the Theory, the Science and the Legal History*. [Updated version of *Merlin and Solomon: Lessons from the Law's formative Encounters with Forensic Identification Science*, 49 *Hastings L J* 1069 (1998)],
- Saks, M.J. & Faigman, D.L. (2005) Expert Evidence After *Daubert*. *Annu Rev Law Soc Sci*, 1, 105-130.

- Saks, M.J. & Faigman, D.L. (2008) Failed Forensics: How forensic science lost its way and how it might yet find it. *Ann Rev Law Soc Sci*, 4, 149-171.
- Saks, M.J., Faigman, D.L., Kaye, D.H. & Sanders, J. (2009) *Annotated Reference Manual on Scientific Evidence (2nd Ed.)*. Thomson Reuters,
- Saks, M.J. & Koehler, J.J. (1991) What DNA Fingerprinting Can Teach the Law About Forensic Science. *Cardoza L Rev*, 13, 361-372.
- Saks, M.J. & Koehler, J.J. (2008) The Individualization Fallacy in Forensic Science Evidence. *Vand. L. Rev.*, 61, 199-220.
- Saks, M.J. & Koehler, J.J. (2005) The Coming Paradigm Shift in Forensic Identification Science. *Science*, 309 (5736), 892.
- Saks, M.J. & Risinger, D.M. (2003) A House With No Foundation. *Issues Sci Tech*, Fall, 35.
- Saks, M.J., Risinger, D.M., Rosenthal, R. & Thompson, W.C. (2003) Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States. *Science & Justice*, 43(2), 77-90.
- Sales, B.D. & Shuman, D. (2005) *Experts in Court: Reconciling Law, Science, and Professional Knowledge*. American Psychological Association,
- Sato, Y. & Ohshima, T. (2000) The expression of mRNA of proinflammatory cytokines during skin wound healing in mice: a preliminary study for forensic wound age estimation (II). *Int J Legal Med*, 113(3), 140-145.
- Schmid, R.E. (1988) Identical Snowflakes Cause Flurry. *The Boston Globe*, 15 June 1988, pp
- Schossor, R.H. & Kendrick, J.P. (1987) Dermatologic photography. *Dermatol Clin*, 5(2), 445-461.
- Science Technology and Law Panel Policy and Global Affairs (2002) The Age of Expert Testimony: Science in the Courtroom. 2002
- Scientific Working Group on Friction Ridge Analysis Study and Technology (SWGFAST) (2011) Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint). *Version 1.0*,
- Scientific Working Group on Friction Ridge Analysis Study and Technology (1997) Proposed SWGFAST guidelines. *J Forensic Identification*, 47, 425.
- Severinsson, E.I. & Borgenhammar, E.V. (2003) Expert views on clinical supervision: a study based on interviews. *Journal of nursing management*, 5(3), 175-183.
- Sexton v State* 93 S.W.3d 96 (Tex. Crim. App., 2002)
- Shannon, C.R. (1984) Royal Commission of Inquiry in Respect to the Case of Edward Charles Splatt (S. Aust) Government Printer
- Sheasby, D.R. & MacDonald, D.G. (2001) A forensic classification of distortion in human bite marks. *Forensic science international*, 122(1), 75-78.
- Sheehan v Daily Racing Form Inc* 104 F.3d 940 (7th Cir 1997)
- Shkrum, M.J. & Ramsay, D.A. (2007) *Forensic pathology of trauma : common problems for the pathologist*. Humana Press, Totowa, N.J.
- Shuman, D.W., Whitaker, E. & Champagne, A. (1993) An Empirical Examination of the Use of Expert Witnesses in the Courts - Part II: A Three City Study. *Jurimetrics J*, 34, 193.
- Shynkaruk, J.M. & Thompson, V.A. (2006) Confidence and accuracy in deductive reasoning. *Memory and Cognition*, 34(4), 619-632.
- Sim, J. & Wright, C.C. (2005) The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257-268.
- Slovenko, R. (1998) The Daubert Sequelae. *Int J Evidence and Proof*, 2, 190.
- Smith v State* Ark. LEXIS 527 (Unreported, 1997)

- Snowden, S.O. (2001) That's My Holding and I'm Not Sticking To It--Court Rules That Deprive Unpublished Opinions of Precedential Authority Distort the Common Law Note. *Wash. U. L. Q.*, 79, 1253.
- Snyder, C.R. & Fromkin, H.L. (1980) *Uniqueness, the human pursuit of difference*.
- Sognnaes, R.F., Rawson, R.D., Gratt, B.M. & Nguyen, N.B.T. (1982) Computer comparison of bitemark patterns in identical twins. *J.A.D.A.*, 105, 449-451.
- Sombat, J.M. (2001) Latent Justice: Daubert's Impact on the Evaluation of Fingerprint Identification Testimony Note. *Fordham L. Rev.*, 70, 2819.
- Sparrow, M.K. (1994) Measuring AFIS matcher accuracy. *The Police Chief*, April, 147.
- Spencer v Commonwealth* (1989) 384 SE 2d.785
- Sperling, J.H.D. (2000) Expert Evidence: The Problem of Bias and Other Things. *The Judicial Review*, 4, 429.
- Srihari, S.N. (2009) Quantitative Assessment of the Individuality of Friction Ridge Patterns: Final Report submitted to the US Department of Justice.
- Stanford Encyclopedia of Philosophy. <http://plato.stanford.edu/entries/popper/Participant#SciKnoHisPre>, accessed on 22 July, 2009.
- Starr, J.M. & Hall, R. (2010) Predictors and correlates of edentulism in healthy older people. *Curr Opin Clin Nutr Metab Care*, 13(1), 19-23.
- Starr, V.H. & McCormick, M. (2000) *Jury Selection*. Aspen Publishers,
- Starrs, J.E. (1994) Bite Marks in the Courts. A review of 1994's Appellate Court Decisions. *Scientific Sleuthing Review*, 184 (4), 4.
- Starrs, J.E. (1999) Judicial Control over Scientific Supermen. Fingerprint experts and others who exceed the bounds. *Crim Law Bull*, 35(3), 234-276.
- State of Maryland v Bryan Rose* No. K06-0545 (Md., Baltimore Co. Cir. Oct. 19, 2007).
- State of Maryland v Bryan Rose* No. K06-0545 (Md., Baltimore Co. Cir. Feb 21 2008). Memorandum Decision Re: Reconsideration
- State of New Hampshire v Richard Langill* No. 05-5-1129 (Sup. Ct. Rockingham Jan. 19, 2007)
- State Rail Authority of New South Wales v Earthline Constructions Pty Ltd (In Liq)* (1999) 160 ALR 588
- State v Brewer* 2005 WL 1023238 (Unreported)
- State v Fortin* 2007 N.J. LEXIS 333 (Unreported)
- State v Lopez* 1996 Neb. LEXIS 51
- State v Swinton* 2004 Conn. LEXIS 190 (Unreported)
- Stevenson, J. (2005) *The complete idiot's guide to philosophy*. Alpha, New York.
- Stilwell, R.T. (2000) Kumho Tire: The Battle of the Experts Continues. *Rev. Litig.*, 19, 193.
- Stimson, P.G. & Mertz, C.A. (1997) *Forensic dentistry*. CRC Press, Boca Raton.
- Stone, I.C. & Wilmovsky, A. (1984) Evidentiary Basis for Fingernail Striation Association. *J Pol Sci Admin*, 12, 291-296.
- Stone, R.S. (2003) How unique are impressed toolmarks? *AFTE J*, 35(4), 376-383.
- Stoney, D.A. (1988) Distribution of Epidermal Ridge Minutiae. *Am J Phys Anthropol*, 77, 367-376.
- Stoney, D.A. (1991) What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society*, 31(2), 197-199.
- Stoney, D.A. & Thornton, J.I. (1986) A Critical Analysis of Qualitative Fingerprint Individuality Models. *J Forensic Sci*, 31(4), 1187-1216.

- Stove, D.C. (1982) *Popper and After: Four Modern Irrationalists*. Pergamon Press, Sydney.
- Street, A.W., Expert Evidence (*Notes*).  
<http://www.sevenwentworth.com.au/publications/Expert%20Evidence%20-%20Street%20SC.pdf>,  
 accessed on July 06, 2009.
- Sweet, D. & Bowers, C.M. (1998) Accuracy of bite mark overlays: a comparison of five common methods to produce exemplars from a suspect's dentition. *J Forensic Sci*, 43(2), 362-367.
- Sweet, D.J. & Bastien, R.B. (1991) Use of an acrylonitrile-butadiene-styrene (ABS) plastic ring as a matrix in the recovery of bite mark evidence. *J Forensic Sci*, 36(5), 1565-1571.
- Sydneywide Distributors Pty Ltd v Red Bull Australia Pty Ltd* (2002) 55 IPR 354
- Tamarelli, A.W. (1994) *Daubert v Merrell Dow Pharmaceuticals*: Pushing the Limits of Scientific Reliability - The Questionable Wisdom of Abandoning the Peer Review Standard for Admitting Expert Testimony. *Vand L Rev*, 47,
- Tante v Roth* SCNSW, 28 November 1996 (Unreported)
- Taslitz, A.E. (1995) Daubert's Guide to the Federal Rules of Evidence - A Not-So-Plain Meaning Jurisprudence. *Harv J on Legis*, 32, 3.
- The Evidence Project (1998) Proposed Revisions to the Federal Rules of Evidence. October 22 1998,
- The Free Dictionary, Legal Dictionary. <http://legal-dictionary.thefreedictionary.com/>, accessed on 24 June, 2011.
- The Innocence Project. [www.innocenceproject.org](http://www.innocenceproject.org), accessed on 02 Feb, 2009.
- The Justice Project (2008) Improving the Practice and Use of Forensic Science: A Policy Review
- Thomas Johnson, L., Blinka, D.D., VanScotter Asbach, P. & Radner, T.W. (2008) Quantification of the Individual Characteristics of the Human Dentition: Methodology. *J Forensic Identification*, 58(4), 409-509.
- Thomas Johnson, L., Radner, T.W., Wirtz, T., Pajewski, N.M., Cadle, D.E., Brozek, J Blinka, DD. (2009) Quantification of the Individual Characteristics of the Human Dentition. *J Forensic Identification*, 59(6), 609-626.
- Thompson, W.C. (2009) Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk*, 8(3), 257-276.
- Thomson, K.S. (1991) *Living fossil : the story of the coelacanth*. W.W. Norton, New York.
- Thornburgh, D. (1997) Junk Science - The Lawyer's Ethical Responsibilities Remark. *Fordham Urb. L.J.*, 25, 449.
- Thornton, J.I. (1986) The Snowflake Paradigm. *J Forensic Sci*, 31(2), 399-401.
- Thornton, J.I., Peterson, J.L. & Legget, A.S. (2007) The General Assumptions and Rationale of Forensic Identification. In *Modern Scientific Evidence*, (Eds, Faigman, D.L., Saks, M.J., Sanders, J. & Cheng, E.K.) Thomson West,
- Thornton, R.N. & Jolly, R.D. (1986) The objective interpretation of histopathological data: an application to the ageing of ovine bruises. *Forensic Sci Int*, 31(4), 225-239.
- Ungvarsky, E.J. (2006) Remarks on the Use and Misuse of Forensic Science to Lead to False Convictions. *New Eng. L. Rev.*, 41, 609.
- United States of America v Taylor*. 464 F.2d 240 (2nd Cir. 1972)
- United States v Addison* 498 F.2d 741 (D.C. Cir 1974)
- United States v Brown* 557 F.2d 1194 (2d Cir. 1978)
- United States v Crisp* 324 F.3d 261 (4th Cir. 2003)
- United States v Dixon* 413 F.3d 520 (5th Cir. 2005)
- United States v Downing* 753 F.2d.1224 1239-1241 (3d Cir. 1985)

