CONSIDERING FORENSIC SCIENCE: JUROR DECISION MAKING AND UNVALIDATED IDENTIFICATION EVIDENCE

Charlotte Rachel Scobie

Supervisors

Dr. Carolyn Semmler, Dr. Michael Proeve

School of Psychology

University of Adelaide

This dissertation is submitted for the award of Doctor of Philosophy

June 2018

Formatted in accordance with the
American Psychological Association publication guidelines (2010)

To my Mum, Dad and Grandparents.

Thank you for supporting and believing in me.

“Circumstantial evidence is a very tricky thing," answered Holmes thoughtfully. "It may seem to point very straight to one thing, but if you shift your own point of view a little, you may find it pointing in an equally uncompromising manner to something entirely different.”

- Arthur Conan Doyle, *The Boscombe Valley Mystery,*

Contents

List of Figures

List of Tables

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed:__ ____

Date: 13/06/2018

Charlotte Rachel Scobie

Abstract

Rapid scientific advances mean that new techniques and areas of research are being used by crime labs to test forensic evidence, but as innovations grow, so does fear that invalid science will make its way to the courtroom. If jurors and judges are not informed of the threats to validity that are associated with identification evidence they are at risk of overestimating the reliability of that evidence. The overarching goal of this thesis was to investigate whether scientifically informed opposing expert testimony or cross-examination will educate jurors about unreliable forensic science, and whether there are individual differences that will affect the perception of forensic evidence.

Study one investigated whether opposing expert testimony could educate jurors about anthropometric facial comparison evidence. In addition, participants' scores on measures of epistemological sophistication and argument skill were used to test for direct effects on verdict, and indirect effects through ratings on a measure of methodological reliability. Path analysis did not show support for relationship for the two individual difference measures. Opposing expert testimony was able to reduce ratings of the reliability of the forensic evidence, indicating that presenting participants with criticisms of unreliable forensic evidence is capable of informing jurors of limitations.

Study two aimed to replicate the main findings of the first study while testing a different measure of individual difference: bias towards forensic evidence, as captured by the Forensic Evidence Evaluation Bias Scale (FEEBS, Smith & Bull, 2012, 2014). Opposing expert testimony reduced reliability, and scores on the pro-prosecution subscale of the FEEBS led to higher ratings of reliability, indicating that when the participants were predisposed to see forensic evidence as highly trustworthy and conclusive they were more likely to convict. Qualitative analysis of responses justifying verdict choice showed that opposing expert testimony was informative, but that many participants struggled with understanding scientific methodology and had unreasonable expectations about forensic science.

Study three tested whether scientifically-informed cross-examination would lead to reduced reliability. Three types of forensic identification sciences were used: anthropometric facial comparison, fingerprint, and voice identification. Participants read through expert testimony regarding one of the three types of evidence, and then either scientifically-informed cross-examination, or questions that focussed on the qualifications and experience of the expert. Multi-group analyses and individual path analyses were conducted. Only the relationship between examination type and evidence type was different between groups, and scientifically-informed cross-examination did not affect ratings of reliability. Scores on the FEEBS affect neither reliability nor verdict. This suggests that differences in testimony and either the origin, or complexity, of criticisms towards evidence may have a large impact on verdict.

This thesis contributes to furthering our understanding of juror decision making regarding unreliable forensic evidence as it has demonstrated that perception of reliability, even if based on substantially biased or incorrect reasoning, will have the largest impact on verdict choice. The findings will be useful to researchers looking into the best ways of educating jurors and judges, as well as those calling for validation studies of forensic sciences.

Acknowledgements

To my supervisors, thank you both for your guidance, and assistance with developing my research and writing skills. Dr. Carolyn Semmler I am immensely thankful for your unwavering support, inspiration, and many a needed (loving) kick up the bum, and to Dr. Michael Proeve for encouragement and insightful questions. We got here in the end!

To those who helped along the journey: Margaret Prysak for assistance in coding participant responses, Sam Lane for data entry, James Alexander King for thoroughly checking all my references, Rebecca McEwen for proof-reading and fixing my legal terminology (as well as answering a hundred or so clueless questions), Matayo Moshi for putting up with my constant adjustments to the diagrams throughout the study, and to all the friends I hassled into recruiting and advertising my studies over the years. The professional staff at the School of Psychology and at the Barr Smith Library for administrative assistance and help with research. Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

A very special thank you to Louise Omer, Joy Dalay-On, Jessica Lockery, Kathleen Wang, Stephanie Clapp and (again) Matayo and Sam, Rebecca and James: thank you for board game nights where we laugh about cabbages, a thousand and one in-jokes, too many coffees, 2am cheese toasties, 1006 messages about video games, curated playlists, just because cookies, picnics by the river, hugs, couches to sleep on, and putting up with me never finish anything I start. Except this thesis. You were there when I needed, put up with me, you made me laugh, you've supported me, and you got me through in mostly the same level of sanity as I started. I love you all. And probably owe you all a few meals (and I promise to finally finish that tv show).

And to my mum and dad, Jean and Rob Scobie: thank you. Not just for the proofreading, lots of laughter and packed lunches, but for the love, compassion, and faith you've shown throughout my university career. You're why I'm sitting here typing this today.

.

1 INTRODUCTION

> *The law's greatest dilemma in its heavy reliance on forensic evidence, however, concerns the question of whether – and to what extent – there is science in any given forensic science discipline (National Academy of Sciences, 2009, p. 87).*

Trial by jury is a valued part of the Australian legal system (*R v Marshall*, 1986), one of the main ways lay persons are involved in the administration of justice (Goodman-Delahunty & Tait, 2005). It is an important component of democratic rule (Lempert, 2001), and famously referred to as "the lamp that shows that freedom lives" (Devlin, 1956, p. 164). Despite being used only in a minority of trials that reach court, the jury is still a common feature of trials involving the most serious crimes (Findlay, 2009; Goodman-Delahunty & Tait, 2005). But the role of the jury has always been susceptible to challenges and the need for reform (Chesterman, 1999). One of the most critical questions for current psychological research is whether a jury made up of ordinary persons is capable of understanding and applying complex scientific evidence to reach accurate verdicts (Lincoln & Lindner, 2004; McClellan, 2012). Critics of jurors argue they are "scientifically illiterate, irrational, and subject to emotional biases" (Goodman-Delahunty & Tait, 2005, p. 61), and two Australian judges, Justices Duggan and McClellan, have recently argued that trials with scientific evidence may be becoming too complex for jurors (Ackland, 2011; Keller, 2011).

Rapid scientific advances mean that new techniques and areas of research are being used by crime laboratories to detect and test forensic evidence. Laypersons place a large amount of trust in the various types of identification evidence (Garrett & Mitchell, 2016; Lieberman, Carrell, Miethe, & Krauss, 2008), considering forensic sciences to be widely available, persuasive, and highly reliable. However, jurors are being asked to evaluate increasingly complicated scientific evidence. In Australia, as a function of special status based on their training, study, or experience, scientists are allowed to give evidence when it is decided that jurors will be unable to interpret evidence without help (e.g. Evidence Act 1995, s79), which is similar to the post-*Daubert* amended Federal Rule of Evidence 28 U. S.C. (1975) in the United States, and "substantial helpfulness test" in the New Zealand Evidence Act (2006). While evidential complexity presents a major challenge, not all evidence is based on reliable scientific research, and there is escalating concern at the amount of unreliable expert evidence that is admitted into Australian, North American, and British courts (Cole, 2017; Edmond, 2012; Koehler, 2018; Lockhart & Satya-Murti, 2017; Porter & Kennedy, 2012; Thompson, 2008). In North America, the risk of substandard and unreliable forensic science was highlighted by the National Academy of Sciences in their report on forensic science (NAS, 2009), and reiterated by the recent PCAST report (President's Council of Advisors on Science and Technology, 2016).

Jurors are charged with evaluating conflicting evidence to reach the best verdict possible; a task that can be complicated by the addition of contested scientific evidence. If evidence based on flawed science is admitted to a trial, then jurors are vulnerable to poor decision making with the potential for miscarriages of justice. This concern is supported by findings from the Innocence Project, where unvalidated or improperly used forensics have been found to be the second-greatest cause of wrongful convictions (Innocence Project, 2016). Garret and Neufield (2009) claim that in the testimony of 137 prosecution experts taken from trial transcripts where wrongful convictions had occurred, 60% included misrepresentative empirical data, or unsupported conclusions. This is likely to be only a small portion of the trials where faulty forensic science may have contributed to a guilty verdict: in most cases no biological evidence is left at the scene, or it degrades or is destroyed before it can be reanalysed, and there is therefore no chance for DNA testing to suggest innocence (Dwyer, Scheck, & Neufeld, 2003). In Australia, recent cases where unvalidated forensic science contributed to convictions that were subsequently overturned include *R v Keogh, R v*

Wood, and *Chamberlain* v *The Queen.* What is even more concerning is that some studies are indicating that wrongful convictions due to flawed forensics may be increasing (DiFonzo, 2005). But despite attention from the legal community and media, little change has occurred to improve the basis or standards for forensic science. Known instances of flawed testimony are more likely to be labelled as the fault of a small number of fraudulent or careless scientists (Thompson, 2008), rather than systemic problems within the field of forensic science. Until substantial change occurs across forensic science disciplines, evidence of suspect reliability will continue to be placed before jurors. This compels us to know if jurors are capable of evaluating complex testimony to come up with accurate probative weight – including accounting for its flaws. Convincing a jury that forensic evidence may be unreliable is made additionally difficult by researchers having little insight into how each individual juror will assess the probative weight of the evidence.

Individual jurors respond to evidence differently, based on individual biases, ideas, and knowledge (Ellsworth, 1993): how else could twelve jurors seeing the same trial not return a unanimous decision every time? Social science and legal researchers have yet to propose an adequate account of the impact of individual differences on jurors. Previous studies attempting to predict jurors' verdicts based on single or combinations of stable characteristics have noted inconsistent findings (Ask, Reinhard, Marksteiner, & Granhag, 2010; Smith & Bull, 2012), and may have little utility in predicting verdicts (Ellsworth, 1993). However, De La Fuente, De La Fuente, and Garcia (2003) argued that this may be because individual differences may be mediating variables, rather than having direct effects, and further investigation of indirect effects may provide valuable insight into juror decision making.

But even if these differences can be identified, unreliable evidence is still reaching jurors. Several 'gate keeping' methods exist to assist in decision making: judicial instructions (typically given at the end of a trial) have been found to have little impact on jurors (Cutler, Dexter, & Penrod, 1989; Dartnall & Goodman-Delahunty, 2006), and cross examination (Kovera, McAuliff & Herbert, 1999; Krauss, Lieberman & Olson, 2004; Kraus & Sales, 2001) and the admittance of rebuttal (or "opposing") expert testimony show mixed results (Eastwood & Caldwell, 2015, Jones & Kovera, 2015; Levett & Kovera, 2008, 2009) . Within the wider corpus of jury decision-making literature, only a small number of studies have looked at the effectiveness of the latter two methods in educating jurors about unreliable evidence. And despite wide variances in the types of unreliable evidence used in the simulated trials, variables examined, and ways of assessing the effectiveness of the educational method, they are regularly labelled as ineffective. More research is need to investigate the potential, and the limitations, of opposing experts and cross-examination, and to continue work investigating individual differences that will affect perception of unreliable forensic evidence.

The overarching goal of this research is therefore to investigate whether scientifically informed opposing expert testimony or cross-examination will educate jurors about unreliable forensic science, and whether there are individual differences that will affect the perception of forensic evidence.

There are four key research questions (RQ) to address:

RQ1 **Can opposing expert testimony inform jurors about unreliable forensic science?**

RQ2 **Can cross-examination that focuses on the methodology behind forensic evidence, rather than the qualifications and expertise of the expert, inform jurors about unreliable forensic science?**

RQ3 **Is the effect of cross-examination on juror decisions similar to or different from that of opposing experts?**

RQ 4 **Are there individual differences that will affect jury decision making concerning unreliable forensic evidence?**

1.1 Overview of the Thesis

There are 5 chapters in this thesis.

Chapter 1: The current chapter presents an overview of why identification sciences are unreliable, and discusses previous findings from psychological studies on jury decision making research as it applies to forensic science.

Chapter 2: The second chapter examines RQ1 by presenting an empirical study testing the ability of opposing experts to educate mock jurors. Anthropological Facial Comparison was used as an example of unreliable forensic evidence. The study also address RQ4 by examining argument skill and epistemological sophistication, two individual difference variables that may impact the effectiveness of education.

Chapter 3: The third chapter presents study two which was a follow up of the first, but was conducted online and looked at a different individual difference: bias towards forensic evidence, addressing RQ1 and RQ4.

Chapter 4: The fourth chapter presents the final empirical study and investigated RQ2, the potential for scientifically informed cross examination to inform jurors, but also compared three different forensic identification sciences: anthropometric facial comparison, finger prints, and voice identification. Bias towards forensic evidence was also included in the predictions and analyses. Comparisons with the first two studies are discussed, addressing RQ3.

Chapter 5: The final chapter contains the conclusions and future recommendations.

1.2 Forensic Identification Evidence

Forensic identification sciences involve the comparison of two pieces of evidence to make source attributions: determining whether two items have consistent or inconsistent features, and ultimately indicating whether they have a common origin (Inman & Rudin, 2001). The origin of these sciences is as far as back as the Middle Ages, when the King's Court would call on experts to determine legal remedies for criminal activities, property disputes, and taxation (Bowers, 2014). In a modern criminal trial, testimony concerning forensic identification evidence such as DNA profiles, latent fingerprints, bite marks, tool marks, hair and fibre analysis, or facial comparison, among others, are commonly used to link a defendant to a crime (Hardy, 2007). Forensic identification evidence is a valuable tool in detecting and identifying perpetrators, and advances continue to make more evidence available for analysis: the use of Closed-Circuit television is increasing in most Australian states (Wilson & Sutton, 2004), increasing the potential number of crimes with photographic footage available, whereas current biological analyses are now able to test material samples as small as one nanogram (Caddy & Cobb, 2004). However, as the science and technology used by investigators becomes more specialised, and therefore harder to understand, it is crucial that we know if it is still within the capabilities of an ordinary juror to understand and incorporate into verdict decisions.

Judge Hand wrote that experts were used in legal disputes in the fourteenth century (1901), while Schofield (1997) found records of experts being used as early as 1620. There has been increased use of experts since the end of the nineteenth century due to rapid technological advances (Janusz, 2000). Today forensic science, and the experts presenting it, hold a dominant and unique position in the criminal legal system as well as in the public imagination. Scientific experts act as educators to decision makers and the justice system, and are intended to objectively analyse and communicate the evidence after "dispassionate evaluation" (Turvey, 2013, p. 40). Typically, expert evidence supports proof of a crime, and may help to identify the offender(s) (Stone, 1995). Increasing use is being made of expert forensic evidence, particularly by the prosecution, who are typically better able to afford such services (Stone, 1995). However, as Edmond, Hamer, and Cunliffe (2016) attest, prosecutors are typically not good at presenting forensic science, “warts and all” (p. 399), which over-inflates the probative value of the evidence whilst ignoring flaws within the testing processes. Ironically, the most scientifically sound procedure (DNA profiling) is also the most regulated (Giannelli, 2007), although even DNA evidence is vulnerable to many of the same sources of error as other identification sciences (Thompson, 1995). The next section will discuss the six main criticisms levelled at identification sciences by legal commentators, as it is important to distinguish these fields from other unreliable evidence, and to show the sophisticated methodological standards of scientific inquiry a jury is expected to be able to understand.

1.3 Unvalidated Science, Trusted Evidence

The NAS report argues: *"The degree of science in a forensic science method may have an important bearing on the reliability of forensic evidence in criminal cases. Two very important questions should underlie the law's admission of and reliance upon forensic evidence in criminal trials: (1) the extent to which a particular forensic discipline is founded on a reliable scientific methodology that gives it the capacity to accurately analyse evidence and report findings and (2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, the threat of bias, or the absence of sound operational procedures and robust performance standards" (National Academy of Sciences, 2009, p. 87).*

The field of forensic science, as a whole, has veered towards claiming instances of discovered wrongful convictions as being the fault of "bad apples", individual examiners with poor training, rather than acknowledging that cognitive bias or prejudice can lead to mistaken conclusions, and can affect any forensic scientist (Thompson, 2008). The world of forensic "science" is different to that of a normal science: findings are not backed up by the same standards in methodology, and it is rare to see controlled experiments, duplication, or peer review (Moriarty & Saks, 2005). Many scientists are aware of issues in the field. In a meta-analytic review of 18 international studies of forensic examiners, Fanelli (2009) found that while 1.97% of scientists admitted to having fabricated or modified data, 33.7% reported having used questionable research methods. While most scientists said they were ethical and honest researchers, the majority also said that, on average, 72% of their colleagues had used questionable practices. Clearly, unvalidated forensic science is systemic, and of major concern.

Identification evidence plays an important role in criminal investigations and trials, but if either members of the justice system or laypersons are unaware of the shortcomings then there is a risk of distorted confidence in the reliability of that evidence. It is a responsibility of the courts to inform fact finders of the method of production of the evidence, and of any information that would assist in assessing the probative weight. The following list explains the main criticisms levelled at forensic identification sciences currently in use in Australia, the United States, Canada, and the United Kingdom, and how they affect the reliability of the evidence given by experts.

i. Cultural Pressures and Practices

Turvey (2013) has gone as far as to say that law enforcement culture is incompatible with forensic science and scientific integrity, and Thornton and Peterson (2007) describe what they call a "fundamental conflict" in how science is used by the legal system. They argue that "the classical goal of science is the production of *truth*, while the goal of law is the achievement of *justice*" (p. 4, emphasis not mine). The cultural pressures put upon most forensic analysts mean that they are likely to adopt the aims of the police and justice system (Swann Jr & Giuliano, 1987): identifying perpetrators and ensuring justice is done. This ideology has led to forensic scientists adopting practices and cultural ideals that are described below. Each increases the chances of mistaken identifications. Due to an overabundance of cases very few investigations are able to undergo the meticulous, methodologically robust examinations that should be standard (Cooley, 2004).

Forensic experts will be motivated by a wide range of goals and pressures which will not always be compatible with disinterested scientific inquiry. Pressure to carry out analyses quickly can lead to changes in the way examinations are undertaken. Goals and motivation have a heavy influence on human behaviour and cognition (Moskowitz & Grant, 2009), and the social environment is a powerful source of goal-activation (Chartrand, Dalton, & Cheng, 2008), with studies showing that this is at work in the forensic laboratory. Testing is typically done with the intent of forming a case for the prosecution (Thornton & Peterson, 2007), which has the potential to diminish the objectivity of examiners. Ask et al. (2011) found that when experienced criminal investigators were exposed to norms promoting efficiency rather than thoroughness, they engaged in shallower appraisals of material provided to them, and the investigators became less willing to consider conflicting evidence provided at a later stage in the investigation. Goal activation happened outside the investigators' awareness: they were unaware that they had been primed towards a particular goal.

The cultural pressure of working within the justice system, whether it is due to a high profile case or institutional pressure, makes all investigators, including forensic experts vulnerable to "tunnel vision" (Martin, 2002). Findley and Scott (2006) describe tunnel vision as an amalgamation of heuristics and biases that lead parties involved in a criminal case to focus on a single suspect so that evidence is filtered into whether it supports or contradicts their involvement. They also argue that it is the overriding problem leading to wrongful convictions.

Experts are frequently in direct and open contact with the police (Saks, Risinger, Rosenthal, & Thompson, 2003), and it is customary for experts to receive incriminating materials with only a single set of comparison images or samples, and to be told of police suspicions (Edmond, 2013b). Tests are done to see if samples from a suspect match those from a crime scene; these are primarily only confirmatory tests, which are more consistent with the comparison strategies used by people in everyday life, not scientific inquiry (Swann Jr. & Giuliano, 1987). In a review of four crime labs in the US, researchers found that fewer than 10% of reports dissociated a suspect from a victim or crime scene (Peterson, Mihajlovic, & Gilliland, 1984). Tunnel vision can also work by lowering the threshold by which identification decisions are made (Risinger, Saks, Thompson, & Rosenthal, 2002), which is most “potent” when evidence is ambiguous. Again, this increases the chance of a wrongful conviction.

Providing forensic examiners with single comparison samples, and informing them – either directly or indirectly – that that sample is from a suspect, does not protect against confirmation bias (discussed below). While contextual influences will likely remain part of forensic science, little research has examined if jurors are aware of the differences between conventional "white coat" scientists and forensic examiners, and if this affects their view of the reliability of evidence, or their reception to other criticisms about forensic methodology. Interestingly, the popular view of forensic scientists on television is as an examiner who is heavily involved in the investigation of the crime and who is likely to be aware of the suspect’s identity (Cole, 2015; Saks & Koehler, 2008).

The final damaging aspect of the culture of forensic sciences is the idea, even unspoken, that forensic sciences are capable of individualisation: the reduction of the potential group of donors (which could be people or items such as weapons, glass, paint etc.) to a single potential source (Inman & Rudin, 2001). The idea that it is possible to find a combination of characteristics in an object, person, or animal, that have so much variance, you can link evidence from a crime scene to the source to the exclusion of all others in the world (McQuiston-Surrett & Saks, 2009). Saks and Koehler (2008) argue that "the concept of individualisation, which lies at the core of numerous forensic science subfields, exists only in a metaphysical or rhetorical sense. There is no scientific basis for the individualisation claims in forensic sciences... [it is] sustained largely by the faulty logic that equates infrequency with uniqueness" (p. 205). The assumption of uniqueness of a source, and that forensic evidence is able to find it, means that scientists can make definitive, confident claims.

Why is this a problem? To argue for individualization (or a very low chance of a match with a different individual), experts would need to establish that the axioms of probability theory hold for the type of features that make up the individual. One of these axioms is that the probabilities or frequencies of each feature within the population is independent from every other feature. This then allows use of the product rule to then come up with a probability that the identified individual could be mistaken for someone else within that population (Saks & Koehler, 2008). As Cole (2011) points out, a large corpus of authors and researchers have argued that it is unclear if there is a type of evidence that could individualise (e.g.Saks & Faigman, 2008; Saks & Koehler, 2008; Stoney, 1991). Unfortunately, in many places “individualisation” is used synonymously with “identification”, making it even more difficult for laypersons to distinguish between the two.

ii. Proficiency Testing and Peer Review

In most forensic identification sciences experts are not required to undergo any testing to prove aptitude or expertise (Edmond, 2013b), although the basis of an expert’s qualifications may be challenged at trial. The exception is latent fingerprint analysis, although even then there have been criticisms raised of the testing protocols (Mnookin, 2003). More and more voices call for the implementation of accreditation of forensic laboratories, standardization of written protocols for techniques, certification of examiners, and external proficiency testing (Giannelli, 2008; President’s Council of Advisors on Science and Technology, 2016). Proficiency testing would support claims that experts had a high quality of performance, were aware of the influence of psychological factors influencing identifications, and would assist in calculating error rates, for Type 1 and 2 errors (Koehler, 2017). Ultimately, it would help establish that identifications were reliable.

Peer review of methods used by an individual examiner or laboratory would also make sure that other experts in the field considered the techniques reliable, and, ideally, an internal peer review system would systematically check results of analyses. Budowle et al. (2009) argue that the best system to implement for peer review of individual cases would be that of blind verification, where a second examiner is "blind" to knowing the conclusions drawn by the initial expert, or the case history. In the field of DNA profiling, work is being done to test sequential unmasking of DNA evidence, where an initial expert is informed of the case and makes decisions about which samples are more meaningful to analyise, and then a second examiner compares samples without knowledge of the case (Peterson, Lin, Ho, Chen, & Gaensslen, 2003). Either method could reduce reduce the chance for bias. While debate continues over the implementation of a peer review process, blinding, or masking proceudres, the current system of of analysis leaves comparison results vulnerable to error, and no experimental studies have attempted to explain the vulnerability to jurors.

iii. Lack of criteria

While most types of forensic analysis do use computers as a part of processing the evidence, the ultimate judgement is made by a human examiner (Kassin, Dror, & Kukucka, 2013). Even in DNA analysis, which uses instruments to show genetic characteristics, or alleles in samples, (Thompson, Ford, Doom, Raymer, & Krane, 2003), data can still be ambiguous to the point where examiners can differ in interpretation (Dror & Hampikian, 2011). However, DNA matching does contain objective standards for what makes a match, whereas in many types of analysis there are no definitive, clearly-written standards for determining a match or exclusion (Thompson & Cole, 2007), so that each analyst is left to make a judgement without guidelines. Examiners themselves decide what is "sufficiently similar" without needing to meet a set of standardised requirements (Kassin et al., 2013).

Some disciplines rely on scales, or degrees of certainty, where terms such as "reasonable medical certainty", "excluded" or "inconclusive" are used (American Board of Odontology, as cited in Thompson & Cole, 2007). However, Thompson and Cole have argued that while this may bring a unified phrasing into the courtroom, it does not change the subjective nature of the analysis (Thompson & Cole, 2007). Martire et al. (2014) argue that verbal descriptions may lead to differences in understandings between jurors, as there is a wide range of evidence showing that meanings given to words can vary between people, and between contexts (Brun & Teigen, 1988; Budescu & Wallsten, 1995); and this has been found in studies of forensic evidence (McQuiston-Surrett & Saks, 2008; Nance & Morris, 2005). McQuiston-Surrett and Saks (2009) found that telling jurors the defendant was the source of a trace, or that the trace and reference sample were “similar in all characteristics” both set probative value of evidence so high that it might as well have been a ceiling effect. Each response was interpreted as being so indicative of guilt that there was little chance of seeing variance between conditions given that such a high proportion of participants in each gave a guilty verdict. They also found that in some cases jurors took the opposite meaning to what was intended. No studies have examined how jurors respond to being told forensic techniques are "subjective", but when there are no set standards or criteria for making a decision, and when there are no clear links between results and methods of reporting, jurors are vulnerable to misunderstanding the strength of identification evidence.

iv. Lack of Validation Studies

With the exception of DNA, there has been no systematic verification of any field of forensic pattern recognition (Dror & Cole, 2010). The National Academy of Sciences report concluded that ‘in most forensic science disciplines, no studies have been conducted of large populations to establish the uniqueness of marks or features’ (National Academy of Sciences, 2009). This should leave testimony incapacitated: identification evidence, theoretically, depends upon the ability to say how (un)common features found in trace evidence are within a population (Hardy, 2007). If no studies have examined the proportion of individuals in possession of a characteristic (in isolation or in combination with other features), then an examiner has no basis upon which to declare whether evidence displaying that feature is likely to come from a given suspect. Most forensic domains can point to large amounts of literature about how their trace evidence of interest is detected and classified, but the research on source attribution is lacking (Thompson & Cole, 2007). This criticism applies even to fingerprints, where no data exists that estimates how common or how rare certain patterns are (Zabell, 2005), so analysts are unable to provide random match probabilities (Thompson, Ford, et al., 2003), and there has been no systematic measurement to support the notion that a given configuration of matching points is consistent (Mnookin, 2003). Having validation studies would be the best basis for proving accurate assessments of probative value (Edmond & Roberts, 2011).

v. Error Rates

Error has different meanings between a research laboratory and a courtroom. Budowle et al. argue: "An error rate in the context of a scientific discussion is defined as a continuous, repeatable, consistent action that yields a predictable level of false positive or false negative results in casework" (2009, p. 801). As every forensic science will have a nonzero probability of error, expert opinions need to have proficiency testing (of Type 1 and 2 errors; Koehler, 2016) and validation studies so that they reflect some level of uncertainty (Martire, Kemp, Sayle, & Newell, 2014; National Academy of Science, 2009; Thornton and Peterson, 2007). Formed based on the frequency of a trait within a population and also on the accuracy of the comparison method used to examine it, error rates, when combined with information about the likelihood of a failure to exclude two samples, would be an appropriate and meaningful way to communicate the strength of forensic evidence. Without error rate information, judges will be unable to evaluate reliability, and jurors cannot reliably estimate probative value (Koehler, 2017). As Thompson, Taroni and Aitken point out (2003), ignoring the possibility of a false positive could lead to considerable errors being made by factfinders, particularly when the other evidence in a case is weak, as more confidence is placed in identification evidence than it deserves.

There is increasing interest in using probability frameworks for evaluating and estimating opinion evidence (Berger, Champod, Curran, Dawid, & Kloosterman, 2011). In the UK, the Association of Forensic Science Providers (2009) has issued a set of numerical and verbal expression standards, based on a likelihood ratio, with the idea that this would help standardise reporting and assist jurors in weighing evidence. However, as Martire et al. (2013) point out, the two probabilities (reflecting the ratio of likelihoods of the evidence being consistent with the prosecution or defence's case) do not leave room for estimation of an error rate, showing that even among leading bodies of forensic examiners aiming for reform, recommendations from numerous commentators are still ignored.

Koehler (2017) has recently highlighted the necessity of error rates looking at both Type 1 and 2 errors. He argues that on the occasions when examiners do take proficiency tests they are "so easy, unrealistic, and otherwise unlike casework, that even the test manufacturers have said that error rates on these 'proficiency tests' should not be used to estimate casework error rates" (p.2). Studies specifically designed to measure both types of error rates will help experts, legal professionals, and juries to be able to more accurately assess the reliability, and therefore the probative weight, of forensic evidence.

vi. "Examiner Bias": confirmation and expectancy bias

Biases affect all decision makers, and as such, all examiners (Thompson, 2008). The idea that perceivers are affected by their internal desires and that expectations will influence observations was even noted by Francis Bacon in 1620 (see Bacon, 1900). In normal situations heuristics and biases operate effectively and efficiently, and are better than Bayesian reasoning for making judgements and decisions (Gigerenzer & Selten, 2002). In the social world, a biased person can do "better" than one relying on rational judgement (Cosmides & Tooby, 1994). But for a forensic examiner, bias can have tragic consequences.

Cognitive psychology has identified confirmation biases which will affect how individuals search, perceive, and judge evidence, so that if an individual expects that evidence will point in a certain direction, they are then more likely to seek out evidence that supports their initial beliefs, and interpret ambiguous evidence in the direction of that theory (Kassin et al., 2013; O'Brien, 2009; Plous, 1993), something common in identification analysis. Evidence that does not support an expected (or desired) hypothesis is ignored or given less weight and attention, and alternate explanations are not considered (Nickerson, 1998). A large amount of literature establishes that expectations can influence attention, memory and interpretation of subsequent information (Gilovich, 1991; Roese & Sherman, 2007; Stangor & McMillan, 1992), and that this effect can occur in a vast range of situations including taste preference (Lee, Frederick, & Ariely, 2006), judgements about humour in cartoons (Wilson, Lisle, Kraft, & Wetzel, 1989), and even the ratings of the taste of turkey (Makens, 1965). The biasing effect is more influential when the stimuli being perceived are ambiguous (Balcetis & Dunning, 2006; Trope, 1986). People are often oblivious to how their expectations and hopes have biased their perception (Ehrlinger, Gilovich, & Ross, 2005; Pronin, Gilovich, & Ross, 2004; Pronin, Lin, & Ross, 2002). These phenomena are well established in cognitive psychology literature, and forensic science is one of the few scientific fields to not benefit from awareness of effects of observation and the impact of potentially biased information (Risinger et al., 2002).

Despite the use of sophisticated analyses, identification evidence relies on a human examiner to make the ultimate judgement (Cole, 2002). The ubiquitous effect of confirmation bias means that forensic experts are susceptible to *examiner bias:* “when the results of a forensic examination are distorted by the context and mental state of the forensic scientists, to include subconscious expectations and desires” (Cooley & Turvey, 2007, p. 62). Different cognitive psychologists use *context effects, observer effects*, and *expectancy effects* to refer to the same experiences (Neisser, 1976; Risinger et al., 2002; Rosenthal, 1966), although Cooley and Turvey (2007) argue that they are describing the same phenomenon. Although they consider themselves immune (Kassin et al., 2013), examiners can be influenced by both explicit and implicit cues of what may be observed in evidence so that they become more likely to perceive a pattern in the expected way (Risinger et al., 2002). Years of experience in the criminal investigation field do not protect against non-conscious influences (Ask et al., 2011).

Cole (2005) has argued that an expectation bias may be behind errors in fingerprint identification. Drawing on a study by Risinger, Saks, Thompson, and Rosenthal (2002), he suggests that merely doing the act of searching for similarities (as opposed to differences) leads an examiner to be more likely to see common features within fingerprint patterns. Confirmation bias, however, may be present from the beginning of an examination. As it is rare for examiners to receive more than one questioned sample for comparison, they can intuit that the source of the sample is already considered by the police to be a suspect (Cole, 2005). Examiner bias is a concern for inter-rater reliability across experts, as well as intra-rater reliability over time (Kassin et al., 2013). In a within-subjects study of latent print examiners, 89.1% of individualisation decisions were repeated, and 90.1% of exclusions, with most of the variability occurring when cases were borderline (Ulery, Hicklin, Buscaglia, & Roberts, 2011). In an examination of exposed cases of fingerprint misidentification, 18% of cases had been corroborated by an expert for the defence: bias was able to overpower both the actual evidence and any potential incentive to disagree with the original interpretation (Cole, 2005). In a recent study looking at criminal stereotypes, participants were given information that is typically available to forensic examiners, such as race, gender, and the type of crime (Smalarz, Madon, Yang, Guyll, & Buck, 2016). Participants were more likely to find "matches" when the crime fit the stereotyped suspect, even if the prints were not actually a match.

Little research has been done to investigate the ways cognitive biases can affect forensic examiners (Edwards, 2009), despite the NAS report strongly recommending it (National Academy of Sciences, 2009). In one notable study, O'Brien (2009) found that mock investigators asked to generate a hypothesis early on in an investigation were more likely to show confirmation bias, and more likely to agree to further enquiries focused on that suspect. They also showed shifting opinions about determining guilt, so that the investigation was more likely to support an initial suspicion.

The mental image of the scientist is "a man in a white coat working in a laboratory, approaching his task with cold neutrality, and dedicated only to the pursuit of scientific truth" (*R v Ward* (1993) 96 Cr.App.R. 1 at 52). The idea of a scientists as a disciplined, disinterested observer only reporting the truth is as pernicious as it is popular. Identification evidence presents several psychological concerns, as it concerns complex judgement and decision-making processes – by the examiner conducting the analysis, and then by jurors who must correctly interpret and assess the evidence in a trial. Cooley and Turvey (2007) point out that forensic sciences ignore the impact of observer effects while simultaneously inflicting cultural pressures on examiners, leading to high ambiguity and expectation. This prevents the development of standards and measures that could assist in protecting examiners from bias, and which could also provide valuable information from validity studies and proficiency tests that could be used to provide error rates.

This list has presented the six main criticisms targeted at forensic identification evidence. Whichever gate-keeping measures are present in a trial need to make sure that jurors are aware of both the criticisms and the influence they could have on identification evidence. Kassin, Dror, and Kukucka (2013) have recommended that legal decision makers be informed about both the procedures leading to identification decisions, and the information that was accompanying the materials to be tested, including whether the examiner may have been exposed to other inculpatory evidence. To the author's knowledge, these criticisms have been presented to participants only once previously (Eastwood & Caldwell, 2015, which is discussed in the next chapter).

Some fields are making progress in establishing standards or procedures to make evidence more reliable (for example, Haber and Haber (2008) on validating fingerprint comparison, with Cole (2008) and Mnookin (2008) as a response; and Peterson et al. (2003) for a discussion of blind DNA proficiency testing). However, as a whole, in Australia and internationally, forensic identification experts testifying in court remain incapable of, and typically resistant to, accurately informing fact finders of how reliable any adduced evidence will be. Until these criticisms are addressed, and if forensic evidence is seen as being more reliable and trustworthy than it really is, then there is a danger that jurors will misuse the evidence. This could ultimately lead to inaccurate verdict decisions, and at worst, mistaken convictions. The current best protection against this will be to find a way to educate jurors about forensic science methodology and analysis. But what do jurors know about forensic science currently?

1.4 Juror Knowledge About Forensic Evidence

Despite the large amount of jury decision making literature, there has been little traction in developing theoretical frameworks to illustrate the cognitive processes leading to a verdict choice. As it is a criminal offence to seek information from a juror in Australia and the US, very few studies have used real jurors (Goodman-Delahunty et al., 2008). Most information about juror perception of forensic evidence and experts comes from experimental studies and anecdotes from members of the judicial system.

In a rare, recent exception, Rhonda Wheate (2006) was able to interview jurors from the Australian Capital Territory, who had sat on two panels of the same trial. The defendant was accused of sexual assault and burglary, but the first jury were unable to reach a verdict in regard to any of the charges, and so the prosecution called for a second trial. DNA profiling testimony was presented; however, the first charge related to the issue of consent as the defendant confirmed a sexual encounter had taken place, thereby making the presence of his DNA profile at the crime scene unsurprising and not inculpatory. Jurors were interviewed regarding their expectations and experiences. Wheate noted that responses confirmed jurors' belief in the importance of forensic evidence: jurors "seemed to fixate on forensic details and what the forensic witnesses 'ought' to have been able to explain" (p. 141). Most of the respondents told the interviewer that the DNA was important or very important to the trial, even though it did not implicate the accused directly, and was considered more important that other evidence in the trial, which included an alibi, testimony from two medical doctors, testimony from the police, and other witnesses. As other forensic evidence was deemed "missing", jurors began conjecturing about why, including the accused being able to cover his "tracks", or blaming poor performance from the police and crime scene investigators. One juror reported that the panel had "felt very let down" (p. 138) by lack of conclusive forensic evidence, and some jurors who wanted to rely on the profiling evidence to guide their judgement of guilt, felt that they couldn't give a guilty verdict even if they thought the accused had committed the crime. Without DNA evidence, they would not say the case was proven beyond reasonable doubt.

As part of a series of studies looking at the perception of forensic evidence, Lieberman et al. (2008) asked mock jurors to rate the accuracy and persuasiveness of different DNA evidence. It was rated as the most accurate type of evidence with the average rating of 93.7% accuracy, and 92.6% and 89.5% persuasiveness if found on the victim or crime scene respectively. Fingerprint, hair and video tape evidence were also rated as very accurate (90.1%, 88.7%, and 88.1%). In contrast, eyewitness statements had the lowest ratings (62.8%), with suspect confessions (73.7%) and victim statements (66.8%) viewed as only slightly more accurate. A general category of "expert testimony by scientists" received an 81.5% accuracy rating. Garret and Mitchell (2016) took a slightly different approach, as they not only wished to know what jury-eligible laypersons thought about the reliability of forensic evidence, but also how lawyers viewed the evidence, and how those lawyers thought the average juror would view the evidence. When considering fingerprint evidence, jurors were consistent with the Lieberman et al. (2008) study, with 76.5% rating the evidence as "reliable" or "very reliable", and 95% saying that fingerprints were unique. In contrast, while most lawyers considered fingerprint evidence to be somewhat to very reliable, over 30% responded that it was somewhat to very *un*reliable, and only 53.4% said that fingerprints were unique. Lawyers thought that jurors would be much less sceptical: 55.3% said that jurors would consider the evidence "almost infallible", and 36.6% said they would have "considerable confidence".

Interestingly, slightly fewer laypersons said that DNA would be unique (91.57%), and even fewer lawyers considered it so (78.4%). When asked about what jurors would think, most lawyers (83%) indicated that jurors would think DNA is “almost infallible”, and almost all of the remaining said it would be regarded with “considerable confidence”. Although the differences between laypersons and lawyers is surprising, the numbers still indicate that jurors will place a lot of faith in forensic identification evidence. That only the two studies reported here were able to be located also highlights the need for more modern research.

The next section will explore what is known about juror decision making and expert evidence, and describe the three models most commonly used to interpret juror responses to forensic evidence.

1.5 Juror Decision Making

Two groups of models are widely invoked within the jury decision making literature: The Story Model and Cognitive Consistency Theories, and the Elaboration Likelihood and Heuristic-Systematic Models. Each has formed the basis for hypotheses examined within previous jury research. They will be briefly discussed, because while no single model provides a basis for predicting whether an individual (or group) will give a guilty or not guilty verdict, each has been used to interpret and explain the cognitive processes working towards that decision.

1.5.1 The Story Model and Cognitive Consistency

One of the most prominent and widely used models in the jury-decision making literature, the Story Model, argues that jurors will seek to organise and understand the evidence presented at trial by forming a narrative (or multiple narratives), with an emphasis on the cause and intent in the story, linked together by a conceptual schema (Pennington & Hastie, 1986, 1992, 1993). Stories are informed by jurors' previous experiences, ideas about human behaviour, and beliefs about the social world. Once formed the narrative is compared to the possible verdicts presented by a judge (Pennington & Hastie, 1986). When examining evidence, Pennington and Hastie propose that each piece is organised so that it is linked to another; pieces of evidence that do not fit in with a particular episodic schema are discarded, and "missing" elements are constructed if the juror cannot draw upon material from the trial

(1986). According to the theory, confidence in a particular story comes from comparing analogous events and experience, and evaluating alternate explanations (Pennington & Hastie, 1993). Kuhn, Weinstock and Flaton (1994) propose that up to a third of jurors adopt a single story and stick to it throughout a trial, compared to more rational jurors who consider and evaluate alternative stories. Individual differences in story construction lead to differences in verdicts; differences in world knowledge can also lead to different items of evidence being seen as more or less credible (Pennington & Hastie, 1986).