- United States v Starzecpyzel* 880 F.Supp. 1027 (S.D.N.Y. 1995)
- United States v Young* 316 F.3d 649 (7th Cir. 2002)
- United States v Byron Mitchell* Criminal Action No. 96-407, US District Court for the Eastern District of Pennsylvania. July 1999 (Unreported)
- United States v Cooper* 91 F.Supp.2d 79 (D.D.C., 2000)
- United States v Cromer* 2006 WL 1430210 (W.D.Mich., 2006)(Unreported)
- United States v Diaz* 2007 US Dist. LEXIS 13152
- United States v Fisher* 2002 US Dist. LEXIS 22385
- United States v Fujii* 152 F.Supp. 2d 939 (ND Ill 2000)
- United States v Green* 405 F.Supp.2d 104 (D.Mass. 2005)
- United States v Havvard* 117 F. Supp. 2d 848, 849 (S.D. Ind. 2000), aff 'd, 260 F.3d 597 (7th Cir. 2001)
- United States v Hines* 55 F.Supp.2d 62 (D.Mass. 1999)
- United States v Jones* 107 F.3d 1147 (6th Cir. 1997)
- United States v Llera Plaza* 179 F.Supp. 492 (E.D. Pa. 2002)
- United States v Llera Plaza* 188 F. Supp. 2d 549 (E.D. Pa. 2002)
- United States v Monteiro* 407 F.Supp.2d 351 (D.Mass., 2006)
- United States v Nichols* 169 F.3d 1255 (C.A.10 (Colo.),1999)
- United States v O'Driscoll* Not Reported. 2003 WL 1402040 (M.D.Pa.)
- United States v Parks* No. CR-91-358-JSL (C.D. Cal. Dec. 10, 1991)
- United States v Paul* 175 F.3d 906 (11th Cir. 1999)
- United States v Prime* 220 F.Supp.2d 1203 (W.D.Wash. 2002)
- United States v Reaux* No. CRIM A. 01-071, 2001 WL 883221 (E.D. La. July 31, 2001) (Unreported)
- United States v Rutherford* 104 F.Supp.2d 1190 (D.Neb. 2000)
- United States v Saelee* 162 F.Supp.2d 1097 (D Alaska 2000)
- United States v Santiago* 199 F.Supp.2d 101 (2002)
- United States v Santillan* 1999 WL 1201765 (N.D.Cal. 1999)
- United States v Sullivan* 246 F.Supp.2d 702, 703 (E.D.Ky. 2003)
- United States v Van Wyk* 83 F.Supp.2d 515 (D.N.J. 2000)
- United States v Williams* 506 F.3d 151 (2007)
- Upadhyaya, C. & Humagain, M. (2009) The pattern of tooth loss due to dental caries and periodontal disease among patients attending dental department (OPD), Dhulikhel Hospital, Kathmandu University Teaching Hospital (KUTH), Nepal. *Kathmandu Univ Med J (KUMJ)*, 7(25), 59-62.
- Ursano, R.J., Fullerton, C.S., Vance, K. & Kao, T.-C. (1999) Posttraumatic Stress Disorder and Identification in Disaster Workers. *Am J Psychiatry*, 156, 353-359.
- US Department of Justice (2006) A Review of the FBI's Handling of the Brandon Mayfield Case
- Velevski v The Queen* (2002) HCA 4
- Vickers, A.L. (2005) Daubert, Critique and Interpretation: What Empirical Studies Tell Us about the Application of Daubert. *USF. L. Rev.*, 40, 109.
- Vogel, B. (1995) *Wenn der Eisberg zu schmelzen beginnt. Einige Reflexionen über den Stellenwert und die Probleme des Experteninterviews in der Praxis der empirischen Sozialforschung* (If the iceberg begins to melt. Some reflections about the significance and problems of interviews with

- experts in the practice of social research). In *Experteninterviews in der Arbeitsmarktforschung*, (Eds, Brinkmann, C., Deeke, A. & Volkel, B.) Bundesanstalt fur Arbeit, Nuremburg,
- Walker, D. & Myrick, F. (2006) Grounded Theory: An Exploration of Process and Procedure. *Qualitative Health Research*, 16, 547-559.
- Waters, N.L. & Hodge, J.P. (2005) *The Effects of the Daubert Trilogy in Delaware Superior Court*. Ph.D. thesis, National Center for State Courts.
- Wayman, J.L., When Bad Science Leads to Good Law: The disturbing irony of the *Daubert* hearing in the case of *US v Byron C. Mitchell*. [www.engr.sjsu.edu/biometrics/publications\\_daubert.html](http://www.engr.sjsu.edu/biometrics/publications_daubert.html), accessed on 04 March, 2010.
- Weinstein, J.B. & Berger, M.A. (1987) *Weinstein's Evidence*.
- Weir, B.S. (1999) Are DNA Profiles Unique? *Proceedings of the Ninth International Symposium on Human Identification*, 1999
- Weiss v The Queen* (2005) 224 CLR 300
- Welsh Child Protection Systematic Review Group, Bites Review. Available at [http://www.core-info.cf.ac.uk/oral\\_inj/bites/bites\\_key.htm](http://www.core-info.cf.ac.uk/oral_inj/bites/bites_key.htm), accessed on 24 Oct 2010.
- West, M.H., Barsley, R.E., Hall, J.E., Hayne, S. & Cimrmancic, M. (1992) The detection and documentation of trace wound patterns by use of an alternative light source. *J Forensic Sci*, 37(6), 1480-1488.
- West, M.H. & Frair, J. (1989) The use of videotape to demonstrate the dynamics of bite marks. *J Forensic Sci*, 34(1), 88-95.
- Wheate, R. (2006) Australian Juries and Scientific Evidence. *Aust J Forensic Sci*, 38, 75-84.
- Wheate, R. (2007) *Jury Comprehension and Use of Forensic Science*. Ph.D. thesis, Univeristy of New South Wales, Australian Defence Force Academy.
- Whitman, G. & Koppl, R. (2010) Rational bias in forensic science. *Law, Probability and Risk*, 9(1), 69-90.
- Whittaker, D.K. (1975) Some laboratory studies on the accuracy of bite mark comparison. *Int Dent J*, 25(3), 166-171.
- Whittaker, D.K., Brickley, M.R. & Evans, L. (1998) A comparison of the ability of experts and non-experts to differentiate between adult and child human bite marks using receiver operating characteristic (ROC) analysis. *Forensic science international*, 92(1), 11-20.
- Wickstrom, G. & Bendix, T. (2000) The Hawthorne effect: what did the original Hawthorne studies actually show? *Scand J Work, Env & Health*, 26(4), 363-367.
- Wilkins, J.S. (1998) What's in a Meme? Reflections from the perspective of the history and philosophy of evolutionary biology. *Journal of Memetics-Evolutionary Models of Information Transmission*, 2(1),
- Williams, D.A. & Hahn, S. (2002) Appropriateness of using Visual Line-Up Guidelines in the Preparation of Bite Mark Analysis. *American Academy of Forensic Sciences*, 2002
- Williamson, D. & Harding, K. (2004) Wound Healing. *Medicine*, 32(12), 4-7.
- Wilson, P. (1994) Lessons from the Antipodes: Successes and failures of forensic science. *Forensic Sci Int*, 67, 79.
- Witzel, A. (2000) The problem-centered interview. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 1(1), 2000
- Wolf v Ramsey* 2003 US Dist. LEXIS 10249
- Wood, J.J.R.T. (2003) Forensic Science from the Judicial Perspective. *Australian Bar Review*, 23, 137.



## References

Woolf, T.R.H.t.L., Access to Justice: Final Report to the Lord Chancellor on the civil justice system in England and Wales. Available at: <http://www.dca.gov.uk/civil/final/contents.htm>, accessed on 07 July, 2009.

Yin, R.K. (1994) Case study research: Design and methods (2nd ed.). Sage Publications, CA.