The Story Model has been used to examine how jurors reach verdicts in civil cases (Horowitz, Bordens, Victor, Bourgeois, & Forsterlee, 2001), and murder trials (Wiener, Richmond, Seib, Rauch, & Hackney, 2002), in investigations of how confidence is linked to arguments (Green, Green, & McCloy), and to improve juror understanding of judicial instructions regarding the death penalty (Smith & Haney, 2011). Podlas (2006b) has argued that the CSI Effect can be theoretically understood through the frame of Story Model, with pre-trial bias about forensic evidence influencing the way a narrative concerning a crime analysis and investigation is constructed, which in turn may influence the probative weight given to evidence, and then verdict choice.

Application of the model is widespread; nevertheless there are limitations with using it to interpret how jurors utilise a particular testimony or piece of evidence. The original study material provided minimal evidence: based on *Commonwealth of Massachusetts v. Johnson*, the defendant was charged with first-degree murder after a fight broke out between Johnson and the victim, Caldwell. The main dispute in the trial is what happened leading up the fight (i.e. whether Caldwell had threatened Johnson with a razor earlier in the night, whether Johnson had his knife out to protect himself, etc.). The way the mock jurors constructed the events leading up to the death would affect verdict choice. However, there is no scientific evidence other than the pathologist saying a man was stabbed, and the cross-examination only seems to be asking questions of the defendant and a friend. The number of participants in the original studies was also quite low: only 26 participants were interviewed. Follow up studies also feature varied design: Weinstock and Flaton (2004) have suggested that many of the studies demonstrating support for the Story Model may be seeing narrative-based decision making because of the low amount of stimulus materials they include, whereas studies with more evidence may lead to demonstrations of alternate decision making styles. Typically, the model has not been tested on evidence approaching the complexity of what would be seen at a trial (Horowitz et al., 2001). And as is most relevant to the studies presented in this thesis, the model hasn't been used to look at how jurors understand complex scientific evidence (Smith, Bull, & Holliday, 2011), which may "transcend" the influence of a typical narrative structure (Edmond, 2011a), although the model has been used to examine the differences in evidence ratings more generally.

While the model does not provide a good framework for understanding or predicting cognitive processing, it does emphasise the role individual differences play in decision making concerning verdicts, and many subsequent studies have interpreted findings through the Story Model lens. One of the main ideas is that the use of different attitudes, life experiences, and general knowledge can lead to different conclusions about the same evidence (Pennington & Hastie, 1986; Smith et al., 2011). However, the model cannot predict how differences will affect story construction, or how stories will affect evidence, and developers and successive researchers have rarely included measures of individual differences in any empirical test.

Simon, Snow, and Read (2004) are proponents of the Cognitive Consistency theories, which they argue can be used to explain jurors' perceptions of evidence when they have to make multiple decisions, especially when it might concern ambiguous evidence. The core concept in relation to jury trials is that decision makers do not like incoherent representations or constructs, and will seek to build coherent ones, by using a network of positive and negative relationships amongst variables (some things in a story go together in a schema, and others do not). Consistency theories are argued to extend the Story Model by focussing on the holistic approach to decision making, but go beyond juror decision making into other tasks.

In a series of experiments, the authors tested evaluations of evidence that were separate from (as vignettes) or connected to legal cases, and measured the effects of a participant's verdict choice on evidence evaluation (Simon et al., 2004). When viewing the vignettes, participants gave moderate ratings to the strength of the evidence. However, when the same information was presented as part of a legal case the evidence received a much greater distribution of ratings, connected to the verdicts participants had chosen: the evidence strength ratings were changed, so that they clearly supported one verdict above another. These bidirectional shifts occurred when the strength of the evidence was manipulated, when participants changed their minds about the guilt of the accused part way through the trial, and when participants were assigned to an initial verdict. The researchers argued that this occurred as an 'adaptive mechanism' so that participants could show confidence and reduce cognitive dissonance, and would also incorporate beliefs and background knowledge. If this effect is occurring in cases concerning contested forensic evidence, it suggests that jurors with a strong belief in the guilt of a suspect prior to hearing evidence may interpret forensic evidence as being more reliable, or may be more likely to disregard the testimony of an opposing expert so that they avoid experiencing cognitive dissonance.

1.5.2 Elaboration Likelihood and Heuristic-Systematic Models

In a contrast to the Story Model, the elaboration likelihood model (ELM; Cacioppo & Petty, 1983) and the heuristic-systematic models (HSM; Chaiken, 1980; Chaiken & Eagly, 1989) have been used to describe information processing within trials, and have undergone large amounts of empirical research within and outside of the jury decision making domain. Each model assumes that an individual wants to hold correct information, but will differ in their motivation and ability to expend cognitive effort in determining the quality of a message (McAuliff, Kovera, & Nunez, 2009). When presented with information that clashes with previously held beliefs, people will analyse that information more carefully than if the information was consistent , and will use as much energy as is needed to become confident in a judgement (Chaiken & Eagly, 1989; Chen & Chaiken, 1999), but this depends on motivation and ability.

In the central route of processing described by the ELM and the systematic processing of the HSM, cognition involves high effort and vigilant consideration of information, with only messages deemed to be of high-quality being persuasive (Petty, Cacioppo, & Schumann, 1983). If, however, motivation or ability to process a message is diminished, a listener or reader may resort to peripheral or heuristic processing, which will take less cognitive effort. Rather than examining the content of the message and assessing quality or how valid the arguments are, people may rely on mental short cuts, or heuristics, such as expertise, likeability or attractiveness (Chaiken & Maheswaran, 1994a) to make judgments of the quality of a message. When processing is shallow there can be an increased effect of biases and a reliance on heuristics (Chaiken & Trope, 1999). Information can be processed in either an objective or biased manner, depending on the prior attitudes and knowledge held by the individual (DeWitt, Richardson, & Warner, 1997).

Cooper, Bennett and Sukel (1996) have argued that in a courtroom, jurors should generally be motivated to closely examine evidence, but that their processing ability may sometimes be limited. When comprehension of a complex message is low, expertise may be more persuasive (Ratneshwar & Chaiken, 1991). DeWitt, Richardson, and Warner (1997) were the first to investigate ELM processing with novel scientific evidence, although they focussed on knowledge about brainwashing in cults or new religions, a topic about which some mock jurors were likely to have strong, misinformed opinions. In their mock trial, an expert for the plaintiff introduced evidence that did not pass the rules of general acceptance (*Frye v. United States*), nor was it falsifiable (*Daubert v. Merrell Dow Pharmaceuticals*). Participants also completed the Need for Cognition scale (NC, Cacioppo & Petty, 1982), which is discussed in the individual differences section below. When jurors who were not motivated to process centrally (indicated by low NC scores) were confronted with ambiguous scientific evidence they showed less processing of arguments, (as ELM would predict). However, jurors who were highly motivated (high NC) and had pre-existing biases used testimony from the plaintiff's expert to exacerbate their previous biases even though they had heard reliable opposing expert testimony. In contrast, one study found that that expert testimony concerning insanity did not affect mock jurors' attitudes and final decisions if prior to the trial they had strong beliefs about the insanity defence (Greenberg & Wursten, 1988). Jurors with stronger beliefs were less likely to use heuristic processing, showing high motivation.

Having high motivation, however, does not mean that arguments or evidence will be considered systematically: an individual must also have the ability to engage with the argument. In complex trials increased quantity of information can decrease jurors' self-reported ability to understand trial information, and decrease confidence (Heuer & Penrod, 1994), and may decrease recall of probative information (Horowitz, ForsterLee, & Brolly, 1996). When experts present complex information, jurors may consider it easier to align with the opinion of experts (Petty & Cacioppo, 1986), using the status as a heuristic. High source expertise can lead to a decrease in systematic processing, although this effect may be strongest when the information and argument is new (Kumkale, Albarracín, & Seignourel, 2010). Additionally, older beliefs about forensic science (such as evidence being reliable, and individualisation being possible) will influence new information, as past attitudes can lead to high confidence and less effortful processing of criticisms. A combination of presenting complex scientific information into a trial, and a juror's trust in the opinions and reliability of experts, may lead to less effortful processing of testimony (Cooper et al., 1996). Several researchers have used the models to explain cognitive load and the effects of complex expert (and opposing expert) testimony on jurors. Complexity and the expert's credentials have been found to affect mock juror decisions (Cooper et al., 1996; Greenberg & Wursten, 1988). As predicted by the heuristic model, higher complexity of expert testimony is directly related to the juror's reliance on an expert's credentials in reaching a decision, irrespective of the content of the expert's testimony. Recently, Jones and Kovera (2015) used the two dual processing models to explain how use of a demonstrative (a visual aid flow chart explaining how to achieve internal validity in an experimental study) could reduce cognitive load, and therefore reduce reliance on heuristic processing. It should be noted that this is simulation research, designed to look at the "how" of juror decision making (Bornstein et al., 2017), and further research is needed to examine differences in processing between mock and real jurors, and whether mock jurors spend the same or greater amounts of effort processing information.

The models above cannot be used to predict verdicts in criminal trials. Yet each has had a lasting impact on jury decision making research. The Story Model has emphasized the role of individual differences in evidence evaluation, and how these differences can lead to alternate verdicts being formed, but the lack of prescriptive abilities means the model has little to offer in terms of helping legal personnel understand the best way to educate jurors. The model can argue that jurors will organise information into narratives and how conflicting and "missing" information will be integrated, but it cannot describe how and why forensic evidence will be assessed differently to other evidence and how this affects verdict choice, nor suggest ways to test which individual differences will influence evidence evaluation.

Although the Cognitive Consistency Model extends the Story Model by providing explanations of global vs individual item assessment, it has had limited testing with jury decision making. The Elaboration-Likelihood and Heuristic-Systematic models have been particularly popular in interpreting juror responses to complex expert testimony, and more recently have been used to describe the methods by which opposing expert testimony and visual aids might encourage more accurate processing among jurors (Park & Feigenson, 2013), and whether jurors would be able to detect an internally invalid study (McAuliff et al., 2009). This makes the dual processing model useful for explaining the influence of experts' credentials and of the role of bias within decision making, but it remains difficult to predict or measure when an individual juror will use the central or peripheral pathways or engage with heuristics (although this is true for all uses of the model throughout applied psychology). All three models describe different aspects of cognitive processing concerning evidence. However, none provide a framework for predicting verdict choice or how contested expert evidence will be received. The mechanisms through which expert testimony affects judgement, and then effects verdict choice, remains inadequately understood (Krauss & Sales, 2001).

Jurors make the final analysis of the value of forensic evidence. Having faith that jurors have reached appropriate verdicts relies on the legal community having assurance that the appropriate probative value has been placed on any forensic evidence presented in a trial. As scientific advances grow further away from the domain of general knowledge, jurors are more likely to need the expert assistance to understand and evaluate evidence (Wheate, 2006). More research is needed that examines how jurors will respond to forensic evidence, and in particular, identification evidence with potentially trustworthy reputation. yet unproven validity. The following section describes the current research examining jury decision making and expert evidence.

1.6 Jurors and Unvalidated Science

Scientific evidence, with its aura of certainty, can help jurors justify a legal decision. Humans feel a need to achieve justice (Ghoshray, 2007), but placing an undue amount of probative value on forensic evidence could lead to inaccurate decisions being made. A recent survey of legal personnel found that non-forensic evidence and even police reports are becoming less important in guiding the strategies used by the prosecution, as forensic evidence becomes more important to securing convictions (Stevens, 2008). If forensic identification evidence is being used to direct the way cases are presented, then it is important to know whether jurors can evaluate unvalidated evidence successfully and are not overawed by the expert presenting it. Petty and Cacioppo (1986) argued that source expertise may lead to a decrease in processing: "if people anticipate that others will ask for their opinions, but the issue is one on which they have only moderate knowledge, they may feel that it is relatively risk-free to identify with an expert's opinion but that the arguments of a source whose credibility is suspect require scrutiny" (p. 207). This is at the heart of the danger of unvalidated forensic science: that jurors who are unequipped or incapable of evaluating forensic evidence will accept an expert’s opinion without carefully weighing the evidence. But is the concern justified?

Most research into the use of expert witnesses with juries has looked at types of social framework testimony, where general social science research results are used to assist jurors in their interpretation of specific facts within a trial (Hurwitz, Miron, & Johnson, 1992; Vidmar & Schuller, 1989). The most common use of empirical research into this type of evidence has looked at experts in eyewitness memory (Hurwitz et al., 1992), who typically present research on potential sources of errors and bias, and how memory works over time (e.g. Desmarais & Read, 2011b; Leippe, 1995). However, the best way to present aggregated social science to juries in order to help them assess the reliability and applicability of the research is an area of ongoing debate (Monahan, Walker, & Mitchell, 2009; Walker & Monahan, 1987), and because it is not representative of all expert evidence types, empirical research is needed regarding juror use of different types of expert scientific evidence. Replication of effects is also needed with the use of different stimulus materials: research has shown that jurors react differently to evidence when it is presented in different case types, and different levels of severity of offence (Kim, Barak, & Shelton, 2009).

Jurors report that statistical, technical and scientific evidence is challenging (Cecil, Hans, & Wiggins, 1990), although this is unsurprising as even forensic practitioners overestimate their abilities to understand and interpret likelihood ratios (De Keijser & Elffers, 2012). Studies examining scientific reasoning skills of laypersons show substantial difficulties in judging complex statistical or technical information. Without assistance or training laypersons are unable to recognize inadequate sample sizes (Fong & et al., 1986; Tversky & Kahneman, 1974), or sample bias (Hamill, Wilson, & Nisbett, 1980), and are unable to recognize the need for a comparison or control group (Mill, Gray, & Mandel, 1994). Studies of laypersons set in jury-trial contexts have shown that jurors have difficulty understanding complex statistical information (Faigman & Baglioni, 1988; Kahneman, Slovic, & Tversky, 1982) and are unwilling to give verdicts based only on statistical information (Niedermeier, Kerr, & Messé, 1999; Wells, 1992). Some studies report that jurors under-weigh statistical information when other evidence is available (Kaye & Koehler, 1991; Schklar & Diamond, 1999). However, Smith, Penrod, Otto and Park (1996) found that while mock jurors in their study under-weighed evidence overall, there was a large amount of variation between individual jurors in whether they over- or under-weighed evidence. Jurors may not be able to assign an accurate weight to data that is presented to them, with researchers arguing that jurors may be confused by the logic behind source probability and estimates of a feature occurring in a population (McQuiston-Surrett & Saks, 2009). In their study comparing forms of forensic testimony, McQuiston-Surrett and Saks (2009) found that participants confused the logic behind these two numbers, where a high number of matches in a population should have led to a lower probability the defendant was the source.

Individuals struggle with probabilities less than 1% (Camerer & Kunreuther, 1989; Halpern, Blackman, & Salzman, 1989), which is common in the presentation of forensic science (Koehler, 1997; Pyrek, 2007). In contrast, studies have found that when information is given using verbal expressions of probability, the interpretations of each expression can vary widely (Beyth-Marom, 1982; McQuiston-Surrett & Saks, 2009). In one experiment comparing different ways of presenting forensic evidence to jurors, telling jurors that the expert had found a "match" or that samples were "similar-in-all-microscopic-characteristics" lead to the highest probabilities given by the participants that the crime scene sample belonged to the suspect, compared to providing subjective numerical probabilities (McQuiston-Surrett & Saks, 2008). Recently, based on recommendations that jurors be presented with verbal equivalents of likelihood ratios, Martire et al. (2014) compared numerical and verbal expressions and found that numerical likelihood ratios were used by jurors in a way more in line with the intended message given by an expert, and that verbal expressions of low strength evidence may even decrease belief in the guilt of the defendant.

Some research has found support for juror reasoning skills. Fong et al. (1986) found that methodological training could improve the ability of jurors to give answers to everyday reasoning problems (e.g. drawing conclusions about a population based on data from a sample). In a review of 13 complex civil cases, with some involving scientific expert testimony, Lempert (1993) found that juries made some errors, but that these could mostly be traced back to problems with presentations by attorneys or jury instructions. In experimental studies, mock jurors were more likely to give guilty verdicts when finding 20% of the population shared an enzyme type with the assailant and the defendant than when the statistic was changed to 80% (Smith et al., 1996), showing appropriate understanding of the scientific evidence. Another study looking at how groups would judge the diagnostic value of a bullet match found that mock jurors deliberating in groups were able to distinguish between statistical evidence correctly, but that this was due to a subset who were confident in statistics (Kaasa, Peterson, Morris, & Thompson, 2007). Goodman-Delahunty and Hewson Goodman-Delahunty and Hewson (2010) found that while initial layperson knowledge was low, expert evidence substantially increased knowledge about DNA, and mock jurors were able to apply their learning to novel scenarios. Those with less formal education had lower pre- and post-presentation DNA knowledge scores, and also showed less improvement.

To date, McAuliff and colleagues (McAuliff & Duckworth, 2010; McAuliff & Kovera, 2008; McAuliff et al., 2009) have conducted the only studies looking at multiple threats to the validity of forensic science evidence, beyond simply missing a control group. The trials presented to jurors concerned a hostile work environment case, where the featured expert's study looked at the effect of viewing sexualised advertisements on male behaviour. In the 2008 study, the ecological and internal validity of the research was manipulated; participants were either from a trucking company (and therefore similar to the work place of the plaintiff, for high external validity), or undergraduates, and the study had a control group (for high internal validity) or no control group. Additionally, the study showed general acceptance by having been published in a peer reviewed journal, or no acceptance. Participants also completed the Need for Cognition scale (Cacioppo & Petty, 1982). The ecological validity and general acceptance of the study made no significant difference in verdict choice. High NC jurors rated the study more highly when there was a control group, but low NC jurors were unaffected.

In the follow up 2009 study, the investigators manipulated internal validity via missing control group and external validity again, and also looked at whether mock jurors could detect a confound (different female confederates interviewed different groups), or demonstrated experimenter bias (where the experimenter knew which group had viewed the sexualised ads). The only threat detected by mock jurors was a missing control group, and decisions did not vary between high- and low-external validity studies. The researchers suggested that confounds and experimenter bias may be harder to understand than control groups, as they require a more sophisticated understanding of research design. In a previous study, mock jurors had used external validity as a heuristic when evaluating expert credibility (Kovera, McAuliff, & Hebert, 1999); however, the authors note that the sample in this study was made up of undergraduates, and that while there is little support for differences between students and public samples in juror decision making, it remains a topic of concern.

Finally, McAuliff and Duckworth (2010) performed a similar study but changed the case materials. Participants read through a trial transcript concerning child sexual assault allegations where a defense expert presented material about suggestibility of interviewing. The same internal and external valid/invalid conditions were used from previous studies. Participants were able to detect missing control groups, but not the confound or experimenter bias. Mock-jurors in the latter two conditions were also not attentive to publication status of the fictional researchers' work, as this made no difference. The researchers concluded that jurors could recognise the importance of control groups across a variety of trial types, sources of expert, evidence types, and operationalisations of missing control groups.

Overall, studies of juror's unaided recognition of methodological flaws shows that most threats will remain undetected or unappreciated, as jurors have repeatedly shown they are unable to evaluate sophisticated research design without assistance. If there are no challenges we can presume that jurors will remain uninformed about limitations of evidence (Edmond & Roberts, 2011). The person(s) providing the help would need to understand the limitations and problems with the particular area of forensic evidence, and be able to communicate these issues to the jury in a way that explains their importance, relevance, and how to consider them in light of the case (Edmond et al., 2016). In Australia's adversarial system, there are three commonly used methods of helping warn jurors about the dangers of unvalidated forensic evidence: judicial gatekeeping, the use of opposing experts, and cross examination.

1.6.1 Judicial Gatekeeping

Judges, like lawyers, are not trained or required to delve into issues concerning the methodological aspects of forensic science, and those who engage with the critical scientific literature concerning forensic evidence are the exception, rather than the rule. This means that most have only a “superficial” understanding of research issues (Edmond et al., 2016), and are ill equipped to see whether there is agreement in a scientific community, if it even exists (Edmond & Roberts, 2011). Nonetheless, surveyed judges were highly confident in their ability to act as gatekeepers (Dobbin et al., 2002; Gatowski et al., 2001), and their understanding of complex, technical evidence (Shuman, Whitaker, & Champagne, 1994).