APPENDIX

**2012**

---

**Guidelines**  
for the conduct of  
**Bitemark Analysis**  
in  
**Australia**

---

Version 1.0

## **Guidelines For Bitemark Analysis**

<b>1. Evidence Collection from the Victim .....</b>	<b>1</b>
<b>2. Initial Assessment of Bitemark Evidence.....</b>	<b>8</b>
<b>3. Evidence Collection from a Suspect .....</b>	<b>11</b>
<b>4. Comparative Analysis .....</b>	<b>15</b>
<b>5. Terminology .....</b>	<b>18</b>
<b>6. Reporting .....</b>	<b>22</b>
<b>7. Contextual bias, cognitive bias and observer effects.....</b>	<b>25</b>
<b>8. Presentation of bitemark evidence in court .....</b>	<b>28</b>
<b>9. Education, training and continuing professional development.....</b>	<b>31</b>
<b>Sample Bitemark Report.....</b>	<b>33</b>

## Guideline 1

---

# Evidence Collection from the Victim

---

### Preamble

Collection of bite mark evidence constitutes the first and most crucial phase of bite mark analysis. The nature of bite mark evidence means that attempts beyond the initial encounter with the latent mark are rarely presented, due to the transitory nature of the wound pattern and the rapid degradation of evidence quality that generally follows. Collection of bite mark evidence should endeavour to obtain maximal information so as to enable meaningful analysis and comparison of the wound pattern with a potential source at a time when access to the original material evidence is no longer possible. Physical bite mark evidence is of varying quality. Every effort must be made to capture the true nature of the evidence at the time of collection by recording as much physical detail about the injury as possible at the time of examination of the victim.

### Scope

This guideline details recommendations for the collection, recording and documentation of physical bite mark evidence from a living or deceased victim for subsequent use in either qualitative or comparative analysis. Detailed protocols for photography in the forensic setting are beyond the scope of these guidelines, and practitioners should refer to appropriate texts and documents for more detailed advice in this regard.

**1.1 Supervision.** In light of the sporadic nature of bite mark casework, it is likely that police photographers, crime scene technicians, medical personnel and other professional groups who may encounter such evidence may have had little personal experience in the collection of bite mark evidence.

- 1.1.1 Odontologists should make themselves available, where practical, to assist in the collection and recording of bite mark evidence in order to ensure that appropriate steps are taken to preserve accurate images of the latent marks.
- 1.1.2 Odontologists should direct, where appropriate, appropriately trained photographers in the recording of bite mark evidence. Use of a forensic photographer is desirable in the majority of circumstances, as they are likely to have high-quality photography equipment on hand, in addition to knowledge of how to best use it to achieve the desired results. Such records will also form part of the official forensic documentation for any given case.

- 1.1.3 Some odontologists may prefer to take 'unofficial' or 'personal' photographs using their own digital camera. While such photographs may provide a useful record for the odontologist's personal case log, their use as images for subsequent forensic analysis is highly questionable given the potentially basic nature of the photographic equipment and the lack of official chain-of-evidence custody of the resultant images. The use of these images for forensic analysis should only proceed where it was considered not feasible to have photographs of the latent mark taken by a trained forensic photographer. Such situations are to be avoided where possible.

**1.2 DNA swab.** There is little doubt that DNA analysis provides a far more robust means of identification than analysis and comparison of latent marks to source objects. During the process of biting and also during kissing and sucking, saliva is deposited on the skin's surface. The literature has demonstrated that this trace evidence is present in sufficient quantity and quality to enable PCR-based typing of the DNA that is present in saliva from white blood cells and sloughed epithelial cells. DNA from residual saliva should be considered as the primary means of identifying perpetrators when bitemark evidence is encountered, however, several factors influence whether or not DNA evidence is likely to be encountered in these situations, such as the material state of the body at the time of discovery, whether or not the victim has washed or been washed since the mark was inflicted, and so forth.

- 1.2.1 DNA evidence collection should proceed as soon as practical after the discovery of bitemark evidence, but does not then negate the need for careful written and photographic documentation of the latent mark as reliance on DNA evidence from saliva may prove fruitless due to the many factors that influence the preservation of biological trace evidence.

- 1.2.2 Collection of DNA evidence from bitemark sites has been well documented. The most consistent, robust method of collecting trace DNA from saliva is that described by Sweet et al using a double swab method.

- 1.2.2.1 This method involves first using a cotton swab that has been moistened with distilled water to collect DNA by swabbing the area from the centre of the latent mark outwards, using a circular motion. The swab should be 'rolled' over the surface of the wound, rather than being used in a scrubbing manner. Extension slightly beyond the periphery of the marks is desirable in order to collect salivary DNA that may have been left by the lips. A second, dry swab is then employed in the same manner as the first in order to collect the moisture left by the first swab.

- 1.2.2.2 Both swabs should be air dried (>30 mins) at room temperature prior to submission for analysis. They should not be stored in plastic bags or containers, but should be kept cool and dry in order to reduce bacterial growth that would otherwise further contaminate the sample. They may be refrigerated or frozen during storage.

- 1.2.3 A reference sample from the victim may prove useful in order to interpret a mixed sample and can be taken at this time using a standard buccal smear. This is

especially useful in cases where the injury has drawn the victims blood, thus contaminating the area where the perpetrator's saliva is most likely to be found.

**1.3 Preparation of the bitemark area.** It is sometimes necessary to prepare the wound in order to allow accurate photographs to be taken.

- 1.3.1 Washing of the wound should only be carried out after the swab for DNA has been taken, and after the wound has been properly documented photographically in its unwashed state. Despite the fact that photographs of the unwashed wound are unlikely to prove useful from a comparison perspective, they form an important aspect of chain-of-evidence and may be useful for other members of the forensic team.
- 1.3.2 Areas of the body that may have excessive or dark-coloured hair, such as the head, arm, leg or genital regions may need to be shaved in order to accurately portray the bitemark. Extreme care should be taken not to nick or damage the skin further, and the importance of pre-procedural photographs cannot be overstated.
- 1.3.3 Examination of suspected marks on the deceased should occur when the cadaver is as close to room temperature as possible. This may not be practical when the body is severely decomposed, but should be considered in other cases.
- 1.3.4 Excision of whole areas of skin that contain a suspected bitemark injury have been reported in the literature, in order to preserve them for later documentation and analysis. Excision of tissue in this regard requires an injunction from the coroner in most jurisdictions in Australia. Despite several techniques being described, the potential for distortion in this process outweighs any advantages it may have for forensic comparison. While research continues in this area, and practitioners in the US are particularly noted for this technique, it remains unsupported by any advantages over the use of good digital photographic records for comparison, and cannot be reasonably endorsed as a method of bitemark evidence collection in Australia.

**1.4 Photographic records.** The gold standard for comparison of a suspected bitemark to a dentition is carried out using photographs of the injury. Direct comparison of a modelled dentition to the wound surface is perhaps unnecessarily traumatic for a living victim, and even if carried out on a deceased victim, is unlikely to produce any useful, reproducible evidence that can then be tendered in court.

- 1.4.1 As per guideline 1.1.2, forensic photographers should be employed where practical, under the supervision of the attending odontologist, to photographically document the injury. A single photograph is unlikely to yield useful information; proper recording of the suspect mark should involve several (at least three) photographs, although it is recognised that ultimately only one or two of these may be used for comparative analysis with a suspected dentition.
- 1.4.2 The minimum recommended photographic equipment that will provide useful images suitable for forensic comparison consists of:

- 1.4.2.1 A SLR digital camera with a 6 megapixel resolution should be considered the minimum technical requirement. Ideally, a 10 megapixel resolution capability is desirable, as this is a close approximation to standard 35mm film.
  - 1.4.2.1.1 Digital images should be stored as lossless compression files, such as TIFF. While JPEG is commonly used for standard photography, it is not suitable for storing forensic images as it is a compression format and results in the continued loss of digital information each time the image is saved.
  - 1.4.2.1.2 If it is impossible to shoot TIFF images at the scene then initial JPEG files should be converted to TIFF files and these used as the primary images.
- 1.4.2.2 A camera with a macro lens will enable close-up shots to be taken with maximal resolution.
- 1.4.2.3 A ring flash, or any multiple flash angle setup that illuminates the wound from multiple angles is superior to a single external flash that illuminates the wound from only one angle. A fixed internal flash is unsuitable for use in forensic photography. Where only a single external flash is available, oblique application of the flash will assist in avoiding 'flash burnout'.
- 1.4.2.4 Film cameras are now rarely encountered, and should be considered inferior to digital cameras given the subsequent digitisation of the photograph that will be necessary in any case, in order to carry out a forensic comparison of the wound. Where it is only practical to employ a film camera, 35mm film should be considered the minimum standard.
- 1.4.3 Orientation photographs that clearly show on what part of the body the wound is located should be taken. These should be taken from reasonable distance and are not meant to demonstrate the finer details of the injury, but to provide contextual information (such as the relative position of the injury) that may prove useful in later analysis.
- 1.4.4 Close up photographs should be taken that demonstrate the finer details of the injury. As a guide, the minimum wound-to-lens distance for 'close-up' photographs should be that which allows individual tooth or bruise marks (when present), to all appear in focus, when considering a single arch. The minimum 'useful' unit of the wound is thus considered to be a single archform, either upper or lower, and thus the entire arch should be in focus in close-range shots. Due to the diffuse nature of cutaneous inflammatory responses, there is little detail likely to be recorded in a cutaneous wound pattern that will correspond with the surface features of individual teeth, thus close-up photographs that demonstrate only one or two discrete tooth marks, with the remainder of the arch out-of focus, are unlikely to be useful. This may not necessarily be true for bite marks in or on inanimate objects, and thus the

odontologist should use his or her discretion when ordering more extreme-range close-ups in these cases.