In the US, judges are the “gatekeepers”, deciding whether evidence should be admitted based on relevance and reliability. Yet, they have shown difficulty in applying the Daubert standards of admissibility in assessing the scientific reliability and validity of research (Dobbin et al., 2002; Gatowski et al., 2001), with Gatowski and colleagues finding only 5% of surveyed judges understood falsifiability and 4% understood error rates. A survey of members of the Australian judiciary found that 70% of respondents had heard expert evidence they had not understood, and 20% reported having difficulties evaluating conflicting expert opinions (Freckelton, 1999). Judges have also shown difficulty in recognising flawed expert testimony (Kovera & McAuliff, 2000). A self-report survey found that judges are more likely to rely on qualifications and general acceptance in reviewing psychological testimony (Dahir et al., 2005), and Edmond (2010) has argued that in Australia, judges presume that questions about credibility, reliability and validity will be addressed adequately during cross examination, by using rebuttal witnesses, or through judicial directions at the end of a trial. However, when they do give instructions they are hard to understand and difficult to apply (Brewer, Harvey, & Semmler, 2004; Lieberman & Sales, 1999; Semmler & Brewer, 2002). This is also despite studies showing that admitting evidence to court increases its credibility according to juror’s ratings (Schweitzer & Saks, 2009). In *R v Tang* (2006, ‘*Tang*’), the presiding judge, Spiegelman

CJ, said: “The focus of attention must be on the words ‘specialised knowledge’, not on the introduction of an extraneous idea such as ‘reliability’”. Edmond and San Roque argue that “specialised knowledge” should be clearly defined, and should require an expert provides proof that a method is reliable by way of validation studies and proficiency tests (2014). The ongoing debate shows that neither in the US nor in Australia, does the judicial role of gatekeeper efficiently stop unreliable evidence from reaching jurors.

Judicial instructions, provided at the end of a trial, have been recommended by the Australian Law Reform Commission to help jurors understand the use and nature of probabilistic evidence. However, they have been shown to be unsuccessful at reducing conviction rates or reducing trust in DNA evidence (Dartnall & Goodman-Delahunty, 2006). The few experiments that have tested the effectiveness on judicial warnings have shown they are ineffective at helping jurors make more accurate decisions (Cutler et al., 1989; Dartnall & Goodman-Delahunty, 2006). Edmond has argued that real judges typically cannot give enough information to help jurors assess complex evidence (2011a). While the number of studies looking at judicial instructions is small, the surveys of judicial knowledge indicate that the judiciary may not be the best method of educating jurors about the dangers of unvalidated science.

1.6.2 Cross-Examination

Lawyers have little advantage over the judiciary in understanding forensic methodology, as legal schools are ineffective at teaching methodological reason or statistical skills (Lehman, Lempert, & Nisbett, 1988). Many lawyers must use general acceptance of the discipline to evaluate the strength of evidence (Kovera, Russano, & McAuliff, 2002). As mentioned above, Edmond has argued that judges rely on cross-examination to reveal flaws in identification evidence, but when the limited number of equivalent studies of those given to the judiciary were presented to attorneys, the results show that they may also struggle to understand complex methodology. Lawyers find scientific principles and language hard to understand (Cashman & Henning, 2012), and without knowledge of scientific methods and of the various forensic disciples they will be unable to accurately assess the validity of a particular expert's testimony. In interviews concerning confidence in presenting and arguing against DNA evidence, Cashman and Henning note that only a small number of lawyers said they knew "enough" about DNA evidence, with the vast majority indicating that statistics and random match probabilities were difficult to interpret and to explain to jurors. However, other studies have found that on self-evaluations attorneys rated themselves as better at assessing evidence than other attorneys would be (De Keijser & Elffers, 2012; Wechsler, Kehn, Wise, & Cramer, 2015). They also assumed that experts made errors (misrepresenting credentials, suppressing evidence or withholding it, just not their own experts (Wechsler et al., 2015). Evidently, not only are lawyers struggling to understand the complex forensic evidence being presented in the courtroom, but they also struggle to accurately assess their own levels of knowledge. Despite these difficulties, more studies have asked the question of whether cross-examination can educate jurors. However, that literature review will be included in Chapter 4 (section 4.3).

The first study was therefore designed to answer **RQ1**, by seeing if opposing experts were capable of educating jurors about the dangers associated with an area of unreliable evidence. However, before beginning a discussion of whether opposing expert testimony is effective, we must also establish the best way to measure that effectiveness.

1.6.3 Sense and Sensitivity (or Scepticism)

Eyewitness evidence is the leading cause of wrongful convictions (Innocence Project, 2017b). As jurors may be unaware of the factors influencing accuracy of testimony (Magnussen, Melinder, Stridbeck, & Raja, 2010), experts are frequently used to increase jurors' knowledge. This issue has considerable attention from psychological and legal researchers (for examples see Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Desmarais & Read, 2011b; Leippe & Eisenstadt, 2014). However, the methods used to study whether experts can increase juror discernment have varied, and not all the effects have been in the directions that commentators wanted. Researchers therefore needed to distinguish between the intentions and the results achieved.

Cutler et al. (1989) categorised three effects of expert testimony on juror decision making in eyewitness trials: no effect (where the jurors are not influenced), scepticism (where all evidence of that type is seen as suspicious), or sensitivity (where jurors are educated and capable of assessing factors that would influence the accuracy of an eyewitness). When jurors become sceptical it is inferred that they have understood the basic message of an expert but have not applied the specific relevant points to the particular case. Cutler, Dexter, and Penrod (1990) argued that "sensitization requires substantially more cognitive effort than scepticism: sensitization involves re-evaluating eyewitness evidence, whereas scepticism involves placing less belief in the accuracy of the identification. Thus, sensitization can be viewed as a consequence of increased elaboration, and scepticism as a consequence of decreased elaboration". This approach of valuing sensitivity has since been extended beyond eyewitness witness testimony research to studies looking at forensic evidence, including the effectiveness of multimedia presentations on mock juror understanding of DNA evidence (Hewson & Goodman-Delahunty, 2008), actuarial versus clinical testimony about dangerousness (Krauss & Sales, 2001), and child sexual abuse testimony (Buck, London, & Wright, 2011). However, this is not the best approach to studies looking at identification evidence as the fields currently stand. Rather, there is a need to focus on the change in jurors' approach to the evaluation of the evidence that is required. This focus is also commensurate with the courts' view of the role of the expert. Namely, in Australian courts, experts are seen as aides or teachers, engaged to educate and enlighten the jury regarding testimony that is beyond common sense (Ligertwood, 2015).

When the expert evidence presented in a trial concerns eyewitness evidence or is based on a single study or small amount of research (such as in Levett & Kovera, 2008, 2009), it is important for potential jurors to be able to distinguish between valid and invalid evidence, so that the probative weight can be adjusted accordingly. In these studies, sensitivity is valued because the factors influencing eyewitness evidence will vary between contexts, as will the design of single psychological studies. In contrast, anthropometric facial comparison evidence, like most types of identification evidence apart from DNA profiling, suffers from threats to validity across the entire field (Edmond, Biber, Kemp & Porter, 2009). Jurors also believe there is a very low risk for false positive errors being made by experts: Kohler (2017) found that jurors considered the risk of a false positive error occurring for a range of five identification sciences to be less than one in a million. Therefore, the goal then is to see if opposing experts are effective at helping jurors to more accurately assess the evidence by directly measuring the perception of the facial comparison as a field of evidence. Testing for scepticism will be the first step of establishing the effectiveness of methods educating jurors.

It is clear that jurors consider identification evidence credible and trustworthy, but research has yet to establish the amount that this differs between jurors. The influence of personal characteristics on the perception of unreliable science has also been under researched. As mentioned in the introduction, research into individual differences so far has found mixed support for direct effects on verdicts.

1.6.4 Individual Differences

Researchers investigating the potential for individual differences in predicting jurors' verdicts have noted inconsistent findings (Ask et al., 2010; Smith & Bull, 2012), with some researchers suggesting attitudes and personality traits predict such a small amount of variance that in most cases they are, for practical purposes, useless (Ellsworth, 1993). Studies of actual jury panels found that less than 2% (Hastie, Penrod, & Pennington, 1983) and 10% (Mills & Bohannon, 1980; Moran & Comfort, 1982) of variance in verdicts could be explained by demographics such as education, political orientation, occupation, gender, age, and prior service. There may be specific cases where attitudes have had more influence such as death penalty cases, sexual assault, or racial crimes (Ellsworth, 1993; Garvey et al., 2004), but for most trials there is no support for a direct relationship between general demographic details and verdict. However, there are possible relationships when considering more proximal individual differences that are relevant to the processing of information.

Need for Cognition

Need for Cognition (NC) is the tendency to seek out and engage in mentally demanding activities (Cacioppo & Petty, 1982). High-NC individuals are more likely to engage in effortful thinking about information than low-NC individuals, who prefer to use less cognitively taxing ways of reaching decisions, e.g. using heuristics, taking the opinions of others, etc. (Cacioppo, Petty, Feinstein, & Jarvis, 1996). Due to the understanding that evaluating scientific evidence will be cognitively effortful, it is not surprising that relationships between NC and contested expert evidence have been investigated previously. In trials with complex scientific evidence, those higher in NC may expend more effort evaluating the methodological design behind an expert's evidence, and potentially have the reasoning skills to assess it more accurately. Leippe, Eisenstadt, Rauch, and Seib (2004) found that when evaluating eyewitness evidence NC scores interacted with case strength: high NC mock jurors were able to find flaws in the prosecution even when the defence had a weaker case, as more thorough evaluation found reason for reasonable doubt. However, moderate NC participants were more likely to give guilty verdicts than low NC when the prosecution had a strong case, suggesting a more complicated, curvilinear relationship. McAuliff and Kovera (2008), in their study mentioned above, found that when an expert's study was internally valid (rather than missing a control group), high-NC jurors gave higher ratings to the research quality of the expert's evidence and were more likely to find the defendant liable, but low-NC jurors did not. However, while they were statistically significant, NC scores accounted for less than 5% of variance.

The studies of NC on jury decision indicate that it may have an effect on the way jurors consider scientific evidence. However, while the previous results are in need of replication and extension, the studies in this thesis looked at alternative individual differences. This is because the direct relationship appears to be quite weak, and little research has looked at whether NC can be improved through training, which makes it of limited use in regards to developing educational methods for jurors.

Meta-analyses of Individual Differences

Devine and Caughlin (2014) recently conducted a meta-analysis of studies looking at personal attributes in jurors and defendants to see which factors were involved in guilt judgements. Two hundred and seventy-two published and unpublished studies were examined, making it the most recent, and largest, examination of the role of juror individual differences. The characteristics were chosen based on Devine's (2012) "Director's Cut Model" multi-level theory of juror and jury decision making which is based on the Story Model. The Director's Cut Model proposes that jurors build mental models: situational specific representations used to help make sense and reach decisions, e.g. a juror might use a mental model of a scientist handling evidence and make judgements of the chance the testing was done objectively. Initial representations are determined by the juror and defendant characteristics, and any information given prior to evidence presentation in a trial, including media pre-trial publicity, and opening statements. These interact with trial elements to create stories. Six characteristics of jurors were focussed on in the analysis, and all had weak to moderate overall relationships with verdicts: education level and prior experience as a juror did not differ significantly from zero (although the authors note that very few studies had hypothesised that education level would have any effect), while authoritarianism and trust in the legal system had effects large enough for practical significance, although the relationship between authoritarianism and verdict varied depending on the particular measure used, and on the case type. Under close inspection, the effect of gender (with female participants slightly more likely to convict) was due to case specific variables, with 44% of the effect coming from adult or child sexual assault cases. Juror gender and NC also had weak overall effects ($\bar{r}$ = .08 and -.07, respectively). The researchers also note that the latter effect is unsurprising as while the studies suggest a weak association between NC and verdict, with low-NC participants more likely to convict, NC was rarely predicted to correlate with verdict, although there is support for interactions with other variables related to the complexity of a trial.

Overall, Devine and Caughlin noted that the predictive value of the characteristics looked at was small, and would only make a difference in a statistically small number of cases where a jury panel was already close. They recommended that more studies should use community members or venire persons (potential jurors who appear in court for selection), and there was a need to extend research on individual difference measures. Particularly, they recommended further research into psychological measures that would reflect attitudes towards the legal system, rather than just observable measures such as race and SES.

The final characteristic of individual jurors to be discussed is the CSI Effect, which has dominated a considerable part of the dialogue concerning juror decision making, both in the media and in legal commentary.

1.6.5 CSI Effect

Two of the most archetypal television shows to feature forensic science, CSI: Crime Scene Investigation and Waking the Dead, started airing in 2000 (Cole, 2015), and the last two decades have seen the genre flourish, with ever more shows focusing on forensic evidence and how it can be used by the police to solve crime (Cavender & Deutsch, 2007; Nolan, 2007). At one point CSI was one of the most watched television shows in the world (Michael, 2007), with several knock-offs and spinoffs also becoming popular around the globe (Cole & Dioso-Villa, 2009). Yet the shows are accused of showing a distorted view of science, and surveys of those working in the American legal system have shown concern that juries may be heavily influenced by the effect (Cole & Dioso-Villa, 2009; Houck, 2006; Hughes & Magers, 2007; Maricopa County Attorney's Office, 2005). A survey of 290 prosecutors, public defenders, and judges found that 79% could cite specific cases where they thought jurors had been influenced by forensic TV (Robbers, 2008). Over 85% of responders also felt their jobs had changed, with more time needing to be devoted to discussing forensic evidence and the reality of TV shows, and 70% felt jurors had unrealistic expectations of evidence (with any lack of forensics being seen as “sloppy” police work). The greatest concern was the number of respondents who recalled examples of jurors discounting eyewitness or victim testimony.

Many jurors will have had little experience with or exposure to the judicial system outside of the media or entertainment (Hallgrimsdottir, Phillips, & Benoit, 2006), and falsified, or even glamorized, depictions of forensic science may inform and influence jurors’ perceptions before they even start a trial. Researchers (Smith & Bull, 2012), legal personnel (Stevens, 2008), and police officers and investigators (Stinson, Patry, & Smith, 2007), have voiced the concern that this will affect jurors’ assessments of evidence and testimony.

CSI main character Grissom says in one episode "Evidence only knows one thing: the truth. It is what it is” (Antonio & Lipsitz, 2000). CSI-type shows feature an exaggerated and stylised version of forensic sciences (Cole, 2015). Evidence is presented as certain and objective, unlike the conflicting and inconclusive forensic evidence real jurors will see in court (DiFonzo & Stern, 2007), and examiners are methodological, unbiased, and most importantly, never identify the wrong person (Cooley, 2004); a match in trace evidence is synonymous with certainty (Kruse, 2010), and typically reinforces the idea that identification evidence can be used to individualise a suspect (Saks & Koehler, 2008). Evidence is typically portrayed as being distinctly separate and removed from the examiner. In an examination of the first season, Kruse found only three moments when the characters discussed the need to "interpret" evidence; it was more likely that a member of the team would get a "hunch" and then go to find evidence to support it (2010). Forensic scientists are also depicted as being "brilliant", pioneering, and far more reliable and honest than police officers and lawyers (Kruse, 2010).

Within the show, new technologies for finding or processing evidence (including highly unrealistic ones) are praised. Depictions have created positive reputation for the experts and examiners who work in forensics (Hoffner & Buchanan, 2005). Harrington, in her analysis of the connections between CSI and American national identity, has argued that the show simultaneously affirms the ability of law enforcement to achieve justice and find the true perpetrators of crime, to "represent society in a quest for justice", and that "CSI follows a familiar pattern that mythologizes the investigators as servants of truth and protectors of society" (p. 8). The shows also depict other forms of evidence as unreliable, and witnesses as deceptive and untrustworthy (DiFonzo & Stern, 2007).

Research into the persuasiveness of fictional stories have found that messages and depictions may influence viewers because readers or viewers are less likely to engage in forming counterarguments or to scrutinize messages (Prentice, Gerrig, Chaiken, & Trope, 1999), which is made more difficult when immersed in a fictional story (Green & Brock, 2000). Previous research has found that fictional depictions of law enforcement can affect opinions of the legal system (Podlas, 2002). People need to be motivated *and* able to reject a message when it is delivered via entertainment (Gilbert, Krull, & Malone, 1990). There is evidence that the likelihood may differ between those with high- and low-NC: for example, Cao (2015) found that labelling videos as fictional, versus non-fiction, lead low-NC participants to be less likely to form arguments against messages shown in the same video. Research into pre-trial bias also suggests that mock jurors have good memory for related information obtained prior to a trial, but can also become highly confident that it was obtained *during* the trial (Ruva & Guenther, 2015), which is problematic if this extends to depictions of forensic analysis.

The “CSI Effect” has been given many different interpretations in the media, judicial commentary, and in the social science community. Podlas (2006) lists three common definitions, and Cole and Dioso-Vila (2006) identify six, all resulting in changes to the way forensic evidence is perceived due to watching CSI-type programs. Maeder and Corbett (2015) present the clearest definition of the most widely cited consequence of the effect, calling it “the perception commonly held by lawyers, judges, police officers, and even the general public that, due to the apparent availability of forensic evidence on crime television shows such as CSI, jurors may be either unwilling to convict in the absence of such evidence or overly reliant on it when it is presented” (p. 84).

Tyler (2006) has suggested that the effects on each juror can be mixed or contradictory, with exposure to CSI type shows raising expectations of evidence so high that there might be more acquittals when forensic evidence is not present. Alternatively, when scientific evidence is present, jurors may place so much trust in the expert and their work that the probative value and reliability of the evidence is overestimated. However, despite the considerable amount of discussion about the possible influence of the effect, little empirical evidence has been found to support its existence, which is unsurprising given the varied methodologies and approaches used by researchers. The following section will describe the main approaches taken to examining the CSI Effect so far.

Empirical Evidence of the CSI Effect

In the first published study of the effect, Podlas (2006b) asked undergraduate students to give a verdict concerning a rape case where the issue was that of consent, and there was no issue of whether intercourse occurred or not. There was no statistically significant difference between CSI viewers and non-viewers, with Podlas concluding there was no effect against prosecutors. However, Mancini (2011) points out that as scientific evidence was irrelevant to guilt, the study was more a test of whether mock jurors could recognise the inappropriateness of the analysis.

In two studies, both of over a thousand people called for jury duty in different counties, Shelton and colleagues (Kim, Barak, & Shelton, 2009; Shelton, Kim, & Barak, 2006) found very high expectations that forensic evidence would appear in court. In the earlier study, just under half the surveyed jurors expected forensic evidence to appear in every case, with around 74% expecting it in trials involving murder or sexual assault. However, watching CSI or CSI-type shows only marginally increased the levels of expectation.

In the 2009 study, Kim et al. searched for indirect effects using path analysis. Exposure to CSI-type shows did not affect juror's willingness to convict, but there was an indirect effect through expectations that scientific evidence would be presented. This did lower willingness to convict, but only when non-scientific evidence was presented. However, this effect was not found when eyewitness evidence was presented, as there was neither a direct nor indirect effect of CSI watching on willingness to convict. Those who watched CSI were also likely to watch other forensic-type shows, documentaries and crime news shows. When exposure was changed to include other forensic-science type media, the results changed such that there was an indirect and direct effect on the dependent variable, with higher exposure leading to a greater expectations of scientific evidence. This led to a *decrease* in willingness to convict when there was circumstantial or eyewitness evidence only, but a significant *increase* in willingness to convict overall, even after controlling for expectations. Some demographic characteristics affected responses also: raised expectations were not only due to watching CSI, but also age and education, with those who were old or had a lower education level having a higher willingness to convict. This reversal shows that studies looking at the CSI Effect must be extremely careful not to lump shows using forensic sciences together. The authors argued that the change could be due to some of the shows approaching the crimes and stories from the more "sympathetic" view of the victims. While the study had a large, jury eligible, sample, they only used short statements to describe cases rather than simulating a jury trial. The researchers however suggested there might be a "tech effect" throughout society that increased expectations about scientific evidence.

Schweitzer and Saks (2007) presented short transcripts of criminal trials to college students and asked them to rate their perceptions of the trial and evidence. More frequent viewing of forensic science type fiction, but not of a more general crime type, led to higher confidence in verdicts than non-viewers, and better self-reported understanding of tasks performed by forensic scientists, which in turn made viewers more sceptical of inappropriate or inconclusive evidence. Frequency of viewing was not important, with this effect being found in those who watched the shows as little as once per month. No effect was found in verdict choice, with the researchers arguing that the measures testing forensic science more directly were more sensitive to individual differences than verdict.