- 1.4.5 Wounds near flexor or extensor surfaces should be photographed with the limb in various positions, at least three shots at both minimum and maximum extension, in addition to one of mid-flexion. Such photographs will enable an assessment of the likely distortion of the subsequent mark to be made. If the victim is co-operative, a photograph of the limb in the position that he or she states it was when the wound was inflicted may prove useful.
- 1.4.6 Photographs should be taken both with and without the presence of a scale. The inclusion of photographs taken from the same orientation but *without* the scale is important to demonstrate that no significant features are being accidentally hidden by the scale instrument.
- 1.4.7 An ABFO No. 2 scale should be considered the standard scaling tool in this regard. The ABFO No. 2 scale should be placed in the same horizontal *and* vertical plane as the central feature of the patterned portion of the injury. Where no ABFO scale is available, a standard linear forensic scale should be used. Plastic or wooden rulers should be avoided, unless there is no other reasonable choice. Stick-on scales, that follow the curve of the limb, should not be used as they do not allow accurate scaling correction in the digital image, although again, in the absence of anything else, they are better than having no scale at all.
- 1.4.8 Every effort should be made to ensure that photographs of the injury are taken with the long axis of the lens perpendicular to the area of interest in order to minimise distortion.
- 1.4.9 When there are two distinct upper and lower arch marks, it may be necessary to re-locate the scale relative to each mark, being careful to orientate it correctly in accordance with the area of the wound being photographed, and take separate photographs in order to ensure accurate recording. This is particularly true where the suspected bite mark is on a severely curved surface, such as an arm, where the arch marks are likely to be on significantly different planes. This process is far more accurate than placing the scale in the same plane as the centre of the bite mark, which actually contains no patterned features useful for analysis, and which results in having the scale in neither plane that the most useful forensic features (i.e. the marks that have been made by teeth) are located.
- 1.4.10 When the odontologist deems it necessary to take separate images of the upper and lower portions of the injury, the long axis of the lens must likewise be perpendicular to each feature of interest in each photograph. This means that each of the photographs will be taken from a slightly different angle in order to preserve this orientation for each portion of the injury.

**1.5 Impressions of the bite mark area.** Generally speaking, the relative rarity of three-dimensional information associated with cutaneous bite injuries makes the routine recording of the wound via an impression unnecessary. However, if on inspection the



odontologist feels there is three-dimensional information associated with the injury, impressions should be taken.

- 1.5.1 A low-viscosity polyvinylsiloxane (PVS) impression medium should be used in order to obtain maximum accuracy and dimensional stability. Alginate should *not* be considered a suitable material for impressions of suspected bite mark injuries.
- 1.5.2 The initial layer of low-viscosity (light-body) material should be supported by adequate bulk of high-viscosity material in order to protect the impression from tearing and distortion. Heavy-body PVS of the same brand as the low-viscosity material is adequate in this regard, and is preferred as it binds chemically to the low-viscosity material, reducing the potential for dimensional inaccuracies through slippage and shrinkage. Care should be taken not to distort the area of the bite mark if using putty-consistency material. Other methods include supporting the low-viscosity material with self-cure acrylic resin or dental stone, however, these have the disadvantage of being exothermic in their set, and on completion they are easily separated from the recording medium. These factors introduce potential errors associated with the dimensional accuracy of the resulting impression.
- 1.5.3 Once the impression is fully set, it should be poured in high-strength Type-IV dental stone. Plaster of Paris is not of sufficient strength or resistance in this regard, given the amount of handling that the resulting models may be subjected to.

**1.6 Documentation.** Documentation of the recording procedure is an essential step in order to maintain an accurate custodial record of the incident. As a minimum, the following is recommended:

- 1.6.1 The details of the examining odontologist and any other persons present or acting as a witness to the examination, including the time and date of the examination.
- 1.6.2 A detailed description of the injury, including notes on its size, orientation, and location on the body. A description of the type of tissue that the mark is on will also allow an assessment of the likelihood of distortion in the mark, such as the contour of the area, the presence of substantial muscle, adipose or elastic tissue beneath the mark, and whether or not the mark is close to any flexor / extensor surfaces or joints.
- 1.6.3 The number and type of photographs taken.
- 1.6.4 A sketch diagram of the site of the injury, with key measurements such as inter-canine width labelled where possible. Such first-hand information will help verify the accuracy of the subsequent photographs during the analysis phase.
- 1.6.5 A brief case history may prove useful, however, the potential for this information to bias the analysing odontologist needs to be weighed against its potential to aid in answering the forensic question. If the question is one of corroboration of the injury as a bite mark, or of verification of an assault victim's version of the events leading to such an injury, then a case history may provide useful details for the analysing odontologist to compare. If the question is one of identity, the less information

regarding the nature of the attack the better, in order to avoid potentially biasing information that may influence the analysing odontologist's opinion.

## Guideline 2

---

# Initial Assessment of Bitemark evidence

---

### Preamble

Odontologists may be called upon to provide an assessment of a cutaneous injury upon suspicion that it may have been caused by teeth. Assessment of this evidence involves consideration of several factors that will ultimately lead to an opinion regarding the qualitative nature of the injury. Initial assessment of a bitemark injury is best performed during examination of the injury itself, as opposed to assessing photographs, however, it is recognised that this opportunity rarely presents itself. For the purposes of qualitative assessment, assessment of the injury via good-quality photographs should be considered only slightly less advantageous than viewing the injury itself.

### Scope

This guideline outlines the factors which an odontologist should consider when performing an initial assessment of an injury suspected to be a bitemark.

**2.1 Is the injury a bitemark?** The primary question in the initial assessment of cutaneous injuries is that of whether the injury represents a mark made by teeth or not. This approach should be considered in a logical way, assessing class, subclass and individual characteristics of the injury pattern, without the need to resort to 'suspicion', or 'experience' as reasoning for the opinion.

2.1.1 *Class characteristics.* The class characteristic of a bitemark is the presence of one or more arc-like cutaneous wounds, representing the approximate dimensions of a human dental arch form. An area of central ecchymosis is a variable class characteristic of a bitemark that may not be present in all injuries. Such an area is not necessarily representative of suction during the act of biting, is almost certainly *not* due to tongue-thrusting, but is likely to be a result of injury to small blood vessels as they are compressed during the bite.

2.1.2 *Sub-class characteristics.* The sub-class characteristics of a bitemark are represented by discrete portions of the injury that could reasonably be ascribed as having been made by individual teeth. It is mainly on the comparison of sub-class characteristics, such as position and degree of rotation, that associations are made between a dentition and a bitemark in human skin.

2.1.3 *Individual characteristics.* Even more discrete details that reveal information about individual teeth, such as areas attributable to notching and differential wear of teeth, are considered individual characteristics. It is doubtful that individual characteristics are represented with any accuracy on a cutaneous bitemark injury due to the nature of human skin, however, individual characteristics may be represented in bitemarks on inanimate objects.

**2.2 Human or animal.** It may be difficult to ascertain the origin of a bitemark as being human or not. Odontologists should make this call on the basis of class and sub-class characteristics, in addition to considering factors such as the size and shape of the injury.

**2.3 Possibility of self-infliction.** Odontologists should consider the possibility of self-inflicted bitemarks. The position of the wound on the body may obviously indicate whether the wound could have been self-inflicted or not.

**2.4 Ageing.** While most odontologists have training and experience in assessing cutaneous injuries, there is little evidence to support the notion that bitemarks can be aged with any degree of accuracy. Statements beyond that of claiming that an injury is either 'old' or 'recent' cannot be reasonably supported.

**2.5 Adult or child.** Arch width has often been cited in the odontology literature as a defining feature of an adult versus child bitemark. Such calls need to be made with caution as there is wide overlap between the possible ranges of inter-canine width for adults and children. Additionally, arch-width has been cited as one of the most distorted metrics in bitemarks on human flesh.

2.5.1 Indicative inter-canine widths of less than 25mm have been generally considered to represent that of a child's arch, but measurements of less than 21mm have been recorded for adults.

**2.6 Level of force.** No study has yet been conducted on the amount of force necessary to produce different types of bitemark injuries, which range from diffuse bruising to avulsion of portions of human flesh.

2.6.1 The presence of a bitemark, even one of diffuse bruising, implies a significant amount of force was generated during the time of injury. Beyond that, other factors such as the morphology of the teeth and the circumstances of the biting influence the appearance of the wound to an as yet unquantified degree.

2.6.2 The severity of a bitemark is not necessarily linearly proportional to the amount of force used or the degree of violence experienced by the victim, and odontologists should avoid making statements that imply such.

**2.7 Description of bitemark.** Odontologists should provide a detailed description of the suspected injury, even when photographs are taken. Anatomical and pathological terminology should be used where indicated to describe the location and nature of the injury

**2.8 Quality of evidence.** Odontologist should consider the relative quality of both the material evidence as well as the bitemark itself, as this in turn should appropriately moderate any conclusions that are then drawn from assessment of the injury.

2.8.1 Assessment of the quality of the injury itself provides guidance on the potential forensic value that the bitemark has.

2.8.1.1 Injuries of medium severity usually represent the most forensically valuable material.

2.8.1.2 In general, injuries at the low end (diffuse bruising) and high end (complete avulsion) of the severity spectrum carry less forensically significant information. Odontologist's opinions should be appropriately weighted in order to account for this.

2.8.2 Assessment of material quality includes consideration of the technical aspects of photographs offered for assessment including:

2.8.2.1 The presence of an ABFO No. 2 scale, its alignment with the bitemark injury and its position in the same plane.

2.8.2.2 The position of the camera lens at the time the photograph was taken, and the resulting amount of distortion (and whether it is correctable or not).

2.8.2.3 The relative contrast and saturation of the bitemark images.

2.8.2.4 The presence of specular reflection that has potentially washed out surface details in the injury.

2.8.2.5 The number of other photographs of the injury, and the consistency of appearance of the injury in these alternative views.

## Guideline 3

---

# Evidence Collection from a Suspect

---

### Preamble

Collection of dental evidence from a suspect is necessary in order to provide material for comparison with that collected from the victim. Suspected biters are usually cooperative, however, as the dental examination is considered a forensic procedure, their (or their guardian's) consent to undergo such an examination is generally considered necessary. The relevant *Crimes Act* in a federal, state or territorial jurisdiction authorises the carrying out of a forensic procedure on a suspect, serious offender or volunteer with the 'informed consent' of that person. Collection of dental evidence for bitemark comparison falls into this category. Courts, and in some cases, police officers, can compel unwilling suspects to undergo dental examination under various clauses of the *Crimes Act*, however, the legal technicalities of these situations are complex and beyond the scope of these guidelines. It is recommended that forensic odontologists seek advice if confronted with a refusal to undergo a forensic dental examination.

### Scope

This guideline details recommendations for the collection, recording and documentation of dental evidence from a suspected biter.

### 3.1 Consent and authority.

- 3.1.1 Most Australian jurisdictions require that the collection of physical evidence from a suspect be carried out either with a witness present, or that a video of the procedure is recorded. It is recommended that dental evidence collection from a suspect be carried out in the presence of a chaperone where possible, even in the presence of video recording equipment. This may be another health professional or law enforcement officer.
- 3.1.2 There is a legal requirement, under the *Health Practitioner Regulation National Law Act (2009)* for a registered dentist to take impressions and conduct dental charting in person on a living suspect. Such activities *may* be delegated to persons holding other forms of registration under the Act, (such as a dental hygienist or therapist) but only on the direct authority of a registered dental practitioner with whom they have a structured professional relationship. In general, this situation should be avoided, and odontologists should be called to perform this examination. Odontologists have

normally undertaken training above and beyond general dental practitioners, which makes them best qualified to carry out a forensic dental examination of a suspect (however, see section 3.7 below).

**3.2 DNA swab.** A standard buccal swab should be taken for DNA comparison. Some laboratories may prefer a whole-blood sample, and this may vary between jurisdictions.

3.2.1 DNA samples need not be directly taken by the odontologist, but he or she should ensure that such reference samples either have been or will be taken for comparison with the sample obtained from the wound area.

**3.3 Photographic records.** Photographic representation of the suspect's dentition is important to verify the accuracy of the models used in the analysis phase.

3.3.1 It has been recommended that full facial and profile shots be taken, however, their exact purpose remains undescribed in the literature. As a minimum, occlusal surface views, as well as frontal and lateral views of the dentition both in occlusion and with the jaws open should be taken.

3.3.2 An ABFO No. 2 scale should be included where possible in these photographs. It is important that this be present in the same plane as the dentition, and not that of the mirror, when using mirrors to obtain occlusal and lateral views.

**3.4 Dental charting.** A chart of the suspect's dentition should be recorded using standard FDI and charting protocols. It is recommended that a written description of each restoration is also recorded, rather than having to rely on interpretation of charting symbols when analysis proceeds at a later date.

3.4.1 A thorough dental examination should be conducted with a mirror, probe and good lighting, with the odontologist noting the details of any missing teeth, restorations, chips or defects present, especially in the anterior segment.

3.4.2 Maximal opening should be measured and recorded at this stage, as this may assist to exclude or include a suspect when compared with similar measurements taken from the suspect mark.

3.4.3 Occlusal anomalies that may effect the mechanism of biting, such as a severe Class II Division I malocclusion should be noted.

3.4.4 Dental records need not necessarily be obtained, however, they may provide useful corroborating evidence regarding the nature and timing of any restorations carried out that may affect the morphology of the anterior teeth.

**3.5 Impressions and bite registration.** Impressions from a suspected biter are the most essential form of physical evidence for bitemark comparison exercises.

3.5.1 Upper and lower impressions should be taken in the highest-quality medium available. Ideally, they should be taken in a polyvinylsiloxane or other similar rubber-based medium using a putty-wash technique.

- 3.5.2 The use of custom trays is not necessary, as the literature now demonstrates that well-fitted stock-tray impressions using modern PVS materials have minimal dimensional distortion that is unlikely to affect the morphology of the anterior dentition for the purposes of forensic comparison.
- 3.5.3 If a rubber-based material is not available, alginate may be used. Most alginate materials are now relatively dimensionally stable for at least 12 hours, providing they are stored correctly. The manufacturers instructions should be consulted in this regard.
- 3.5.4 Impressions with air bubbles or other defects on the incisal or occlusal portion of teeth must be categorically rejected and subsequently re-taken. It is unacceptable to 'alter' models once they are poured in order to correct for these defects.
- 3.5.5 Impressions should be poured as soon as practical after the procedure in Type IV stone (or another similar hard-wearing dental stone).
- 3.5.6 The use of rubber-based impressions will generally allow a second pour of each impression. In this case, one set of models should serve as a working set, with the other subject to as little handling as possible to serve as master models. It is recommended that master models be mounted on a plane-line (hinge) articulator in order to preserve their condition.
- 3.5.7 Alginate impressions should not be poured a second time. The material is likely to tear during removal from the first model, and in any case will likely undergo distortion during the exothermic set of the dental stone. Where impressions are taken in alginate, it is recommended that two upper and two lower impressions are taken, to serve as master and working models respectively.
- 3.5.8 A record of the suspect's occlusion should also be gathered via a bite registration. Base/catalyst bite registration materials such as BluMousse® (or similar products) should be considered first-line materials in this regard. Wax bites are dimensionally unstable over even short periods of time and are not the preferred means of recording occlusal relationships.
- 3.6 **Bite sample.** Collection of a bite sample is not the same as collection of a bite registration, and fulfils a different purpose. A bite registration is gathered in order to allow accurate association of upper and lower models. A bite sample demonstrates the pattern of the suspect's teeth during their own particular physiologic action of biting.
  - 3.6.1 Bite samples are preferentially recorded in polyvinylsiloxane putty-consistency material. Wax bite wafers are subject to greater dimensional instability and should only be used when putty is not available. A single layer of baseplate wax should be considered unsuitable for recording a bite sample.
  - 3.6.2 A bite sample should not involve a through-and-through bite. There should be sufficient thickness of material between the upper and lower teeth in order that the sample maintains its rigidity once set. The goal is to record the indentations made by upper and lower teeth to a depth of approximately 1mm.



3.6.3 Where wax wafers are used, they should be photographed with an ABFO No. 2 scale in place as soon as they are recorded, in order to provide verification that no distortion has occurred between the recording of the bite and the comparison phase.

3.7 **Third-party evidence collection.** It is recognised that it is not always feasible for odontologists to obtain first-hand physical evidence from a suspect. While this situation is not ideal, occasionally the suspect may need to have impressions taken by a third-party practitioner who in all likelihood will not have had any forensic training.

3.7.1 In the event that a third-party examination provides the only practical means of gaining dental evidence from a suspect, it should be considered essential that the odontologist communicates with the practitioner responsible for collection of this evidence prior to the examination.

## Guideline 4

---

# Methods of Comparative Analysis of Suspect Dentitions to Bitemark Injuries

---

### Preamble

Evidence collected from both victim and suspect are compared in the analytical phase of bitemark comparison. During the process of comparative analysis, the odontologist attempts to draw a conclusion regarding the nature of correspondence between a latent mark and the morphological features of the suspect dentition. This guideline should be read in conjunction with Guideline 7 regarding the potential for contextual bias and observer error in this process.

### Scope

This guideline articulates recommended methods of comparison of a suspect dentition to a bitemark injury. Technical details of how these comparisons should be carried out are beyond the scope of this document. Relevant texts in forensic odontology should be consulted in this regard.

**4.1 Methods of comparison.** A dentition may be compared to a bitemark in both qualitative and quantitative ways.

- 4.1.1 The method endorsed by the literature as being the most accurate is that described by Johansen and Bowers, using an qualitative 'overlay' technique. This should be considered the primary method of comparative analysis.
- 4.1.2 Other means of comparison should be considered supplementary to this method, but may prove useful in corroborating the opinion reached on the basis of qualitatively comparing hollow-volume overlays with the bitemark
- 4.1.3 Odontologists should be comfortable with the principles of distortion and subsequent manipulation and correction of images as described in this method. A failure of this understanding, in addition to not being able to explain the principles and methods involved in its application to a judge or lay jury is likely to prove devastating to the odontologists credibility as an expert witness.

**4.2 The overlay technique.** Production of an overlay should be achieved via the use of digital manipulation software such as Adobe Photoshop®.

- 4.2.1 **Overlays.** Hollow-volume overlays should initially be produced from a scanned bite record, or in the absence of a bite record, a wax impression of the indentations made by the anterior teeth as made by the odontologist from the models of the suspect. This provides the maximum level of detail with minimal 'selection bias' from either the odontologist or the computer software. Scanning should proceed at at least 300 dpi.
- 4.2.1.1 Overlays may be produced via other methods, such as from scanned models or wax scrapings, however, these methods introduce biases that cannot be compensated for. Consequently, they should be considered as supplementary, or secondary means of producing overlays.
  - 4.2.1.2 Compound overlays should likewise be considered as supplementary means of demonstrating the relationship between a dentition and a latent mark.
- 4.2.2 **Bitemark image.** It is usually necessary to correct for distortion in the digital images of the latent mark, as rarely is the photography so perfect as to represent perpendicularity in two dimensions relative to the bitemark.
- 4.2.2.1 Correction for distortion involves both scaling (or re-sizing), in order to obtain a dimensionally comparable image with that of the overlay, and correction of angular distortion. The technique for this correction is well-described by Johansen and Bowers
  - 4.2.2.2 Manipulation of image qualities such as contrast, brightness, colour, and opacity are acceptable, providing a record of these alterations are kept.
  - 4.2.2.3 Dimensional manipulation of the images to correct for any other form of image distortion introduce non-accountable errors and are not endorsed as a means of 'correction for distortion'
- 4.2.3 **Comparison.** The comparative phase of analysis occurs when the overlay and the digitally corrected image of the bitemark are compared with one another.
- 4.2.3.1 The primary means of comparison between a dentition and bitemark image should be one based on the similarity of morphology of the hollow-volume overlay and the digitally corrected image of the injury.
  - 4.2.3.2 The metric, angular and non-metric methods described by Johnsen and Bowers for comparison should be considered supplementary techniques useful for supporting a conclusion based on comparison of the hollow-volume overlay and the corrected image of the bitemark. There is no support in the scientific literature for using these methods as a primary means of demonstrating concordance between a dentition and a bitemark.

**4.3 Other methods of analysis.** The method described by Johansen and Bowers is considered to be the most accurate means of overlay production and its use for subsequent comparative analysis has likewise been demonstrated to be the most reliable.