Maeder and Corbett (2015) have recently criticised studies of the CSI effect because of the focus on the frequency of watching the shows. Maeder and Corbett (2015) cite Potter's (1986) research into perceived realism of television shows, by arguing that how realistic a show is seen to be may moderate the effect of its depiction of events on social judgement. Instead, the researchers attempted to examine the perceived realism of CSI type programs with the idea that some individuals may see the show as being a true representation of forensic science, whereas others would view the shows purely for fun. Perceived realism of Crime TV was related to positive attitudes towards DNA and eyewitness testimony, whereas frequency of watching was not related to either. Perceived realism also led to more weight being given to the DNA evidence. No direct relationship between continuous perceived realism of crime television and guilt certainty was found, but there were indirect effects via attitudes toward DNA and eyewitness memory, and perceived influence of DNA. Contrary to the researchers' predictions, frequency of viewing was related to verdict certainty, with those who watched more crime television *less certain* of the guilt of the defendant, suggesting that frequency of crime show viewing could lead to scepticism towards evidence (as it was related to lower guilt certainty).

Researchers have also conducted more abstract studies of the effect by conducting interview or surveys rather than using simulated trials to evoke responses. Heavier consumers of crime television were more likely to have heard of the CSI effect, and were more likely to report that shows such as CSI help to educate the public about investigative and evidentiary procedures (Hayes & Levett, 2013), were more likely to report higher self-ratings of understanding of DNA evidence (Brewer & Ley, 2010), and gave higher ratings of the accuracy of DNA profiling (Brewer & Ley, 2010; Smith, Patry, & Stinson, 2007). While these studies did not look at the impact on simulated trials or verdicts, they do indicate that viewing CSI type shows influences perception of forensic evidence.

The CSI Effect has gained a large amount of media attention in a short span of time (Cole, 2015), and yet the few empirical studies examining the effect have found mixed results.

Robbers (2010) argues this is unsurprising given the different sampling methods, with some studies using students, some community respondents, and others attorneys. Designs of studies have also varied, along with different definitions of heavy or frequent viewing, making studies hard to compare. Additionally, the focus of many studies has been the decision to convict or acquit, rather than the potential impact on other variables of interest, such as perception of evidence (Robbers, 2008). The main result of the effect appears to be an increased expectation that science might appear in the courtroom, although this is cannot be solely attributed to watching the shows. While the fictional depictions of scientists are far from the reality of real forensic experts, CSI-type shows are not the only source of misinformation: wrongful ideas about scientific findings can also come from journalistic reporting that is unintentionally misleading (Howes, 2015). The most prominent and, arguably, important finding is that current jurors place a large amount of faith in forensic sciences, and are likely to expect that it will be present in a criminal trial. As said by one juror from Wheate's study: "I don't know what case could get through without and DNA evidence at all. We never had DNA in the old days, and I don't really know if I was around then, but we never had it. So how did they actually convict people in those times, without DNA?" (2006, p. 142).

Ellsworth has argued: "discovering and analysing an individual difference that *does* have some predictive value allows us to study the *mechanism*s by which the internal propensity affects the judgement of a legal case" (p. 44), and would go beyond the surface level "stereotyping" of legal practitioners and some social scientists (1993). Investigating characteristics that affect the evaluation of scientific evidence will help legal decision makers evaluate the most effective methods of educating jurors. By focusing on individual differences that would affect juror perception of unreliable expert evidence, which is **RQ4,** the studies conducted will extend research in this area.

1.7 Summary

This chapter has provided an overview of the existing jury-decision making literature as it applies to unreliable forensic science, and a brief introduction to the two most prominent theories of jury-decision making. The theories cannot be used as frameworks to predict verdict choice, but have each contributed to different focuses in previous studies. It has also described the current problems with forensic identification science as general fields to help readers understand the criticisms that have been made by legal commentators, and psychological research, and why this leads to unreliable evidence being presented in courts.

Previous research concerning ways to educate jurors about unreliable forensic evidence, and the corpus of CSI Effect studies and commentary, are both based upon concern that jurors will not weigh forensic evidence accurately. This can be due to a lack of ability but also to preconceptions about forensic science (and scientists). The justice system needs more psychological research aimed at understanding the best ways to educate jurors about forensic science and enabling them to more accurately assess forensic evidence, which is becoming ever more sophisticated and complex. And while identification sciences continue to suffer from a lack of validation studies and standardised practices, subjective methods of interpretation, and use insufficient methods to protect examiners from cultural pressures and examiner bias, there is a crucial need to know if laypersons are able to be educated about these dangers to reliability. Additionally, while research into individual differences is unlikely to be applied in *voir dire* jury selection in Australia, further investigation of individual differences will help researchers and the justice system to build and invest in more effective and educational gate-keeping methods or tools. Rather than asking participants about CSI viewing habits, the studies presented in this thesis will focus on characteristics that affect the ability to evaluate unreliable evidence, such as Argument Skill, Epistemology, and bias towards forensic evidence.

2 STUDY 1

Historically, courts in Australian jurisdictions have admitted the majority of the forensic evidence adduced by the State (Edmond, 2015b), preferring to let the value of the evidence be contested during a trial (Houck, 2015). Forensic science is rapidly expanding and evolving – and yet even long-standing fields such as fingerprinting have yet to propagate practices that would protect evidence from bias, or to design and conduct validation studies (Thompson & Cole, 2007). Jurors, as triers of fact, are asked to consider the probative weight of all evidence adduced, but can only do so accurately if they are aware of the limitations behind that evidence. As argued in Chapter 1, legal commentators and researchers have argued that cross-examination and judicial instructions on expert testimony may not be adequate methods of educating jurors. But Stephen Odgers, SC (2015), has argued in regards to cross-examination that this may be due to the defence lawyer "[lacking] the knowledge of resources to engage in a thorough testing of the evidence of the forensic scientist so as to point out its flaws and weaknesses – with the consequence that the evidence is both ruled admissible and is inadequately challenged in cross examination" (p. 150). These concerns are just as likely to apply to the judiciary, but not to experts from the same, or similar, field of forensic science.

Opposing, or rebuttal, testimony is currently regarded as an ineffective method of educating jurors about forensic science with researchers concluding it will have minimal or no effect on verdict choice. However, this chapter will provide a detailed review of previous studies to argue that opposing expert testimony has not yet undergone the rigorous testing required to justify its value as a method for countering misleading forensic science. The study designs and stimulus materials used by previous studies may have limited the ability of opposing expert testimony to show the potential for educating jurors about unreliable expert evidence.

Each juror will respond to evidence and expert testimony differently, based on attitudes towards their role as a juror, the legal system, science and forensic science. Evaluating evidence relies on weighing and comparing arguments, and judging which are based on valid or invalid claims. The goal in study one was to investigate whether individual difference variables relating to arguments, and ideas about knowledge, would influence the reception of forensic evidence.

The first experiment was therefore designed to be an exploration of whether opposing expert testimony could affect mock-jurors' verdicts in a criminal trial that featured unreliable evidence (**RQ1**), and to see whether verdicts would be affected directly or indirectly by two individual differences: argument skill and epistemology (**RQ4**).

2.1.1 Opposing Expert Testimony: wizards into teachers

Koppl and Cowan (2010) argued that the conflict between experts within a courtroom has the potential to reveal the common sense meaning behind technical ideas of forensic science to jurors. When only one expert testifies about scientific evidence, they dominate the discussion and have control over the amount of information provided: "when they do, they resemble wizards whose power is enhanced by an air of exotic mystery. Strongly opposed experts provide an incentive to behave less like wizards and more like teachers" (p. 254). Utilising an opposing expert means that fact finders are hearing from an informed source who is probably able to provide more detail about the scientific method and an analysis of the evidence featured in a trial.

Five key studies examining opposing expert testimony and contested forensic evidence have been conducted in the last decade. Each study has provided key insights into the benefits, and limitations, of using opposing expert testimony with jurors. However, the overall impression legal commentators have taken from this research is that opposing experts are an ineffective tool for assisting jurors. This is partially due to Cutler et al. (1989)'s sensitivity versus scepticism criterion (explained below) being unduly applied to all jury decision making research involving expert evidence, and the focus being put on change in verdict choice – when in fact the relevant outcome variable may be comprehension of the evidence. The small number of studies and the subsequently limited range of investigated trial and evidence types, means that this topic has not been investigated in a rigorous way.

Levett and Kovera (2008) used opposing experts in a mock case of suspected child abuse. An expert supporting the defence presented evidence concerning the effects of the suggestibility of witnesses, which was based on research that had varying levels of validity. In the research presented by the expert, a study had examined the susceptibility of children to suggestive questions. Three versions of the research were presented: either a valid condition which had a control group, a counterbalanced group (with the events and objects featured in the leading and non-leading questions balanced across children), or a group who were asked leading questions. The expert testified that, based on the research, leading questions led to inaccuracy from the child witness. The opposing expert either addressed the methodology, or did not. In the conditions where they did address the method, they explained the importance of control groups and counterbalancing. In the valid condition, the opposing expert admitted that the study had a valid design. Comparison of the participants' ratings of expert's credibility, trustworthiness, and research quality as well as verdict enabled the researchers to see if an opposing expert (appearing for the prosecution) was capable of increasing juror sensitivity to methodologically flawed science. Participants did not become sensitized to the unreliable evidence, but instead were made sceptical of the expert's findings: guilty verdicts increased, and ratings of the expert's credibility and trustworthiness decreased, regardless of the actual content. The researchers argued that opposing experts were not effective at increasing juror sensitivity to flawed science.

In a follow up study, Levett and Kovera (2009) proposed two potential mediators between opposing expert testimony and liability decisions in a trial concerning hostile work environment sexual harassment: perceived general acceptance of the research the expert was presenting, and the view of opposing experts as "hired guns". Each concept was conceived as a heuristic that would shape the view of an expert. In the mock trials the plaintiff's expert either presented valid or invalid sexual harassment research on men's behaviour after exposure to sexual material. Male participants watched advertisements that either focussed on scantily clad female actors, or on the attributes of the products. In the invalid condition there was no comparison group. In direct or cross examination, the opposing expert then evaluated the study and explained the methods used (and why they made the study valid or invalid). Participants rated evidence strength, and provided ratings of the plaintiff's expert's research, and the verdict.

Levett and Kovera found evidence for both effects occurring simultaneously: the interaction did not affect measures of the trial outcome, but there was an effect on evidence strength and ratings of the expert's research, which were lowest when the research was invalid and this was addressed by the opposing expert. In testing for scepticism, the researchers examined the presence or absence of the opposing expert, validity, and verdict outcome, and found that participants were more likely to find in favour of the plaintiff when there was no opposing expert, indicating scepticism. The presence of an opposing expert also affected ideas of general acceptance of the research, which then affected ratings of probability that the plaintiff had experienced sexual harassment, but this was not influenced by the validity of the research presented. The researchers again concluded that there was limited support for the use of opposing experts in an adversarial form (although suggested further exploration into the use of court appointed experts).

However, there are still areas that remain unexplored by the two studies. The initial study featured evidence of a sensitive nature, as jurors respond to child sexual assault trials in different ways compared to other crimes (for examples, see Cashmore & Trimboli, 2006; Quas, Bottoms, Haegerich, & Nysse-Carris, 2002; Sheahan, Pica, & Pozzulo, 2017). The 2008 study also lacked a condition without an expert (to show a change in verdict when the harassment research was adduced), and had the initial evidence presented by the defence. In the follow up study, the stimulus material was a civil trial. In both studies the scientific evidence was research from a single research study, where jurors need to understand sophisticated design details and their implications for reliability to be able to judge the probative weight of the evidence. Juror evaluations of expert testimony where the research is based on a single experimental study may differ substantially from a juror's opinion of testimony which is based on a field of science. The threats to validation that can apply to the design of a single study (such as missing control groups, experimenter bias) differ to those applying to forensic identification evidence, such as a lack of any validation studies, or lack of standardised training, which were mentioned in Chapter 1 (section 1.3). Importantly, neither study incorporated deliberation.

Goodman-Delahunty and Wakabayashi (2012) examined jury responses to adversarial expert testimony to see how conflicting opinions were used when reaching a verdict as a group. Panels of jurors watched versions of a trial where opposing experts testified about trace evidence of bomb residue. By examining the spoken word content of the deliberations, the researchers found that jurors were not confused by the conflicting explanations or arguments, but took the time to debate the evidence and theories given by the experts. Having adversarial experts increased juror understanding of the forensic evidence after deliberation by reducing evidential errors. Unlike the studies by Levett and Kovera (2008, 2009), the trial did not focus on evidentiary reliability, and the experts disagreed about the conclusions of the evidence rather than the method used to analyse it. There was also no condition with a single expert to measure the effect of a non-contested testimony. However, this study shows that opposing expert testimony has the potential to aid juror understanding without leading to either a sceptical rejection of forensic evidence, or the "white coat effect" (Vidmar, 2005). Jurors engaged with the material in their attempts to judge the probative weight of the evidence. Having an opposing expert added to their understanding, as was suggested by Koppl and Cowan (2010).

Jones and Kovera (2015) recently tested whether including a demonstrative (a text based flow chart designed to help mock jurors understand how to ensure high validity in an experimental trial) would assist in sensitizing jurors to valid or invalid scientific evidence more so than if an opposing expert appeared alone. The study featured undergraduate students, and a 2 (experimenter bias vs no bias) x 4 (no opposing expert, opposing expert who did address validity, opposing expert who did not, and opposing expert who addressed validity and also had a demonstrative) design. The initial expert, appearing for the defence, testified about the impact of stress on a witness's accuracy during identifications, which was based on a real study in which an interrogator used different techniques to evoke levels of stressfulness in a mock prisoner of war camp (Morgan et al., 2004). On the next day of the Morgan et al. study, participants were asked to identify the interrogator from a line-up. Morgan et al. (2004) found that those who had gone through a low stress interrogation were more accurate in their identification.

In the invalid condition of Jones and Kovera's study the experimenters providing the line-up were described as knowing who was in the low- and high-stress conditions, introducing experimenter bias. The opposing expert then addressed experimenter bias, but in the valid condition admitted during cross-examination that there were no internal threats to validity. Mock jurors who were given the demonstration were able to diagnose the validity of the study 71% of the time, whereas the other expert conditions had lower rates of success (49%-58%). Those who heard from an opposing expert rated the quality of the featured study more highly when it was valid rather than invalid, even without the demonstration. However, the sensitivity to validity did not affect verdict choice, with only those who had the demonstrative being less likely to give guilty verdicts. The authors did note that the conviction rates were low overall (32%), and the bias in the defence expert's evidence may not have been enough for participants to change verdicts. It is also unfortunate that the participants were all students, as it would have been interesting to see if there were differences between students and non-students.

Finally, Eastwood and Caldwell (2015) investigated the effectiveness of opposing expert testimony and judicial instructions at helping jurors to identify flaws in hair comparison evidence. Participants read through a sexual assault trial, based on the real life case of Calvin Lee Scott (Innocence Project, 2017a). During the trial, an expert for the prosecution presented testimony concerning a comparison of the defendant's hair and hair found at the crime scene, overstating the value of the match between the hairs. In the study, participants were sorted into five conditions: a baseline which had no prosecution expert or identification evidence, a condition with the prosecution expert but no defence expert or judicial instructions, a condition with defence expert, a condition with judicial instructions, and finally a condition with both. The defence expert and judge directly explained the limitations of hair comparison: that it is a subjective analysis, there are no accepted statistics on the likelihood of similarities being found, two hairs from different donors can look similar, that hairs are not unique. As well as indicating whether they would give a guilty or not guilty verdict, participants also rated the strength of the cases, and the credibility of the prosecution's expert (if seen).

The researchers found that judicial instructions had no effect on verdict choice, which is consistent with previous studies (Cutler et al., 1989; Dartnall & Goodman-Delahunty, 2006; McQuiston-Surrett & Saks, 2009), and there was similarly no difference in case strength or credibility ratings. However, participants who heard the defence expert testify about the limitations of hair comparison were less likely to give a guilty verdict, gave lower prosecution and higher defence strength ratings, and found the defence expert less credible. Eastwood and Caldwell suggested that the differences between the jurors who heard the judge's instructions may have been less influenced because mock-jurors trusted the qualifications and scientific expertise of the defence expert more than the experience of the judge. The researchers concluded that opposing experts were able to educate jurors about the flaws in the identification evidence.

It is possible that opposing expert testimony may help minimise errors in decision making involving scientific evidence. It has been labelled as an ineffective tool for assisting jurors, but it has not been investigated in a rigorous way. Research has typically used case materials where the opposing expert appears for the prosecution and where there is no effect of the initial defence expert witness (Devenport & Cutler, 2004; Levett & Kovera, 2008; Pezdek, Avila-mora, & Sperry, 2010), so that there was no "effect" for the opposing expert to work against. Additionally, in many of these studies the evidence in dispute has been psychological (Krauss & Sales, 2001; Scurich, Krauss, Reiser, Garcia, & Deer, 2015), or from eyewitnesses (Cutler et al., 1990; Devenport & Cutler, 2004; Pezdek et al., 2010). This is a problem because many additional types of evidence involve the testimony of experts and lead to wrongful convictions, and bias and attitudes towards different types of evidence, may produce substantial differences. Eyewitness evidence has been reported by mock jurors to be not as trustworthy as scientific evidence (Lieberman et al., 2008), and legal commentators point out that forensic psychological evidence is viewed differently to "hard" evidence (Wechsler et al., 2015), even if they argue that it should be held to similar admissibility standards (Bernstein, 1995).

The prosecution have access to more resources, scientists, and consultants, as well as being more likely to have access to specialised training and the ability to consult with other prosecutors, compared to defenders (Edmond, Cole, Cunliffe, & Roberts, 2014). Subsequently, when expert testimony has been presented in court it has commonly appeared for the prosecution (Edmond, 2013a; Risinger, 2000; Rozelle, 2007), which has a different position in regards to establishing burden of proof than when evidence is presented by the defence. The prosecution is often able to weave forensic evidence into an engaging and interesting narrative which is supported by other evidence, whereas the defence are more likely to have a less interesting narrative with less evidence introduced (Edmond, 2012). Using a fictional trial where forensic identification evidence is presented by the prosecution will provide more insight into juror decision making by investigating whether opposing expert testimony can be effective when used by the defence.

2.1.2 Anthropometric Facial Comparison

In the criminal justice system, photographic evidence has increasing prominence as technological advances make it easier to record and enhance images (Feigenson & Spiesel, 2009). Mobile phone use is ubiquitous, and extensive instillation of CCTV cameras is expected to continue worldwide (Norris, McCahill, & Wood, 2002). It is estimated that in 2016, the number of digital photographs taken could have been as high as 1.1 trillion (Lee, 2016). This increases the chance that criminal activities will be caught on camera, which, in turn, increases the rate at which this evidence will appear in criminal trials (Porter, 2013).

In Australia, the recent decision made in appeals for *R v Smith* (1999), involved the ability of police officers to give recognition evidence based on captured footage. During an appeal in the high court, the joint judgement declared that the officers "were in no better position to make a comparison between the appellant and the person in the photographs than the jurors" (*Smith v The Queen* 2001 at [11]). However, in 2003 morphological and anthropometric comparisons were used to identify a Pakistani asylum seeker who was claiming to be from Afghanistan in *SHJB v Minister for Immigration and Multicultural and Indigenous Affairs.* As reported by Edmond et al. (2009), the acceptance of this type of testimony by the Refugee Review Tribunal and the Federal Court of Australia encouraged the use of experts to analyse photographic material for identification and subsequently testify in court.

Anthropometric facial comparison is the assessment of measurements of facial features: this can be done through the qualitative analysis of features to determine similarities between two photos, identification of unique features (scars, tattoos), or by the metric comparison of proportional relationships (*R v Tang*). The testimonies of anthropometric facial comparison experts are regularly included in criminal trials as a form of expert opinion evidence: recent Australian cases include *Murdoch v The Queen*, *R v Tang,* and *R v Alrekabi*. In the UK, Bromby (2003) reports that between 500 and 600 expert witness reports concerning facial mapping are produced annually. The admission into evidence of facial comparison expert testimony, however, is contentious. In Australia, one of the leading commentators on evidence law, Gary Edmond, has said that given the current state of the field it should not be admitted into court (Edmond, 2013b).

Edmond et al. (2009) conducted an overview of the legal challenges to the admissibility of expert identification evidence from images in Australian courts. The main challenges regarding admissibility occurred in *R v Murdoch* and *R v Tang*. In *R v Murdoch* the prosecution used images of a man refuelling a vehicle that had been taken at a time and place close to the location of the murder of Peter Falconio and abduction of Joanne Lees. The images "were of such poor quality that the number plate, adjacent to where the person of interest was standing, could not be resolved" (Edmond et al., 2009, p. 346). Chief Justice Martin called the expert, Dr. Sutisno "an impressive witness… highly qualified and experienced in her field" (*Murdoch* at [95]), also saying that the "principles underlying the work of Dr. Sutisno can be readily understood" (*Murdoch* at [112]).