- 4.3.1 Hand-tracing of the incisal or occlusal surfaces of the dentition on clear acetate for use in subsequent comparison with a bitemark image is highly subjective and is not endorsed as either a primary or supplemental method of overlay production.
- 4.3.2 Similarly, the use of life-sized photographs that have been uncorrected for distortion as direct comparisons to the overlay is to be discouraged. This is not to say that these images cannot be used, however, they should be scanned appropriately and corrected for angular distortion prior to any comparison exercise being undertaken.
- 4.3.3 Radiographs produced from radiopaque material applied to a wax bite have been described as a useful method of generating a comparative tool, however its accuracy is still less than that of a digitally produced overlay method. It may have a role as a secondary means of analysis but should not form the primary means of comparison.
- 4.3.4 Use of acrylic templates of the suspect dentition has been described in the literature as a means of comparison with life-sized photographs via physical superimposition of the template over the image, however this has only been demonstrated in the form of a case study. No controlled, comparative study has been conducted regarding the accuracy of this technique and conclusions based on the association between direct placement of models on a cutaneous injury (or photograph of an injury) is highly subjective.
- 4.3.5 The amount of distortion that injuries undergo when subject to excision from the body has not yet been established, but it is reasoned to be potentially significant. Comparison of a suspect dentition with a mark present on an excised portion of skin is likely to produce highly questionable associations. No conclusion of *any* degree of similarity can be reasonably held for comparisons conducted in this way.
- 4.3.6 Other myriad ways have been described both anecdotally and in the literature regarding methods of bitemark injury and dentition analysis that can lead to varying conclusions of association. Only four of these methods; digital production of overlays, hand tracing of overlays, production of overlays from a radiographic outline, or from a photocopy of a model; have been compared for accuracy via a scientific study. The use of any other technique for analysis of the dentition and image should be approached with caution.

## Guideline 5

---

# Terminology Regarding the Similarity Between Bitemark Injuries and Dentitions

---

### Preamble

Anecdotally, it is rare for odontologists to be asked to identify suspects on the basis of a bitemark. The advent of other biological tools such as DNA analysis has resulted in other, more accurate means of proving identity and association between latent marks and source objects. Regardless, there are times when odontologists may be asked to draw inferences regarding the relationship between a bitemark injury and a suspected dentition. The privilege of expert witnesses under our legal system allows such inferences to be drawn, however, this must be balanced with the scientific evidence regarding the ability to associate latent marks with individual dentitions. The literature is scant in this regard, and so a conservative stance is recommended.

### Scope

This guideline specifies endorsed terminology for the conclusionary phase of analysis of a bitemark and suspect dentition.

**5.1 The forensic question.** Odontologists may be asked several questions regarding the nature of the relationship between a bitemark and a suspect dentition. The conclusions reached by the odontologist at the end of the analysis phase need to accurately, yet conservatively, reflect the question posed by the requesting authority.

**5.2 Conclusions of identity.** Conclusions regarding the identity of a suspected biter are rarely justified. Terminology relating to identity, as used in disaster victim identification protocols, with the possible exception of 'exclude' is not endorsed for use as conclusionary statements in bitemark comparison reports.

**5.2.1** The literature does not support the notion that perpetrators can be positively identified from bitemarks alone. The general lack of representation of individual features of teeth that may otherwise be used to identify an individual, such as those used in DVI exercises and as described in Guideline 2.1.3 means that definitive identifications from bitemarks should be considered technically impossible. Accuracy rates for 'identifications' from bitemarks have been described in the order of 60 to 80 percent, however, these studies are deeply flawed, and have all involved small populations attempting to identify individuals on the basis of a very small number of

cases and possible perpetrators. Consequently, use of the terms ‘identified’, ‘positive identity’, ‘match’, or any other words that connote a definitive link between a particular suspect’s dentition and the bitemark are not endorsed as conclusions regarding the relationship between them.

- 5.2.2 Use of the terms ‘probable’, ‘likely’ or ‘more likely than not’ are similarly not generally endorsed when describing the association between a dentition and a bitemark due to the inherent variation in their meanings between practitioners. The exception to this would be where an odontologist is asked to distinguish between multiple suspects as to the relative likelihood of one dentition being the origin of the bitemark when compared to another given dentition. Such conclusions may be warranted in these situations, however they then only refer to the probability within a given closed set of dentitions offered for analysis, and do not reflect the probability of *another* dentition having made the mark outside the range of exemplars given. This must be emphasised when tendering such opinions.
- 5.2.3 The term ‘possible’ may be acceptable, however, is unlikely to provide the agency with any meaningful information without a subsequent definition of what the odontologist defines as ‘possible’ and conversely, ‘not possible’. Similarly, the term ‘consistent with’ is vague and appears to vary widely between individual practitioners. The most accurate non-misleading term to convey the possibility that the dentition and mark may be related while recognising that they also may *not* be related, is that of ‘inconclusive’.
- 5.2.4 There is evidence to suggest that it is possible to exclude suspects on the basis of comparison of a dentition with a bitemark. Such conclusions need to be based on obvious disparities between the bitemark and the geometric arrangement of the suspect’s dentition in both pattern *and* metric analysis.

**5.3 Uniqueness.** Use of the term ‘unique’ or any similar phrase that connotes individuality of the dentition or of bitemark patterns on flesh or inanimate objects remains unsupported by any scientific literature and should be avoided. The terms ‘unique’, ‘individual’ and their related derivatives (‘uniqueness’, ‘individualisation’, ‘individuality’ and so forth) add little to the weight of argument for ‘identifying’ or otherwise potential suspects.

**5.4 Relationship between a bitemark and a dentition.** It is entirely appropriate for odontologists to describe the similarities and differences between a bitemark and the morphological features of a dentition. The hollow-volume overlay provides the primary tool for comparison in this regard. Association between a dentition and a bitemark is made primarily on the basis of accumulation of sub-class characteristics.

- 5.4.1 Terms used to describe the association between a dentition [overlay] and a bitemark may be used such as ‘strong’ or ‘weak’, however, conclusions in this regard should not consist of one or two word phrases. Odontologists should describe exactly what they mean when they use these terms. The following are given as examples:

- 5.4.1.1 *Strong association:* There are a significant number of points of correspondence between the dentition [overlay] of the suspect and the bitemark. There are no features in the bitemark that cannot be correlated with the morphology of the anterior teeth.
- 5.4.1.2 *Weak association:* There are a significant number of points of correspondence between the dentition [overlay] of the suspect and the bitemark, however, there are features present in the mark that do not correlate well with the morphology of the teeth, and/or the metric analysis suggests discrepancies in their spatial arrangement.
- 5.4.1.3 *Minimal association:* There are numerous features present in the bitemark that cannot be correlated with the morphology of the dentition, and/or the metric analysis suggests that there are numerous discrepancies in their spatial relationship. However, there may be some general points of correspondence between the dentition [overlay] of the suspect and the bitemark.
- 5.4.1.4 *Significant disparity:* There are [virtually] no features that correlate between the bitemark and the morphology of the anterior teeth. There are multiple features of the bitemark that cannot be related to the morphology of the anterior teeth, and there are significant metric differences between the bitemark and the dentition in question.

5.4.2 The use of such terminology does not imply that there are minimum thresholds for points of correspondence (or points of disparity) for each of these phrases. Odontologists are entitled to use their discretion when assessing the weight of the relative number of such features, but they should then modify the relative strength of their opinion in accordance with this assessment.

**5.5 Use of statistical and probabilistic terms.** There is little evidence for the use of numerical terms when describing the likely association between a bitemark and a potential perpetrator. Large-scale studies attempting to describe the relative frequency of morphological arrangements of the dentition are yet to be conducted, although there is some data on the frequency of missing teeth. Such data is unlikely to significantly advance the argument for probabilistic association of a suspect with a bitemark.

- 5.5.1 Statements of numerical probability, expressed as either a percentage, decimal or other number are not endorsed given the current absence of relevant population data.
- 5.5.2 Statements of certainty, expressed as a number, percentage, or descriptive term carry no weight towards the acquisition of ground truth, and should be avoided. It is recognised that legal counsel have frequently asked expert witnesses in court 'how certain' they are regarding their conclusions, and in this circumstance the odontologist may be compelled to answer, however, inclusion of such a statement in a written report should be considered superfluous.

- 5.5.3 Statements of similarity should not be given in numerical or probabilistic terms, but should be given as descriptive phrases.



## Guideline 6

---

# Reporting

---

### Preamble

Odontologists should produce written reports of their findings when undertaking any form of analysis of a suspected bite mark. These reports may be tendered as evidence in coronial, criminal or civil court proceedings, and should be written with this in mind. Reports are entitled to reflect the opinion of the expert witness, in accordance with the privileges allowed a professional under the Evidence Acts. However, these opinions should be in keeping with the scientific status of the discipline. Unsupported opinion, conjecture and *ipse dixit* bear no place in bite mark reports. This guideline should be read and applied in conjunction with Guideline 5 *Terminology Regarding the Similarity Between Bite Mark Injuries and Dentitions*. A sample odontology report is included at the end of this series of guidelines.

### Scope

This guideline outlines the recommended form and content of a written report following assessment and analysis of a suspected bite injury.

**6.1 Introduction.** The report should contain a brief introduction that states how the odontologist came to examine the bite mark, including details of:

- 6.1.1 Who contacted the odontologist initially
- 6.1.2 Where the examination[s] took place and who was present.
- 6.1.3 What the examination[s] involved, and what records were taken. Odontologists should include details of the examination of both the victim and the suspect if they were involved in both.

**6.2 Description of the evidence.** The report should give a detailed description of the evidence, including both a description of the suspected bite mark injury and the suspected dentition. The description should use correct anatomical and pathological terms.

- 6.2.1 The description of the bite mark should include a description of where on the victim the mark was found, as well as a general description of its colour and form, including approximate measurements of the size of the injury.

- 6.2.2 An objective assessment of the quality of the evidence should be included in the report. Concerns with the material quality of the evidence should be noted.

**6.3 Statement regarding whether or not the injury is a bitemark.** If the injury is ascertained to be a bitemark, the report should state this as well as the reasons for reaching this conclusion. It should also state whether it would have been possible to self-inflict the injury, and any other information that was gleaned from the qualitative assessment of the injury.

**6.4 Comparative Analysis.** If a comparative analysis was carried out with a suspect dentition, a brief description of the types of analysis conducted should be given.

- 6.4.1 It is not generally necessary to specify how each step of the error correction and overlay production was performed, although the odontologist should bear in mind that they may be asked to explain this in detail should the case proceed to court.

- 6.4.2 This section is not necessarily limited to text. Diagrams should be used to explain points of correspondence, points of disparity and other remarkable features of the injury where considered necessary.