In *R v Tang,* Dr. Sutisno testified, again giving positive identification that Hien Puoc Tang was shown in video footage taken from a robbery of a convenience store. At this time, Dr. Sutisno told the court:

"*The results of the analysis lends support to positive identification. Given my experience in looking at faces and analysing them, matching them, etcetera and building them up as well, I formed the opinion that given that number of matches I'm of the opinion that they're of the same, one and the same."* (*R v Tang* [23])

During cross-examination, Sutisno also argued that there was "no subjectivity in her assessment" (*R v Tang* [78]), arguing that she had followed a strict protocol.

The appeal in *Tang* became the first time Australian appellate courts reviewed anthropometric facial comparison evidence (Edmond et al., 2009). Chief Justice Spigelman wrote that the 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” (*R v Tang* [140]). However, the expert evidence was not excluded, but rather Dr. Sutisno was allowed to describe similarities and differences between the images. In the appeal regarding *Murdoch,* the evidence was found to be capable of describing similarities, but not to give positive identifications. Subsequently, the testimonies of anthropometric facial comparison experts about similarities and differences are now a legally recognized form of specialised knowledge in Australia, although experts are unable to give an opinion on the ultimate issue.

Facial Comparison Evidence: the problems

Studies examining the accuracy of laypersons at identifying unfamiliar faces has shown that even with high quality images matching is error prone: one study found error rates of 30% when looking from video stills to a line-up of photos (Bruce et al., 1999). Henderson, Bruce and Burton (2001) used a staged bank raid captured on CCTV footage and asked community members to match actors to photo-spreads. Follow-up studies used high-quality photographs, video-footage, and disguising hairstyles. Matching ability was highly fallible throughout the series of trials, even when using forced choice: in the fourth study 24% were incorrect when choosing between just two images, and in the final study (which asked participants to judge if a single image showed the actor in the video, closely resembling the task most likely to be presented to experts or jurors) 45% of participants incorrectly said that photos of the same actor were of different people, and 27.5% matched two different actors. Davis and Valentine (2009) found that 17% of participants incorrectly said that an actor from a video was not present (in person or in a photo), even when the footage had been taken minutes before. In the same series of studies, when the time between video and identification increased, so did error rates in target present (44% incorrect) and absent (33%). Giving mock-jurors a warning that the culprit’s identity may have changed over time also lead to some participants becoming less conservative in choosing. Collectively, the studies show that lay persons are highly error prone when matching faces.

Recent work led by Alice O’Toole, and funded by the National Institute of Justice, is aimed at determining how to measure expertise in forensic facial comparisons (Norell et al., 2015; White, Phillips, Hahn, Hill, & O'Toole, 2015). White et al. (2015) conducted the first systematic analysis of facial examiners only two years ago, and they were able to show that examiners had superior recognition compared to laypersons (including when images were inverted, indicating more strategic comparison methods). However, this is just one study, based on only 27 volunteers from a Facial Identification Scientific Working Group (FISWG) conference held in Quantico.

Very little research has been conducted to prove the validity and reliability of the most common photo-matching methods: the forensic pattern recognition techniques commonly used have not been systematically verified (Dror & Cole, 2010; Porter, 2012), and experts are not required to undergo testing to prove aptitude or expertise (Edmond, 2013b). Gary Edmond has previously said that one of the main techniques, superimposition, is “better suited to pop videos than forensic purposes” (2011a, p. 145).

Images taken from CCTV, mobile phones, etc., are not required to meet the strict evidentiary standards of other evidence (Edmond et al., 2009). CCTV camera footage is typically of poor quality (Porter, 2009), and recording and compressing digital images leads to a decrease in fine detail and can increase distortion (Valentine, 2015). The positioning of the camera taking the photo (Mardia, Coombes, Kirkbride, Linney, & Bowie, 1996; Porter, 2009) and perspective may alter the appearance of distances within a photo (Edmond et al., 2009). Porter (2013) points out that most CCTV footage is placed to help security personnel monitor crowd behaviour, and does not take into consideration lighting, framing, height, or angle.

Experts are typically asked to look at images where targets are a considerable distance from a camera, and may be wearing disguises or working in shadow or with competing light sources (Davis, Valentine, & Davis, 2010). Two images of a person will rarely look exactly the same as there is a large amount of within person variability between two images, and the variability is still substantial when compared to differences between individuals (Jenkins, White, Van Montfort, & Burton, 2011). Comparisons are vulnerable to changes in physical appearance and disguise (Edmond, 2013b), and correlations between features on the face may be quite high, precluding individualization (Mardia et al., 1996). If identification is not based on highly individualised features such as scars or tattoos, then the lack of knowledge about the frequencies of features means that experts cannot come to strong conclusions about identity (Edmond et al., 2009).

It is common practice for experts to receive incriminating images with a single set of comparison images, and for the experts to be told of police suspicions before they examine the images (Edmond, 2013b). Ambiguity in the images and direct communication between the police and examiners makes experts vulnerable to the biases mentioned above, an effect which was seen by Charman, Greggory, and Carlucci (2009), when undergraduates were placed in the role of examiners and asked to judge between photos and facial composites. Those who were led to believe a particular member of the line-up had been selected by a fictional witness had increased chances of choosing that person and also had increased similarity ratings.

The probative value of photo-matching evidence is rarely known, and within the court the prosecution seldom has to establish that the evidence is reliable. This means the defence is placed in the position of having to find a way to call it into doubt (Edmond, 2013a). Only a small number of studies have examined how visual evidence influences jurors' decisions, and most fail to control for experimental confounds (Feigenson, 2010). Research on photographic evidence more generally has shown mixed findings of biasing effects: in studies of negligence cases photos of an accident victim lead to increases in the damage award, but not culpability (Oliver & Griffitt, 1976; Whalen & Blanchard, 1982); showing jurors colour or black-and-white photos of a murder victim doubled guilty verdicts compared to a control condition who were not shown photos (Douglas, Lyon, & Ogloff, 1997); and gruesome images of murder victims, compared to neutral images, lead to increases guilty verdicts (Bright & Goodman-Delahunty, 2006). Feigenson (2010), however, also cites two unpublished dissertations that found no effect on sentencing or verdict (Modin, 2006; Nemeth, 2002). Showing videos of a crime scene walk-through during a murder trial (Kassin & Garfield, 1991), or a re-enactment of an accident (Fishfader, Howells, Katz, & Teresi, 1996), did not show any effect on verdicts or liability, either. Kahan, Hoffman, and Braman (2009) found support for an effect of the race and political ideology on interpretation of edited dashboard camera footage, with black and liberal viewers more likely to say there was not enough danger posed to justify deadly force than white and conservative viewers. However, these are studies where photographic images or videos were shown to participants. Because the legal research community has not addressed the use of photographic matching (particularly from security camera footage), the probative weight and significance of introducing these comparisons into a criminal trial is unknown (Biber, 2007). However, the participants in this trial were not shown photographic images to avoid any biasing effects.

This study aims to investigate whether opposing expert testimony can help educate jurors about the potential flaws in facial comparison evidence by improving knowledge about the scientific concepts behind the evidence. However, merely examining verdict rates does not provide enough information about how jurors perceive forensic science or how this changes when the evidence is called into question. In a study also looking at opposing expert testimony, Levett and Kovera (2008) used a series of questions measuring perception of the reliability of the expert evidence, which will be discussed below.

2.1.3 Methodological Reliability

The presence of the opposing expert is not what should be influencing the shift in ratings of guilt. Rather it is how the opposing expert informs and cautions the jurors' perception of the forensic science that should be the key factor in producing a change. The jury decision-making literature documents many studies examining how experts are perceived by jurors: typically, they focus on the credibility of the expert (e.g. Brodsky, Griffin, & Cramer, 2010; Cramer, Harris, Fletcher, DeCoster, & Brodsky, 2011; Neal, Guadagno, Eno, & Brodsky, 2012), communication style (e.g. Bornstein, 2004; Kovera, Levy, Borgida, & Penrod, 1994), the importance of credentials (Hurwitz et al., 1992), or even whether the expert was being paid (Cooper & Neuhaus, 2000). However, these approaches are not suited to an investigation of whether opposing expert testimony is effective at helping jurors.

In their studies looking at opposing experts, Levett and Kovera (2008, 2009) used a series of questions asking participants to rate the validity of the research used by the initial expert featured in their trial. Study one also used those questions (slightly modified from the original to make them appropriate for anthropomorphic facial comparison) as a measure of how reliable and valid the anthropometric facial comparison evidence is seen to be, and defined the score as a rating of methodological reliability. This approach captured the perceived trustworthiness of the scientific results and evidence, and the validity of the conclusions presented from those techniques.

Opposing expert testimony, by providing a critique and commentary on the anthropometric facial comparison, should affect ratings of methodological reliability. In turn, as the forensic science is the most significant evidence linking the defendant to the crime, ratings of methodological reliability should affect verdicts. It was therefore predicted that there would be an indirect relationship between the presence of an opposing expert and verdict choice that would have influence by way of ratings of methodological reliability. Study 1 also investigated whether there could be variables moderating perceptions of this reliability, and if these variables have a direct or indirect effect on verdict choice.

2.2 Individual Differences

Ellsworth (1993), and Devine and Caughlin (2014), have argued for the need to examine individual juror characteristics beyond demographic variables, and to look for attitudes towards the legal system or evidence. Introducing competing experts into a trial means that jurors need to evaluate and resolve competing arguments and opinions. The work of developmental and cognitive psychology has identified the need for jurors to be proficient in argumentation, and to have a sophisticated understanding of what it means to make a knowledge claim (Kuhn, 2001; Kuhn et al., 1994; Weinstock, 1999). The aim of Study 1 was to investigate whether operationalised measures of argument skill and epistemology (an individual's theory of knowledge and the nature of what means to "know" something, Hofer & Pintrich, 1997) would explain variance in juror responses.

2.2.1 Epistemology

An individual's personal epistemology will influence the way they argue, weigh and judge evidence, view claims, and make decisions (Hofer, 2002), potentially having a drastic impact on how they behave as a juror (Kuhn, Cheney, & Weinstock, 2000; Weinstock, 2011; Weinstock & Cronin, 2003), particularly whether a person is motivated to seek alternative interpretations of evidence (Weinstock & Flaton, 2004). There are various definitions of epistemology and explanations for how it develops, but Kuhn and Weinstock argue that personal epistemology involves the "coordination of the subjective and objective dimensions of knowing" (2002, p. 123), with individuals moving from believing that a "knower" is separate from the knowledge and that an objective reality could be observed, to realising that knowledge is constructed and depends on the context. Given the subjective nature of identification experts' judgements (Thompson & Cole, 2007), whether jurors are already predisposed to see knowledge claims as subjective could have a large influence on their reception to opposing expert testimony.

A study of jurors' beliefs about whether it was possible to be completely certain of guilt or innocence has been found to be related to verdict choice (Kuhnet al., 1994). Accepting that knowledge claims have subjective aspects, as well as objective ones, means that knowledge can be evaluated, and arguments judged for their merit (Weinstock & Cronin, 2003). Given the subjective nature of identification experts' judgements (Thompson & Cole, 2007), whether jurors are already predisposed to see knowledge claims as subjective and how they balance competing claims may have a large influence on their reception to opposing expert testimony.

In the hierarchical system used by Kuhn and colleagues, absolutists ignore the subjective nature of claims, and believe there is an objective, external reality they can know with surety (Kuhn et al. 1994). For some individuals, the subjective nature of knowledge dominates. Multiplists will have difficulty finding a way to evaluate knowledge claims: because each knowledge claim is subjective, and has been selected by its holder, all claims are valid. In the final evolution the objective and subjective dimensions achieve equilibrium. Evaluativists believe: "(that) two people can both have legitimate positions...but one position can have more merit...to the extent that that position is better supported by argument and evidence" (Kuhn & Weinstock, 2002, p. 123).

Kuhn and Weinstock (2000) have developed a measure of epistemological sophistication based on research in the cognitive development domain which offers a glimpse at how individuals balance and judge conflicting claims (Kuhn & Weinstock, 2002). It was intended to give a score over three levels of epistemology: absolutist, multiplist, and evaluativist. The researchers' goal was to examine frequency of pattern of responses and what differences occurred across age and education level, but there was no support for a progression towards higher sophistication due to age or education. Just under half (48.5%) of participants showed a pure pattern, achieving the same level of sophistication across all domains, and there were 16 patterns found overall, showing a large variation in responses when they are divided by domains.

In this study it was necessary to alter the scoring method questionnaire so that the results could be used in a path analysis, where categorical variables can only be dichotomous. This is a deviation from the intended use of the questionnaire, as it has not been tested in this way before, and therefore should be seen as only an exploration of whether epistemological sophistication can influence jurors' decision making in regards to forensic evidence, rather than validation of the construct in this context.

2.2.2 Argument Skill

In addition to examining epistemology, Kuhn, Weinstock and colleagues have examined the role of argument skill in jury decision making (Kuhn et al., 1994; Weinstock, 2011), extending prior work looking at reasoning skills more generally. Previous studies have found that argument skills varied widely among individuals, with those achieving higher scores able to reflect on their own views, able to contemplate alternative theories based on conflicting evidence, and construct a framework of potential scenarios rather than sticking to one narrative, as did those with lower skill (Kuhn, 1991). Argument skill was measured as a combined score from aptitude shown in counterargument, the discounting of alternative verdicts, and justification of alternative verdicts. Scores using this measure are related to the proportion of evidence used when justifying a verdict and the ability to consider other verdicts (Weinstock & Flaton, 2004).

Higher argument skill is associated with greater evidence coverage (based on amount of pieces of testimony used when answering interview questions about a trial; Weinstock & Cronin, 2004), indicating that greater skill may lead to more comprehensive consideration of evidence and of alternative verdicts. In a case involving conflicting expert evidence, those with higher skill should be better able to evaluate unreliable scientific testimony, rather than simply siding with one expert over another. Kuhn, Weinstock and colleagues have presented a considerable amount of research using jury decision making tasks as a way of studying responses to complex, conflicting arguments. However the emphasis, and analysis, has focused on the study of arguments in an informal reasoning task rather than applying the findings to juror behaviour.

As an exploratory study, this will be the first to see if epistemological sophistication and argument skill can explain variance in the perception of unreliable forensic evidence. Participants with higher epistemological sophistication or higher argument skill may show more sensitivity to potential flaws in the anthropometric facial comparison evidence and therefore may give lower ratings of the validity of the science used by the anthropometric facial comparison expert, either after hearing opposing expert testimony (for those in that condition), or based on their own evaluation of the evidence after the brief cross-examination of the facial comparison expert. Higher argument skill and epistemology should lead to consideration of alternative explanations of the facial mapping evidence, and decreased certainty, reducing the likelihood the facial mapping evidence will push the prosecution's case beyond reasonable doubt. The final prediction was that there would be direct effects of epistemological sophistication and argument skill on verdict, and indirect effects through methodological reliability.

The aim of Study 1 was to test the proposed model, shown in Figure 1 below, to examine potential direct effects on ratings of methodological reliability and on verdict choice (with an indirect effect on guilt through methodological reliability) as well as direct and indirect effects of argument skill and epistemology on verdict choice. By educating participants about the methodology behind anthropometric facial comparison, there will be a negative effect of opposing expert presence on ratings of reliability. As participants who have better argumentation skills or higher epistemological sophistication will also be better able to recognise threats to reliability within the evidence, higher scores on the measures of argument skill and epistemology will also lead to lower ratings of reliability.

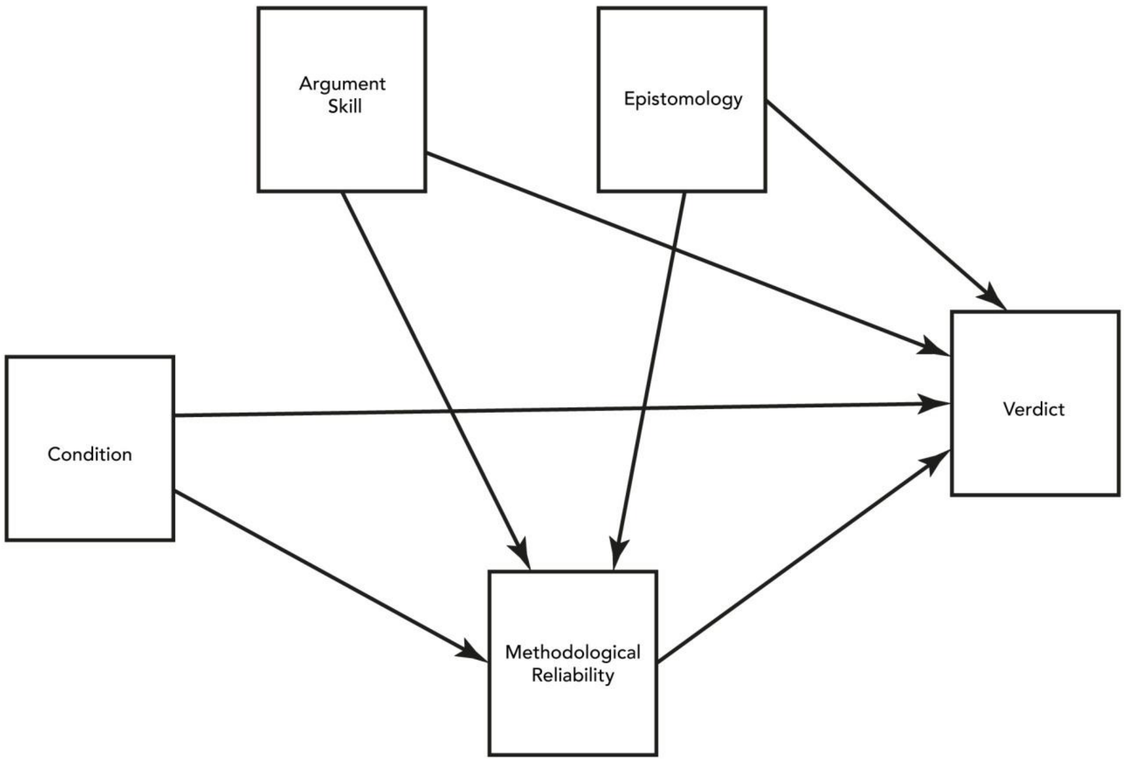


Figure 1. *Path model showing predicted relationships between Condition, Argument Skill, Epistemology, Methodological Reliability and Verdict*

2.3 Method

2.3.1 Participants and Design

The participants (N=124) who began the study were made up of university undergraduates who completed the tasks in return for course credit, and participants recruited from around the university and the wider community (results were combined from the beginning). Community members were recruited by posters placed around the University of Adelaide, emailing a list of people who had previously indicated they were interested in participating in psychological research at the university, and through snowball sampling.

Twelve participants were removed for being non-Australian citizens and thus not jury eligible (jurors are selected from electoral rolls, which limits selection to individuals 18 or over and citizens. Jurors must also not work for certain types of legal services or law enforcement). Eleven participants were removed for missing or short responses. Ten participants were removed for displaying poor comprehension of the material in the trial (i.e. their responses showed lack of attention to the stimulus material rather than difficulty coming to terms with conflicting evidence): two from the control condition, seven from the condition who experienced the prosecution expert testimony, and one from the condition who heard the prosecution expert testimony and opposing expert who appeared for the defence. Ninety one participant responses (M = 27.38, SD = 11.2) remained.

As a group, the participants were slightly more likely to be female (59.3%, with 3 jurors also identifying as transgender), with English as a first language (90.1%). Only three participants had prior jury experience.

2.3.2 Procedure

Participants, either individually or in groups of up to 6 people, were informed that they would be taking the role of a juror. They were also informed they would be answering a questionnaire at the end of the trial, and given guidelines on how to do so. After hearing instructions from a researcher and signing consent forms, each participant was seated at a desk where they were given a booklet which contained the demographic measures, trial transcript, and all the dependent variable questions, short answer questions, and the epistemology questionnaire. Copies of forms and questions are in Appendix A. Participants were told they should read through the trial carefully, but should do so only once, at a comfortable pace. All participants were also told to "consider only the evidence mentioned in the trial transcript". They were not able to take notes or to refer back to the trial transcript while answering the questions.

2.3.3 Materials

Trial Summary

A modified version of *R v Tang* was used as the basis for the trial transcript. *Tang*, a criminal trial heard in New South Wales, featured a defendant charged with an armed robbery committed by three people, two of whom were caught hours later with a bag containing some of the stolen goods. Fingerprints belonging to a third man were found on these items and were later matched with the defendant *Tang*. Anthropometric Facial Comparison was used by a single expert to compare images taken from surveillance footage from the store with images taken of the accused. In the study, all names, dates, locations, and some details such as the type of weapon were altered (e.g. a machete rather than a hunting rifle was used). The trial was chosen because it shows an example of a real case (which subsequently went to appeal) but featured unreliable evidence that jurors would reasonably be able to understand but find questionable.