**6.5 Concluding opinion.** The odontologist should make clear the association [or otherwise] between the features of the bitemark and the morphology of the suspect dentition.

- 6.5.1 There is no need to resort to one or two-word phrases in order to summate the evidence. Odontologists should carefully and precisely explain the conclusions reached in order to accurately portray the foundation for [and limits of] their concluding statements.

- 6.5.2 Where one or more dentitions were compared, a statement regarding which of the dentitions was more likely to have produced the mark may be made, although a statement clarifying that this does not necessarily imply that it was the only dentition that could have inflicted the injury should accompany this.

- 6.5.3 Statements regarding the potential identity of the perpetrator, with the exception described in 6.5.2 above, are not endorsed under these guidelines.

- 6.5.4 Percentage or numerical probabilities that describe the likelihood of the dentition having made the bitemark are not appropriate without reference to relevant population data. Such data does not exist at the present time.

**6.6 Supporting documentation.** Associated diagrams generated during the analysis phase, photographs, overlays and other such media may be included as part of the report should the odontologist feel they are necessary.

**6.7 Signature block and date.** As a legal document, it is usually required that the odontologist physically signs the report and dates it accordingly. Details of postgraduate qualifications can usually be summarised here in post nominal form. There is little added benefit to including details of these qualifications in the text of the report. Should the case proceed to court, details of postgraduate qualifications, education and experience

is usually explored in detail by legal counsel for the benefit of judge and jury, regardless of the text of the report.

## Guideline 7

---

# Contextual Bias, Cognitive Bias and Observer Effects in Bitemark Analysis

---

### Preamble

Psychologists have long recognised the effects of contextual and extraneous information on decision making. Such information renders the subject susceptible to both motivational and cognitive bias, yet it is difficult to assess the extent to which these influence forensic odontologists opinions as there have been no studies to date on this subject. It appears that the current practice of bitemark analysis is rich in sources of potentially biasing influences. Cognitive bias specifically refers to the psychological sway towards one opinion versus another as a result of having information extraneous to the task at hand – in other words, bias induced by ‘knowing’ something.

Context effects — psychological influences on decision making induced by knowledge of circumstantial information extraneous to the immediate task at hand — most obviously give rise to motivational bias, which may be conscious or unconscious, however, they may also give rise to cognitive bias, particularly when there is ambiguity in the choice between two alternative hypotheses. This latter form of bias is easily over-ridden when the evidence presents an obvious choice between two hypotheses, but becomes problematic when evidence is ‘borderline’, of poor quality, or ambiguous.

Bitemark analysis is particularly susceptible to both manifestations of bias due to the context in which it is collected and analysed, which is rich in subliminal information that renders the practitioner susceptible to motivational bias, and; due to the nature of the evidence itself – its ambiguity and potential for interpretation. In addition to the fundamental recognition that odontologists will be subject to some forms of bias, odontologists should also consider ways in which these effects can be minimised as they go about bitemark evidence collection, analysis, and comparison in order to minimise potential biasing effects until there is experimental data available to qualify under what circumstances, and to quantify to what extent, they influence our analysis and interpretation of bitemark evidence.

### Scope

This guideline suggests various ways in which contextual bias and observer effects may be reduced in bitemark analysis.

**7.1 Cognitive bias.** It is important to recognise that biasing influences cannot be simply 'willed away', because by their very nature they are not under the conscious control of the individual.

7.1.1 Awareness of the forms of cognitive bias is not enough to combat its effects on practitioner performance.

**7.2 Role and conformity effects.** The potential to feel like an advocate of justice for the victim [or suspect] can be minimised by engaging as little as possible with the victim, law enforcement agencies and lawyers.

7.2.1 Analysis should be conducted independent of these influences. This is most practically achieved by separating the phases of collection and analysis of odontological evidence.

7.2.2 Where possible, the odontologist who is responsible for collecting the evidence should not be actively engaged in any subsequent analysis. While this is ideal, it is recognised that this is sometimes unavoidable. This suggestion is most practically achieved where multiple odontologists work for the one institution.

**7.3 Emotive influences.** These can further be avoided by limiting the amount of extraneous information available to the odontologist responsible for analysis of the bitemark.

7.3.1 This includes analysis of the bitemark independent of knowledge surrounding the case in order to minimize emotional influences, before any viewing of the suspect's dentition in order to minimize target shifting, and before any other circumstantial evidence is revealed, such as the presence (or absence) of the suspect's fingerprint, or DNA, to minimize confirmation bias.

**7.4 Target Shifting.** This occurs when ambiguous evidence is subject to bias in its interpretation due to prior information gathered about what features to look for in comparison with a suspect dentition. This can be avoided by attempting to identify the potentially relevant features of the mark prior to viewing the suspect dentition.

7.4.1 Potentially relevant features are those that indicate the class, subclass and individual characteristics of the dentition, for example; marks from the upper versus lower teeth; marks that indicate the relative position of the incisal edges and position of the canine cusp tips; unusual or distinctive spatial arrangements, or the potential presence or absence of certain teeth

7.4.2 Analysis of the dentition should also proceed separately from that of the bitemark. Ideally, more than one dentition, from persons who are unknown as to their involvement in the case to the odontologist, should be presented for analysis in order to avoid the generation of a purely confirmatory hypothesis.

7.4.3 Only having analysed both the mark and the dentition separately should the odontologist then attempt to combine this information in a single analytical technique. Following the combined analysis, other information, such as the reported

position of the biter relative to the victim in an assault case, can be revealed so that the odontologist can assess the relevance of this new information to the former conclusion. This process still ensures that the odontologist receives information that affords them the greatest opportunity to generate meaningful conclusions while removing unnecessary and potentially biasing detail.

**7.5 Evidence quality.** Odontologists should consider avoiding analysis of bitemark evidence that is of poor or dubious quality, where the risk for contrast effects is greatest.

7.5.1 The susceptibility to contrast effects is demonstrated when the odontologist gradually begins to 'see' an association between the mark and the dentition, particularly after lengthy analysis.

7.5.2 In these circumstances, there is a real risk that the threshold for determining the significance of a particular feature of the bitemark is lowered as the analysis proceeds, and so the odontologist becomes susceptible to seeing things that aren't there.

**7.6 Avoiding statements of certainty.** Research has long demonstrated that there is no link between certainty and accuracy, and any statement regarding certainty is potentially misleading and irrelevant to proper assessment of the evidence.

7.6.1 Research has demonstrated that practitioners are generally over-confident in their ability to perform, particularly when performing routine or often-repeated tasks.

7.6.2 In addition to the potential for reaching a biased conclusion, overconfidence of the expert carries with it the sequelae of an unconscious biasing effect on juries and judges, who despite claims of impartiality, are still encouraged to include assessment of witness demeanour as part of the process of assessment of the expert's evidence.

**7.7 Independent verification.** Where practical, an independent evaluation of the analysis and comparative phases of the bitemark assessment should be carried out by another odontologist who has had no prior involvement in the case.

7.7.1 This practice follows that of other pattern-matching disciplines and acts as a quality control check. It may also highlight where an odontologist has potentially been unduly influenced by cognitive bias.

7.7.2 The purpose of the review is to ensure that no obvious errors have been made in the conduct of the bitemark analysis. It is unnecessary for the reviewing odontologist to comment on whether or not he or she agrees with the opinion reached, bearing in mind that such opinions should not generally offer statements of identity.

7.7.3 If a verification of a particular conclusion is sought, the reviewing odontologist should undertake their own, independent analysis prior to viewing the original method and conclusion, so as to avoid bias themselves

## Guideline 8

---

# Presentation of Bitemark Evidence in Court

---

### Preamble

Odontologists may be called to appear in court to give verbal evidence regarding the results of their conclusions about the association between a bitemark and the circumstances of a particular crime, including that of the potential identity of the suspected biter. The legal principles regarding the giving of expert evidence are complex, vary between jurisdictions, and are considered beyond the scope of this guideline. It is recommended that odontologists consult a legal text on expert evidence if they wish to familiarise themselves with these principles. Similarly, general principles of expert witness conduct are also beyond the scope of this guideline.

### Scope

This guideline offers guidance for odontologists who are called to give evidence as an expert witness regarding bitemark evidence.

**8.1 Goal of Expert Testimony.** Odontologists should recognise that the goal of expert witness testimony is to assist the trier of fact [a judge or jury] by interpreting complex, scientific, or other specialised forms of evidence.

8.1.1 As part of the judicial process, it is the expert's role to furnish the trier of fact with the tools and information necessary to allow them to draw their *own* conclusions about the significance, or otherwise, of the evidence in question.

8.1.2 It is not considered the expert witness's role to comment on the likelihood of the suspect's guilt or innocence: this is the task that the judge or jury are assigned with.

**8.2 Questions of Identity.** Counsel may press odontologists for opinions on the identity of the suspect. Odontologists should bear in mind the principles espoused in these guidelines when giving answers to these questions.

8.2.1 Expert witnesses are usually allowed to qualify their answers to questions. Counsel are generally not permitted to insist on simple 'yes' or 'no' answers to questions when it is obvious that such answers are potentially misleading.

**8.3 Presentation of Evidence.** Counsel will usually advise on the best way for the expert witness to present their evidence. Such testimony does not have to be purely verbal, and

the use of computers, overhead projectors, diagrams and other physical aids can assist the trier of fact in understanding the odontological evidence.

- 8.3.1 Images from the digital analysis of the bitemark evidence should be considered key aspects of the evidence presented in court. Every effort to demonstrate them to the trier of fact should be made.

**8.4 Common errors associated with forensic expert testimony.** There have been instances, particularly in the United States, of forensic science testimony being excluded due to unreliable testimony by experts. While the admission of forensic science testimony is considered liberal by contrast in Australia, avoidance of the following pitfalls ensures not only assignment of appropriate weight to the odontological evidence by the courts, but also good scientific practice.