Three versions of the trial were used: in the first version the jury was told by the judge that surveillance footage had not been able to be analysed, the second condition contained testimony from an expert for the prosecution discussing the facial comparison evidence, and the third condition contained testimony from an expert for the defence. Jury instructions were included in each trial version. The instructions included directions about the onus and standard of proof being beyond reasonable doubt in the Australian legal system. The trial transcripts can be found in Appendix A.

Anthropometric Facial Comparison Testimony

The testimony used in this study was formed from statements made by the expert who testified in *Tang*, edited to make them easier to understand given the short time-frame participants would have to read them. The expert describes the three main techniques she used to identify the accused: photo-superimposition (overlaying images), looking for unique identifiers (comparing distinctive features), and photo-anthropometry [1]. Photo-anthropometry uses measurements of the distance or angle between facial landmarks (Valentine, 2015), comparing proportional relationships between features. In conclusion, the expert tells the court that "given the number and quality of matches between facial features

[1] In *Tang,* the expert also testified about body-mapping evidence but the Court did not accept that a field existed during the appeal.

on the surveillance footage and pictures [she] was given of the accused, [she] formed the opinion that they show the same person".

Rebuttal Testimony

As Tang did not include a rebuttal expert, the defence expert's testimony was based on one from a second trial, *R v Alrekabi* (2007), which featured testimony from the same photo-anthropometric expert. The defence expert argues that the forensic evidence may have been biased, no second opinion was sought, and that the facial comparison expert was only given one set of comparison images, and was therefore aware that the person featured in them was already a suspect. The defence expert also argues that anthropometric facial comparison is not reliable as a field, as the claims made by the expert (that the images taken from the surveillance camera and the images provided by the police are a "match") are not backed up by empirical measurements – so there is no way of saying how common or uncommon the measurements used in the facial comparison would be in the population. This testimony can therefore be said to make specific arguments against the facial comparison expert. Compared to previous studies looking at opposing experts this is a large number of arguments to present to jurors, however it is an accurate representation of the field of facial mapping as it currently stands.

2.3.4 Methodological Reliability

Six of the statements used by Levett and Kovera (2008) were given to all the participants who saw the prosecution expert, rewritten to make them appropriate for the evidence featured here. Two were reverse scored. The questions were:

I. The methods used by the prosecution expert to identify the suspect were based on good scientific principles;
II. The evidence presented by the prosecution expert was not reliable (reversed);
III. The techniques used by the expert for the prosecution were scientifically appropriate;

IV. The findings of (the anthropometric facial comparison expert) can be used to reliably identify the third offender;

V. In my opinion, the procedure used by the prosecution expert to identify the suspect could protect against bias;

VI. The methods used by the prosecution expert cannot be used to declare a reliable match between images from surveillance footage and images of a suspect (reversed).

Participants responded on a Likert scale (1= strongly disagree, 7= strongly agree). The sum of the answers was used to give a final score out of 42. Versions of the statements have also been used in McAuliff et al. (2009), and Jones and Kovera (2015). Cronbach's α for the current study was .90, which is similar to the scores from previous studies (Levett & Kovera, 2008, α = .92, McAuliff, Kovera and Nunez, 2009, α = .87; Jones & Kovera, 2015, α = .89).

2.3.5 Argument Skill

The short answer questions were: "Why did you choose this verdict?", "What other factors went into your decision to choose that verdict?", "Was there anything in the trial that suggested this was not the proper verdict?", "Why didn't you choose the other verdict?", and "Other jurors might have chosen a different verdict. How might someone who chose the alternate verdict explain their choice?" Spaces were provided to write the response for each question.

Responses were scored using a modified version of coding scheme developed by Kuhn et al. (1994), which was also used in Weinstock (2011), which draws upon theory-evidence coordination. Evaluating one theory in contrast to another means that evidence must be evaluated and reflected on, rather than just integrated into a previously held narrative, or ignored. Higher skill is shown when an individual recognises evidence that does not fit their preferred theory, and then is capable of creating supporting arguments for an alternative theory, and finally being capable of acknowledging arguments discounting their preferred theory.

As there were only two possible verdict options in the stimulus materials, the coding scheme was simplified. A successful answer had to use one or more pieces of relevant evidence drawn from the testimony and either evaluate them or logically argue why they supported guilt or innocence. The first two questions did not prompt for a particular skill but allowed jurors to think about the chosen verdict and the evidence used to support it.

The author and an assistant coded all the short-answer question responses after practicing on a random selection of responses. Inter-rater agreement was calculated using the Kappa Measure of Agreement to account for any agreement due to chance. The kappa value was .79, p < .001, which is a good agreement (Mellis, Peat, Williams, & Xuan, 2001). Disagreements were resolved by consensus to create a final score. Participants could receive a score of 2 (successful), 1 (partially successful), or 0 (unsuccessful) for each of the skills examined, for a total score out of 6. For a successful *counterargument* participants needed to mention one or more pieces of evidence and explain how it might be used to argue against their chosen verdict. To successfully *discount* an alternative verdict participants needed to use the presented evidence to logically argue for their chosen verdict and against the other. Finally, to successfully *justify* an alternative verdict, participants needed to explain how another juror could choose other verdict. Partially successful scores were given if relevant evidence was mentioned and not explained. An unsuccessful score was given if no attempt was made, or if an illogical argument was presented. In some cases jurors already demonstrated the skill when answering a previous question (e.g. discussing evidence another juror would have considered proof of guilt while explaining why they selected innocent), and in this case would still be given a score for displaying that skill. Table 1 shows examples of successful answers to each question.

Table 1. *Argument Skill Questions, Skills and Examples of Successful Responses*

Answer	**Score**	**Explanation for score**
<u>Counterargument</u> **"Was there anything in the trial that suggested this was not the proper verdict?"**		
"I thought that the evidence presented regarding the machete was highly indicative of the accused's guilt. This is a fairly unusual object/potential weapon to have in your possession at the same time a robbery was committed with the same item. I also thought the 'stranger in the shadows' explanation of the third man was a bit lame and reminiscent of film noir rather than realistic evidence."	Successful (2)	The participant gave a not guilty verdict, and identifies two pieces of evidence and explains why they could be used to argue for the defendant's guilt.
"The owner of the store could not specifically identify the accused. The expert defence witness made a number of convincing arguments when discussing the case/bias of Dr. Hamilton".	Partially successful (1)	Participant mentions relevant evidence but does not explain how they affect guilt.
"There was a large amount of evidence against the man on trial"	Unsuccessful (0)	Participant makes no attempt to explain reasoning.
<u>Discounting</u> **"Why didn't you choose the other verdict?"**		
"Even if the forensic CCTV evidence was biased, the machete and fingerprints were still enough to convince me beyond	Successful (2)	Identifies evidence and reason for decision.

reasonable doubt. Also, we only have the defendant's word that there was another man with the accused, and no other evidence to support the claim."		
"Because I believed there was a chance of bias and because the methods used for the technical aspects of the facial recognition do not seem supported by evidence"	Partially successful (1)	Mentions bias and unvalidated methods used by expert but argument does not explain how this does not support the use of the facial comparison evidence as inculpatory evidence.
""Because the evidence provided by the prosecution was both stronger and more convincing."	Unsuccessful (0)	No details or evidence mentioned.
Justification "**Other jurors might have chosen a different verdict. How might someone who chose the alternative verdict explain their choice?"**		
"They could have seen [the facial comparison expert's] testimony as a reason for reasonable doubt and became sceptical at the idea of positively identifying someone through facial comparison/recognition. [The defendant's] story about the alley may also have left room for doubt and sounded plausible."	Successful (2)	Identifies two pieces of evidence, and acknowledges other mock juror's may consider them to lead to enough uncertainty that the prosecution cannot prove their argument beyond reasonable doubt.

“They might find him guilty based on the large amount of evidence presented against him e.g. cigarette preference, the weapon, the evidence presented from the face matching analysis.”	Partially successful (1)	Implies that case strength would be enough for another juror to give a guilty verdict but only lists elements of the trial rather than linking and presenting argument.
“That the facial recognition could not be used as concrete evidence”	Unsuccessful (0)	Mentions facial comparison evidence but does not present argument as to why it is unreliable.

2.3.6 Epistemology

Epistemology was measured using the scale developed by Kuhn, Cheney, and Weinstock (2000). The questionnaire has twelve scenarios, and asks whether only one of the fictional actors could be correct in their view (suggesting either an absolutist perspective), if both could be right (indicating a multiplist view), or if one could be more right than the other (indicating evaluativism). The response given indicates the coordination of the objective and subjective dimensions of knowing. An example is presented below:

Robin has one view of why criminals keep going back to crime.

Chris has a different view of why criminals keep going back to crime.

A) Can only one of their views be right, or could both be right?

- Only one can be right (GO TO NEXT QUESTION)
- Both can be right (GO TO B)

B) If both could be right, could one view be more "right" than the other?

- One could be more right.
- Neither is more "right" that the other.

The measure has been tested with 107 participants, with ages ranging from fifth graders to adults studying at college, MBAs and PhD candidates. The researchers also report good consistency with the longer, interview based Livia task (where participants must evaluate two conflicting historical accounts to demonstrate ability to reconcile conflicting arguments and certainty of a historical account (Leadbeater & Kuhn, 1989). Certainty in an individual's understanding of the Livia task (e.g. whether one account must be correct, or if both can have valid points) has previously been found to be related to mock juror verdict choice, and the amount of evidence used when describing reasons for verdict choice (Weinstock & Cronin, 2003). The short measure used here is described by the researchers as a simpler version, even if it is acknowledged to lose some of the "richness" of the interview. It was intended to give a score over their three levels of epistemology: absolutist, multiplist, and evaluativist, through different domains of thinking (aesthetic judgements, value judgements, truths about the social world, and physical truths). The personal taste domain questions used in the initial study have been removed as the researchers found all participants including the children achieved the highest level possible. This study, however, was not interested in using the separate domains, as no research has investigated whether one domain (e.g. social value claim or physical truths) are more important to juror reasoning above another. Instead, the focus was on the overall balancing of competing arguments and an individual's ability to consider alternative arguments as it relates to verdict choice and conflicting expert opinion (where the opposing expert is considered to be more "right"). In this study, each type of response was given a score (with the absolutist = 1, multiplist = 2, and evaluativist = 3). The ultimate score was used as a measure of epistemological sophistication and will be able to be incorporated into the path analysis.

2.3.7 Pilot Test

Once the trial transcript and questions had been created they were piloted with a group of participants (N = 12). Nine were undergraduate students who participated for course credit, 3 were recruited from outside of the university by word of mouth. The feedback gained from the pilot test was used to modify instructions and to rewrite some of the questions where language was ambiguous or unclear.

2.4 Results

2.4.1 Usefulness and ease of understanding of expert testimony

Participants who read through the prosecution expert's testimony were asked how easy to understand and how useful it was, indicating their answers on a 1-7 Likert scale (ease of understanding M = 5.07, SD =1.61; usefulness M = 4.94, SD = 1.48). The equivalent questions were also asked about the opposing expert (ease of understanding M = 5.57, SD =1.07; usefulness M = 5.31, SD = 1.28).

2.4.2 Case Strength

Participants in all conditions were asked to indicate how strong they thought the case for both the prosecution and defence were. The results are shown in table 1, below.

Table 2. Mean ratings with standard deviations of case strength

	Control	**Prosecution expert only**	**Prosecution and opposing experts**
Prosecution	3.79 (1.32)	4.72 (1.20)	4.49 (1.50)
Defence	4.04 (1.33)	3.72 (1.25)	1.40 (1.26)

A One-Way ANOVA was conducted to compare the effect of the condition on strength ratings. For defence ratings, the effect of condition was not statistically significant, yielding an *F* ratio of $F(2, 91) = 2.38$, $p = .10$, $\eta^2 = .05$, indicating that ratings of defence strength were not different between the conditions. However, for prosecution ratings, the effect of condition was statistically significant, yielding an *F* ratio of $F(2, 91) = 3.38$, $p = .04$, $\eta^2 = .07$, which indicates there were differences between the conditions, although the effect is small.

2.4.3 Verdict

There was a significant association between condition and verdict choice, $\chi^2(2) = 9.41$, $p < .001$, with participants who saw only the prosecution expert having a higher proportion of guilty verdicts (56.2%), than the control group (16.7%), and a substantial reduction in guilty verdicts when the opposing expert testimony was heard (34.3%). This supports the claim that prosecution expert evidence would have an effect on verdict choice, as would the introduction of opposing expert testimony.

2.4.4 Normality of data for use in the path analysis

Assessment of the univariate and multivariate normality was conducted, using only data from the expert conditions. All the values were within Kline's (2005) accepted range of =/- 3 for the use in path analysis, and had skew values < 2 and kurtosis values of < 3. When examining the multivariate normality of the variables to be used in the path analysis, Mardia's multivariate kurtosis was -2.15, with a critical ratio of -1.29. Bentler and Wu (2002) suggest that a value of 3 and above indicates concern. As the value was lower than that, maximum likelihood was deemed acceptable to use.

2.4.5 Correlations

Before conducting the path analysis, the relationships between epistemology, argument skill, methodological reliability, probability, certainty and verdict choice were examined for participants in the expert conditions using Pearson correlations, presented in Table 1. There was a small negative correlation between condition and methodological reliability scores, with participants who read through the opposing expert's testimony giving lower ratings of reliability. Higher reliability scores were associated with more guilty verdicts.

Table 3. Correlations between condition, epistemology, argument skill, methodological reliability, probability, and verdict for expert conditions

	Condition	Epistemology	Argument Skill	Methodological Reliability	Probability	Verdict
Condition	1	-.00	.03	-.24*	-.12	-.22
Epistemological sophistication		1	-.02	.15	.20	.17
Argument Skill			1	-.02	.07	.11
Methodological Reliability				1	.69**	.63
Probability					1	.61
Verdict						

** p <.05 (2-tailed)*

*** p < .001 (2-tailed)*

2.4.6 Path Analysis

It was predicted that the relationship between verdicts and opposing testimony would be partially mediated by reliability ratings. It was also predicted that epistemological sophistication and argument skill would affect ratings of the methodological reliability and verdict choice. As the research questions concerned the ratings of the reliability of the evidence, only the participants from the conditions who read through expert testimony were included in the path analysis (making it a dichotomous variable). According to independent samples *t*-tests, there was no difference in the mean argument skill scores between the participants in the prosecution expert condition (M = 3.03, SD = 1.40) and those who also saw the opposing expert (M = 3.11, SD = 1.25); $t(65) = -.26$, $p = .80$. Nor was there a difference in epistemology, (prosecution expert M = 26.28, SD = 2.70; opposing expert M = 26.26, SD = 3.38; $t(65) = .03$, $p = .97$).

The model had 5 variables, with no latent variables, and no missing data. The path analysis excluded the control condition who saw no expert (and therefore were not able to give ratings on the reliability of the data), therefore making it possible to use opposing expert presence as a categorical variable in the model. The sample size was still small, with 67 participants being included in the path analysis. While there are currently a wide range of "rules of thumb" for determining sample size requirements for path analysis (Wolf, Harrington, Clark, & Miller, 2013), the sample met the requirements of Bentler and Chou (1987). Bootstrapping (maximum-likelihood) with 2000 samples was used to generate 95% confidence intervals. Bootstrapping is a resampling method used to focus on the precision of results (rather than only on significance) where random sampling distributions are created based on the original sample (Byrne, 2010). By repeated re-sampling many times an empirical distribution of the sample statistic of interest is calculated. This allows for improved estimates of variance, the estimation of standard error, and the calculation of confidence intervals. Using bootstrapping means there can be more confidence placed in the size of mediation effects within the proposed model. All analyses were conducted using Amos 21 software. Figure 2 shows the model with standardised estimates.

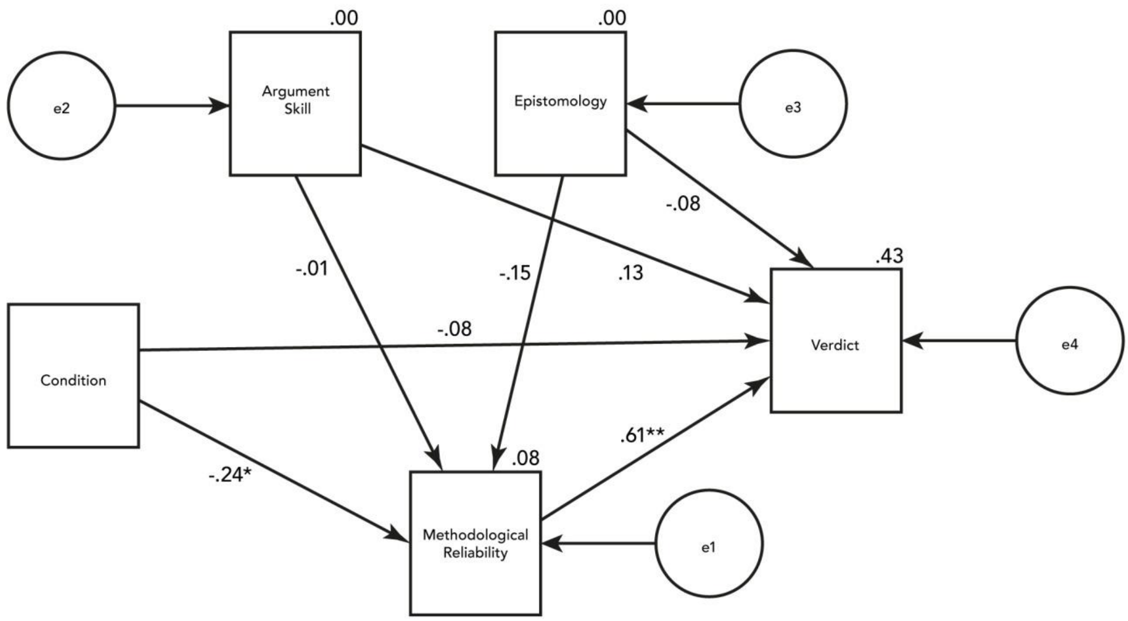


Figure 2. Path model for Study 1 showing standardised estimates

The path analysis shows that while the direct relationship from condition to verdict was very small (and not statistically significant), the relationship from condition to the ratings of methodological reliability was moderate, and that from methodological reliability to ratings of guilt was relatively large. Participants who read the testimony of the opposing expert gave lower ratings to the methodological reliability (*M* = 22.86, SD = 8.3), than those who only heard from the prosecution expert (*M* = 26.72, *SD* = 7.38). The epistemology and argument skill variables had a small, not statistically significant, effect on ratings of methodological validity and guilt ratings. Regression weights, standard errors, and direct, indirect and total effects (along with the bootstrapped values and confidence intervals) are shown in Table 3.

Table 4. Unstandardized and standardized regression weights, and total, direct and indirect effects for condition, epistemological sophistication, argument skill, methodological reliability, probability, certainty and verdict for expert conditions

	Unstandardised				**Standardised**		
	Original Estimate		Bootstrapping		Original Estimate	Bootstrapping	
	b	S.E. b	b (95%CI)	S.E.b	β	β (95%CI)	S.E. β
Condition → Verdict	-.08	.10	-.08 [-.23, .10]	.10	-.08	-.08 [-.23, .10]	.10
Total			-.22 [-.45, .05]			-.22 [-.45, .05]	
Direct			-.08 [-.26, .14]			-.08 [-.26, .15]	
Indirect			-.14[-.29, .00]			-.15 [-.28, .00]	
Condition → Methodological Reliability	-3.85*	1.89	-3.88 [-6.84, -.41]	1.92	-.24	-.24 [-.42, .28]	.12
Total			-3.85 [-7.36, .04]			-.24 [-.46, .01]	
Direct			-3.85 [-7.36, .04]			-.24 [-.46, .01]	

Indirect			....			...	
Methodological Reliability → Verdict	.04**	.01	.04 [.03, .05]	.01	.61	.61 [.45, .73]	.08
Total			04 [.03, .05]			.60 [.43, .75]	
Direct			.04 [.03, .05]			.60[.43, .75]	
Indirect			...			...	
Argument Skill → Verdict	.05	.04	.05 [.05, -.02]	.04	.13	[-.07, .31]	.10
Total			.05 [-.06, .12]			.12 [-.14, .33]	
Direct			.05 [-.03, .11]			.13 [-.07, .31]	
Indirect			-.00 [-.07, .06]			-.01 [. -.17, 15]	
Epistemology → Verdict	.01	.02	.01 [-.02, 04]	.02	.08	[-.15, 28]	
Total			.03 [-.01, .07]			.17 [-.09, .40]	.11
Direct			.01 [-.03, 05]			.08 [-.15, 28]	

Indirect			.02 [-.01, .04]			.09 [-.05, 25]	
Argument Skill → Methodological Reliability	-.08	.72	-.13 [-4.41, 1.25]	.82	-.01	[-.27, 25]	.13
Total			-.08 [-1.72, 1.46]		-.01	-.01 [-.27, .25]	
Direct			-.08 [-1.72, 1.16]		-.01	-.01 [-.27, .25]	
Indirect			…		.00	…	
Epistemology → Methodological Reliability	.41	.31	.43 [-.15, .89]	.32	.15	[-.10, 37]	.12
Total			.41 [-.26, .98]		.15	.15 [-.10, .37]	
Direct			.41 [-.26, .98]		.15	.15 [-.10, .37]	
Indirect			…		.00	…	

Examination of the direct, indirect and total effects confirm the importance of including the ratings of methodological reliability. There is some evidence that epistemological sophistication and argument skill may have a very small direct, and indirect, effect on ratings of reliability and verdict; however the results from the path analysis show that this measure did not reach statistical significance in the model.