- 8.4.1 Odontologists should not attempt to provide statistical information unless such information is obtained via credible, scientific sources. 'Estimates' on incidence, likelihoods and other population statistics should be avoided.
- 8.4.2 Similarly, odontologists should not attempt to indicate the reliability of their technique by referencing incredulous error rates. Judges are finding that witnesses attempting to make an 'absolute' identification of the defendant, identifications 'to the exclusion of all others', or claiming extremely low, or even non-existent error rates are not relying on credible scientific reasoning.
- 8.4.3 Odontologists should adhere to endorsed methods and protocols within the field. Experts who have used their own methods without having good reason not to follow the endorsed protocols have invited controversy over the validity of their methods.
- 8.4.4 Odontologists should be able to appropriately explain their methodology in court. This is particularly true for evidence that has involved digital manipulation of images.
- 8.4.5 Odontologists may be required to tender original documentation of their assessment, or to recall details that were noted at the time. It is recommended that odontologists keep all notes made during the collection, analysis and comparison phases of bitemark assessment.
- 8.4.6 The existence of observer bias has been noted as strengthening the case for exclusion. While it has yet to be a significant issue in Australian courts, (although it has been alluded to in a number of high-profile bitemark cases), it is recommended that documented steps be taken to avoid potential sources of bias during bitemark assessment.
- 8.4.7 Experts must testify within their field of expertise. For odontologists, this field relates to the marks made by teeth as weapons of violence. Odontologists are generally not considered experts in facial mapping, offender profiling, or other peripherally related subjects, unless they can demonstrate by way of training, study or experience that they have specialised knowledge in these fields. One of the most common reasons for excluding, either in part or in whole, expert testimony in Australia is due to the



witness attempting to opine on an area outside that deemed to be within their scope of expertise.

## Guideline 9

---

# Education, Training and Continuing Professional Development

---

### Preamble

The education, training and continued professional development of odontologists in Australia is subject to oversight by regulatory authorities. Statutory authority for the practise of bitemark analysis as a registered Forensic Odontologist is derived from Australian Health Practitioner Act National Law, and enacted by the Australian Health Practitioner Regulation Agency via the Dental Board of Australia. The provisions for maintenance of registration as a forensic odontologist are also specified by the Dental Board, however, this guideline seeks to provide further recommendations specific to the practice of bitemark analysis by odontologists.

### Scope

This guideline provides recommendations for training and CPD requirements in order to maintain currency in the field of bitemark analysis.

**9.1 Qualification as an Odontologist.** The Australian Health Practitioner Regulation Agency (AHPRA) via the Dental Board of Australia have endorsed minimum education requirements in order to be registered as a Forensic Odontologist in Australia.

- 9.1.1 Use of the title *Forensic Odontologist* confers all of the rights of practice associated with this specialist area, including that of bitemark analysis.
- 9.1.2 Continued registration by the DBA is conditional on the maintenance of a 60-hour minimum Continuing Professional Development (CPD) cycle over each three-year period beginning in July 2010.
- 9.1.3 As a registered odontologist, it is *expected* that some proportion of CPD during this cycle will be devoted to continuing education in the area of bitemark analysis and interpretation. The form of this CPD may be in any of the formats recognised by the DBA for this purpose.
- 9.1.4 It is *recommended* that odontologists also undertake some form of expert witness training on a semi-regular basis.

**9.2 Qualification as an expert witness in bitemark analysis.** Legislation and the common law recognizes that witnesses can be considered ‘experts’ and can testify thereto an opinion providing they can demonstrate they have specialised knowledge in the relevant field. Such knowledge may be gained through training, study or experience.

9.2.1 There is no tacit requirement under the *Evidence Acts* in any federal, state or territorial jurisdiction that an expert must be a registered forensic odontologist in order to testify regarding bitemark analysis, providing the court is satisfied that the witness meets the definition of having ‘specialised knowledge’ in the field.

9.2.2 However, the evidence of a witness who is not board-registered is likely to carry significantly less weight in court than that of someone with registration status. In keeping with best practice principles, it is expected that any dentist who wishes to practise bitemark analysis maintains registration as a forensic odontologist with the DBA.

**9.3 Maintenance of currency.** Bitemark evidence is rare in Australia, and presents very limited opportunity to engage in bitemark analysis via actual casework.

9.3.1 Mock casework should be considered an essential component of CPD given the relative infrequent opportunity that most practitioners will have to undertake genuine cases.

**9.4 Provision of education to other related disciplines.** Forensic odontologists should welcome the opportunity to engage with colleagues from other forensic disciplines, including law enforcement agencies.

9.4.1 Education of police photographers, forensic nurses, clinical forensic medicine specialists and rape/crisis centre workers should take priority as these professionals represent the primary recipients of living bitemark victims. Education regarding the appropriate procedures for contacting odontologists, and if necessary, collecting bitemark evidence from victims, is essential for these professionals and will only serve to enhance the services that odontologists can offer.

**9.5 Research.** There is a dire need to undertake further research into the validity of bitemark analysis. It is expected that odontologists who practise bitemark analysis will do their utmost to meaningfully participate in research of this nature as conducted by their colleagues.

## Sample

---

# Bitemark Report

---

At approximately 1800 on Monday 14 September 2010 I was asked to examine an injury suspected to be a bitemark on the body of a deceased 28 year old male, ID number 12333/10 by the on-call pathologist, Dr. William Smith. I conducted an initial examination of the injury at the mortuary in the presence of Dr Smith and the on-call mortuary technician at approximately 1845, my findings of which are detailed below.

### Examination

Examination of the deceased reveals an ovoid wound measuring approximately 35mm x 45 mm, the geometric centre of which is located approximately 12cm proximal to the styloid process of the ulna, on the dorsal aspect of the left arm. The lateral aspect of the wound consists of bruising, the outline of which is irregular in form. The medial aspect of the wound consists of similar bruising, along with four discrete areas of laceration superimposed, measuring approximately 2, 4, 4 and 3mm in length respectively (See accompanying photo). There is an area of contusion in the centre of the wound. The appearance of the wound at the time of examination is strongly suggestive of a human bitemark. A swab from the centre of the injury was taken by the attending pathologist for DNA. Documentation of the wound in addition to my written notes proceeded via photographs taken by the forensic photographer under my direction. There was no three-dimensional detail present in the injury at the time of examination, and it was not deemed worthwhile to take impressions of the area as they would contribute little to further analysis.

### Analysis

Upon obtaining the digital images of the wound, further analysis was undertaken. The photographs of the injury are of good technical quality. There is slight distortion in the images in that the lens of the camera was not positioned perfectly perpendicular to the ABFO scale, however, this effect is assessed as being minimal. The scale is well positioned relative to the injury, and there is good contrast and depth of focus.

The wound has both the class and sub-class characteristics expected of a human bitemark. The lateral aspect of the wound with its irregular outlined bruising suggests that this portion of the injury was caused by upper teeth. The external outline of this portion of the wound is representative of the relative dimensions of human upper anterior teeth. The medial aspect

of the wound demonstrates bruising and lacerations suggestive of the shape and dimensions of human lower teeth. Given this orientation, it would be extremely difficult for the deceased to have self-inflicted this injury. The relative dimensions of the wound are within the range of documented human dental arch sizes. The area of contusion in the centre of the wound represents injury to small blood vessels in this area, which is often caused by compressive forces. Its presence further strengthens the hypothesis that this injury was caused by teeth as this area of tissue would have been compressed during the act of biting.

### **Comparison**

I was further contacted by SGT Jones on Wednesday 15<sup>th</sup> September in relation to the possibility of conducting a comparison of the wound with a dentition of a person suspected of assaulting the deceased. Impressions of the suspect were taken by his local dentist, after consultation with me via telephone. Constable Evans brought the resulting evidence taken from Mr Suspect to my office on Friday 17<sup>th</sup> December. An upper and lower model of Mr Suspect's dentition, in addition to a test bite was handed to me on this date. The test bite was recorded in two trays of putty-consistency polyvinylsiloxane material and appeared to have undergone minimal distortion. It was possible to physically relate the test bite to both the upper and lower models, demonstrating this. A digital overlay was produced from both the test bite and scanned models of the dentition, as per the Johansen and Bowers method. Prior to comparison taking place, copies of the digital images of the injury were also corrected for angular distortion as described by the same authors.

Comparison of the digital overlay produced from the test bite and the injury reveals only a weak correlation between the two. Comparison in several orientations reveals that there is concordance in the general shape and dimensions of the suspect's upper arch with that portion of the injury assessed as having been made by upper teeth. The lack of detail in the aspect of the injury ascertained as having been made by upper teeth renders it difficult to carry out any meaningful comparison with the suspect's upper dentition, other than to comment on its general similarity in size and shape.

While there is general concordance found between the arch morphology of the suspect and a portion of the injury itself, there is no clear or consistent relationship between the marks suspected to have been made by lower teeth and the morphology of the suspect's lower dentition, even when considered in the position that demonstrates the most concordance. There is general agreement in the approximate dimensions of the suspect's lower arch and that demonstrated in the injury, however, the position of the lacerations and other sub-class detail does not conform to any morphological features of the suspect's lower teeth as evidenced in the test bite. There is no particular demonstration in the injury of marks that could be reasonably attributed to the suspect's well-developed canines, which would be expected to leave obvious points of reference. While the most proximal laceration towards

the inferior aspect of the arm of the deceased could conceivably have been made by the suspect's lower left canine, the presence of other lacerations then bear minimal, if any, resemblance to the positional arrangement of the suspect's other lower teeth.

Comparison with hollow-volume overlays produced from scanned models of the suspect's dentition reveals little in addition to the above findings. Again, there was a very general concordance in the size and shape of the upper arch, but there is a lack of convincing concordance between the morphology of the suspect's lower teeth and the sub-class characteristics demonstrated in the inferior aspect of the injury.

### **Conclusion**

The mark found on the deceased as described in this report is highly suggestive of a bite mark of human origin. Comparative analysis of the injury and the dentition of Mr Suspect reveals that while this suspect cannot be definitively excluded as having caused the injury, there is at best only a weak relationship between the morphology of the suspect's dentition and the pattern of injury found on the deceased. The correlation between the relative dimensions of the arch forms in both the injury and the suspect's dentition should be considered of low to moderate significance given the weak relationship demonstrated between the sub-class features of the injury and suspect dentition.

**Dr. R. Black**

BDS Sc Grad Dip For Odont FICD

Forensic Odontologist

12 October 2010