2.5 Discussion

As predicted, the opposing expert influenced perceptions of the reliability of the methods and techniques used by the anthropometric facial comparison expert, and, through this, had an effect on guilt. Argument skill and epistemological sophistication had no substantial effect on ratings of reliability or guilt, either directly or through an indirect effect via methodological reliability.

2.5.1 Opposing Expert Testimony and Methodological Reliability

While the path analysis did show a relationship between the inclusion of an opposing expert and ratings of guilt, the total effect was very small. The prediction that there would be an indirect effect, however, was supported. Ratings of methodological reliability mediated the relationship between the opposing expert and guilty verdicts. There is therefore some support for the ability of the opposing expert's testimony to be used to educate jurors about unreliable science, and for the inclusion of mediating variables in studies of jury decision making: the opposing expert produced lower ratings of reliability, and lower ratings lead to fewer guilty verdicts. The reliability of the identification evidence was also the most important predictor of verdict.

Previous research using opposing experts has shown mixed results: scepticism (Levett & Kovera, 2008), scepticism and sensitivity (Levett & Kovera, 2009), a lack of an effect (Devenport & Cutler, 2004), sensitivity but no effect on verdict (Jones & Kovera, 2015), and the education of jurors about the limitations of unreliable evidence (Eastwood & Caldwell, 2015). As shown in the introduction, previous studies have taken very different approaches to stimulus materials and manipultions, as well as analyses. Despite this, commentators still consider opposing expert testimony to be ineffective. The current research suggests this is not the case.

There are several differences between this research and the previous studies. This study tested for a mediating relationship, rather than just a direct effect, and did not just look at differences in verdict. Additionally, the stimulus material was a case where the initial jury trial had given a guilty verdict. Unsurprisingly, the proportion of guilty to not-guilty verdicts in this study was higher than previous studies: Levett and Kovera (2008) and Jones and Kovera (2015) both had low conviction rates. In these studies the initial expert was appearing for the defence and the subsequent opposing expert for the prosecution (to argue against the study discussed by the defence expert). Eastwood and Caldwell (2015), had a similar conviction rate to this study, they found that 50% of participants in their condition with only the prosecution expert gave a guilty verdict, and 21.21% when there was an opposing expert. The current study had an increase in conviction rates of almost 3.5 times greater in the prosecution condition compared to the control condition. This may have made the effect of opposing expert evidence easier to see, as the anthropometric facial comparison evidence - much like the hair comparison evidence from Eastwood and Caldwell (2015) - was clearly perceived as highly suggestive of the defendant's guilt.

An indirect effect of opposing expert testimony, rather than just a direct effect on verdict choice, had only been looked at once before. In their 2009 study, Levett and Kovera tested an interaction between the condition where their opposing expert had addressed the validity of the research presented by the expert and the actual validity of the research, which made it possible to test for sensitivity. A statistically significant total and indirect effect was found, indicating that those participants who viewed the valid research rated the expert to have higher quality research than those who saw the invalid research. The effect sizes for the results, however, were all small. When studying case strength they found an interaction between opposing expert testimony type and validity, but it only accounted for 2% of the variance. The main scepticism effects, the interaction of opposing expert testimony presence on probability of a hostile work environment and also on ratings of the experts research, were also small, at η^2= .04 and η^2= .02, respectively. In comparison, this study found a small direct effect of opposing expert on verdict, but a moderate effect of hearing an opposing expert on ratings of reliability. There was a large effect of methodological reliability on verdict. Of this, a moderate size portion was due to the opposing expert, as can be seen by the indirect effect.

While this study was not able to test for sensitivity, it does show support for the educational value of opposing experts when there is concern jurors may have inflated beliefs in the reliability of forensic science. Eastwood and Caldwell (2015) claimed that they found support for a defence expert witness being capable of educating jurors about the limitations of hair comparison evidence. Based on the findings of Study 1, it appears that defence experts are similarly capable of educating jurors about the limitations of anthropometric facial comparison evidence.

However, there are limitations to using stimulus materials that reflect the state of hair or anthropometric facial comparison as they are in real trials. By presenting all of the criticisms as part of one testimony it is not possible to see if there are certain arguments that most jurors do or do not understand. By using a single research study, and only changing one aspect of the design to make it invalid, such as experimenter bias (Jones & Kovera, 2015) or having a missing control group or lacking counter balancing (Levett & Kovera, 2008), previous researchers have been able to focus on single, sophisticated elements of study design and to argue whether jurors are able to recognise that one problem. In this study, jurors did not need to demonstrate understanding of how each flaw would affect reliability. It would be expected that this is a more realistic representation of the problem facing jurors generally – they need to recognise the indicators of irrational or illogical conclusions drawn by forensic experts, within the totality of the evidence before them.

There are multiple threats to validity within anthropometric facial comparison evidence, and these were directly related to the evidence given by the expert. The aim of the study was to see if opposing expert testimony would help jurors recognise the limitations of the evidence, and by doing so, allow them to assess the evidence in a less biased manner. Some arguments made by the opposing expert may have been more compelling to certain jurors, but overlooked by others. For example, one juror wrote that they had a background in graphic design and knew how important picture resolution was, which led them to be sceptical of the evidence presented by the facial mapping expert. Having the opposing expert point out that CCTV footage is typically of low quality had been very persuasive, as his experience had made him aware of how hard low resolution footage was to get a clear image. Other jurors mentioned the lack of having a second expert very concerning and focused on that one critique. As is suggested by the Story Model of juror reasoning, individual juror's attitudes and experiences will lead to different narratives about the reliability of the anthropometric facial comparison evidence. When involved in a deliberation stage, each may be able to contribute understanding of a particular criticism.

2.5.2 Methodological Reliability

The ratings of the reliability of the evidence were the highest predictor of verdict choice, with a large effect. Higher ratings of reliability of the evidence lead to a greater likelihood of a guilty verdict. Neither argument skill nor epistemological sophistication had a statistically significant relationship with methodological reliability.

The studies previously using the series of questions have found varied outcomes from participants. Levett and Kovera (2008) found that more positive ratings were given to the research presented by their initial expert (who appeared for the defence) when no opposing expert was present compared to when an opposing expert appearing for the prosecution testified, but the partial eta squared scores of the difference were very low, even when the opposing expert discussed the methodology (with $\eta^2 = .01$ for change in research ratings, and .04 and .03 for trustworthiness and credibility, respectively). However, there was no effect of study validity, and participants gave lower scores when the opposing expert testified regardless of the original expert's testimony's content. McAuliff et al. (2009) observed differences in ratings between their valid and no control conditions, but no differences between conditions containing a confound or experimenter bias. However, Jones and Kovera (2015), similar to the current study, found that scores of reliability (which they called Scientific Quality) were higher when there was no opposing expert present, than when there was an opposing expert who used a demonstration to criticise the initial experts methodology. However, there was no statistically significant difference between the no expert condition, and conditions where the opposing expert either did not address the methodology or did so but with no demonstration.

Using only guilty or not guilty verdict decisions as the only way of measuring the effectiveness of an expert is neither adequate, nor informative. The series of questions used in this study has been successfully capable of revealing juror perceptions of the reliability, or quality, of varied types of research or evidence when it has been used as stimulus materials. It is far more informative than use of verdict (or probability of guilt) alone in judging the effectiveness of opposing expert testimony, as it is capable of revealing more about the mechanisms explaining variance in verdict choice, showing it is largely driven by perception of the reliability of the evidence. Levett and Kovera's (2008) questions provide a measure of whether mock jurors considered the evidence to be reliable, although they are only agreeing with statements about the reliability of evidence rather than having to reflect on the evidence. The next step would be to ask jurors and examine their responses for a more detailed understanding.

2.5.3 Argument Skill and Epistemology

The epistemology questionnaire was designed to measure epistemological sophistication across domains, but to be included in the path analysis it was used as a single, continuous measure of epistemological sophistication. The measure had not been used in this way previously, and this potentially reduced much of its explanatory power as the different domains weren't able to be considered separately. However, even when splitting the questionnaire into its separate domains and checking for correlations with other variables, no relationships were found between epistemology and the dependent variables. Notably, even the few significant correlations showed only a very small, negative correlation between argument skill and epistemology.

As part of their study looking at the use of evidence, Weinstock and Cronin (2003) looked at epistemological level using the Leadbeater and Kuhn (1989) task, and compared levels of sophistication to performance on different argument skills, including the three used in this study. The researchers found an increased likelihood of being successful at each argument skill as the level of sophistication increased, with large differences in success levels appearing between absolutists and the other higher groups. However, in this study the correlation between argument skill and epistemology was small, did not reach statistical significance, and showed a negative relationship. In the path analysis, the relationships between argument skill, epistemology and methodological reliability were also small – although the negative relationship is in the predicted direction, with higher sophistication and skill associated with lower ratings of reliability.

The skills needed to resolve and evaluate competing arguments, and the ability to balance the objective and subjective elements of knowledge claims may influence juror decision making. It may be that the measures used did not capture juror performance adequately, or may be more suited to looking at differences in verdicts and not for contested forensic testimony. Education level was not captured as a demographic detail, but as the study was presented to undergraduates as an option for gaining course credit and the remaining participants heard about the study through either snowballing or being contacted via a mailing list collected by the university, it is probable that the majority of the participants were highly educated and were either studying, or had completed, tertiary study. Education level has previously been associated with epistemological level (Weinstock & Flaton, 2004), and argument skill (Kuhn et al., 1994). By having a highly educated sample there may have been an unintentional restriction on the range of the argument skill variable, limiting the chance of showing a significant relationship.

Several participants also reported difficulty understanding the questions and the format of the Epistemology questionnaire. Weinstock and Kuhn have also measured epistemology using one-on-one interview scenario, which is correlated with results on the multiple domain questionnaire (Kuhn et al., 2000). The interview, however, is time-intensive and has not been used in studies focusing on perceptions of evidence.

The lack of effects involving argument skill may be due to limitations in the measures that emerged throughout data collection. Participants who didn't understand the photo-anthropometric evidence or the nuances in the cross examination and opposing expert testimony may have chosen to write about the other evidence. Doing this did not preclude them from getting a high argument skill score, just as those in the studies using the coding scheme previously did not require participants to discuss *all* of the elements of the trial. Out of the participants who read through the prosecution expert's testimony, only 62.7% made a reference in their responses to the short answer questions concerning the possibility the expert was biased, or that the scientific method the evidence was based on was unreliable. Additionally, the format of the experiment may have reduced participants' motivation to provide more than the bare minimum required to answer the questions. Some participants may have understood more about the limitations of the evidence but only provided a brief answer.

2.6 Summary

In addressing **RQ1**, *Can opposing expert testimony inform jurors about unreliable forensic science?,* this study has investigated the usefulness of having opposing experts appear in a trial. The testimony of a second expert informing mock jurors about the limitations of the method used by the initial expert was able to reduce the ratings of reliability, compared to a condition where only the anthropometric facial comparison evidence was used. In turn, this reduced guilty verdicts. As part of researching **RQ4**, *Are there individual differences that will affect jury decision making concerning unreliable forensic evidence?,* it does not appear that epistemological sophistication or argument skill predict how unreliable forensic evidence will be perceived, or have an effect on verdict. However, this may be due to the particular measures used and a restriction of the range in the measure due to sampling from a university population.

The results support the continued exploration of opposing expert testimony to confirm the role of perceptions of methodological reliability on guilt with a second sample, as well as to continue investigating different individual differences that could influence ratings of reliability. Therefore, Study 2, will continue investigating the impact of opposing expert testimony, while continuing to search for individual differences that may affect perception of unreliable forensic evidence.

3 STUDY 2

The first study showed support for the prediction that opposing expert testimony would be able to help jurors recognise that anthropometric facial comparison was unreliable. However, the individual differences explored did not account for variance in either ratings of reliability or in verdict choice. As the previous studies of opposing expert testimony predominantly argued that opposing expert testimony was ineffective (Devenport & Cutler, 2004; Levett & Kovera, 2008, 2009), the second study was designed to test for a replication of the indirect effect between opposing expert testimony and verdict by way of ratings of methodological reliability. If replicated, this will provide additional support for the argument that opposing expert testimony can be an effective method of educating jurors. The second study was also designed to investigate individual differences that might explain reliability and verdict choice, and, additionally, examine the content of short answer responses for more information on how jurors reached a verdict and their perceptions of the evidence.

The short answer responses from the first study showed that individual participants had a wide range of reactions to the forensic evidence, with participants who only saw the prosecution expert also recognising that the methodology of the facial comparison was flawed. Participants who read through the opposing expert testimony doubted the opposing expert, with reasons including their potential allegiance to the defendant and the lack of analyses performed by the opposing expert. The questions were designed by Kuhn et al. (1994) to test for argument skill, but, as part of evoking arguments from the participants, they also revealed which pieces of evidence, opinions, and connections were considered important. A qualitative examination of the responses, using content analysis, was conducted to more closely examine the way participants viewed the experts and the evidence. This chapter will present the quantitative path analysis results and discussion first, followed by the qualitative results and discussion, and finally a general conclusion.

As there was no support for relationships between the dependent variables and either epistemology or argument skill, but a large effect on verdict of perceptions of reliability, the investigation turned towards individual differences that would influence perceptions of the forensic evidence, rather than general cognitive abilities. The measure included within this study, the Forensic Evidence Evaluation Bias Scale (FEEBS), was recently developed by Lisa Smith, and was designed to explain variance in the perceived strength of forensic evidence.

In an initial study, Smith (2011, see also Smith, Bull, & Holliday, 2011) varied the type of evidence presented to jurors (fingerprint, DNA, or foot ware print), as well as the mobility and relevance of evidence to the investigation to see if the probative value would change. The mobility of evidence refers to how easily an innocent explanation can be given to the presence of the evidence at the crime scene, with more highly mobile items having less evidential worth (Bond, 2007). Investigative relevance depends on the sample in question. Their first study, where no case context was given, found significant main effects for mobility and relevance, but no interaction effect: regardless of the mobility of the evidence, higher relevance evidence was seen as stronger. This led them to argue that jurors may "not understand the need for evidence to be indicative of guilt in order for it to be useful, regardless of its level of mobility" (Smith et al., 2011, p. 413). In a follow up study, mock jurors were provided with enough information to form a narrative about the case (including eye witness testimony, reports from a police officer, forensic pathologist etc.). The inclusion of case contexts increased the perceived probative value of the forensic evidence compared to a control condition across all types of evidence. The effect was strongest when mock jurors had been presented with DNA evidence that was weak to moderate in strength. However, there was no corresponding shift in verdict choice, as it was only in conditions with strong forensic evidence that participants gave a guilty verdict.

Smith and Bull (2014) argued that the overall findings support an interactionist view of juror decision making, where jurors are influenced by the case specific information as well as by their individual characteristics such as attitudes and beliefs. When the case for, or against, a defendant is very strong (or weak), situational factors will be more influential. This is similar to arguments previously made by Devine et al. (2001), that when evidence is seen as strong the effect of biases are reduced, but may increase if evidence is weaker.

The second study used a similar design to the first, including the use of the same simulated trial based on *Tang,* but with two minor changes, (1) a different measure of individual differences was included, and (2) the study was conducted online. This meant that the same hypotheses from the first study were tested.

The decision to move from looking at general cognitive measures such as Argument Skill and Epistemological Sophistication led to an investigation of measures of more specific attitudes that would influence jurors' perceptions of forensic evidence. There are three pre-trial measures of general legal attitudes which have been developed to investigate juror biases that may affect legal decision making. Research into, and criticism of, these three measures led to the creation of the Forensic Evidence Evaluation Bias Scale (FEEBS, Smith, 2011). The following section will briefly describe the measures that inspired the development of the FEEBS, drawing heavily on criticism from Smith's dissertation, to highlight the advantages it has over previous measures.

3.1.1 Pre-trial measures of juror attitudes

Legal Attitudes Questionnaire

Research had shown that jurors high in authoritarianism were more likely to give harsher punishments (Mitchell & Byrne, 1973). Other studies were not consistently able to find support for an influence of personality on punitive behaviour, and little support for an influence on verdict (Kassin & Wrightsman, 1988; Narby, Cutler, & Moran, 1993). Building upon initial research, the Legal Attitudes Questionnaire (LAQ), developed by Boehm (1968), examines whether liberalism-conservatism attitudes will influence juror behaviour. There are three attitude dimensions: Authoritarianism, Anti-authoritarianism, and Equalitarianism. The initial validation studies for this scale used 151 undergraduate students who read through two versions of a murder trial and had to give a not-guilty, guilty of manslaughter, or guilty of second-degree murder verdict. In the "innocent" version of the trial, participants who scored higher on authoritarianism were more likely to convict. In the "guilty" version of the trial, participants scoring highly on anti-authoritarianism were less likely to vote for second-degree murder, and had a preference for manslaughter verdicts.

Kravitz, Cutler, and Brock (1993) made four criticisms of the scale: 25% of the respondents made an error when completing the questionnaire, the positively-scored authoritarian items and negatively-scored anti-authoritarian items were not in the same scale, the items are iterative and therefore not independent, and the anti-authoritarian subscale had weak validity (and some items from the equalitarianism subscale behaved like items from the anti-authoritarianism subscale). Revised versions of the scale have been created (see Cutler, Moran, & Narby, 1992; Kravitz et al., 1993; Moran & Comfort, 1982), which differ from the original by removing unreliable or ambiguous items, form a single score, or scale items separately (Narby et al., 1993). The most reliable revision, by Kravitz, Cutler and Brock (1993), had α =.71. However, Smith (2011) points out that participants only completed the measures, and did not look at a case or evidence, and there is no support for the use of the revised LAQ for predicting juror verdicts.

Juror Bias Scale

Kassin and Wrightsman (1983) developed the Juror Bias Scale (JBS) as a more generalised measure of juror attitudes. It was designed to be a self-report measure of predisposition towards guilt or innocence aiming to identify whether there are pre-trial perceptions that will influence these decisions, and, if they could be identified, whether these perceptions would predict verdict choice. The JBS assumes that jurors reach verdicts by evaluating two decisions, “a) probability of commission (PC), that is, the subjective likelihood (O-100%), given one’s a priori beliefs and the evidence, that the defendant actually committed the crime; and (b) reasonable doubt (RD), that is, the threshold of certainty (O-100%) deemed necessary for conviction” (Kassin & Wrightsman, 1983, p. 426). Items written to reflect the PC dimension were worded to indicate pro-prosecution bias, and RD items to reflect pro-defence bias. Jurors compare the decisions, and if the probability of the former decision is greater than the later, the juror will give a guilty verdict. The final version of the JBS had 17 items, that either measured pro-prosecution or pro-defence attitudes (reverse scored), with higher scores indicating a pro-prosecution bias. Both student and juror samples were tested and it was found to be uncorrelated with measures of belief in a just world and social desirability, but highly correlated with authoritarianism (Kassin & Wrightsman, 1983). The overall finding was that the JBS was a reliable predictor of verdict preference.

Smith (2011) argued that there were inconsistencies in the initial validation studies by Kassin and Wrightsman (1983): in their first study pro-prosecution and pro-defence participants differed in reasonable doubt criteria but not in estimates of probability of guilt; however, in the second study the opposite effects were found. Additionally, when looking at different crime types, there was no statistical difference between those scoring higher on pro-prosecution and pro-defence in terms of verdict choice.

De La Fuente et al. (2003) examined the validity of the scale (the original version translated into Spanish) when the strength of evidence was varied, and the influence on verdict choice before and after deliberation. The researchers gave mock jurors two trials to read through: one containing clear evidence and one containing ambiguous evidence. Scores on the JBS predicted the verdict in the homicide case containing ambiguous evidence, but no effect was shown when the evidence was clear. The researchers suggest that strength of evidence may mediate the association between verdict and bias towards prosecution or defence. Participants also went through a mock deliberation process, in juries that were made up of either those biased towards the prosecution or the defence. The divergence in guilty verdicts between those scoring more highly on pro-defence or pro-prosecution biased scales increased after deliberation (fewer guilty verdicts in the defence-biased juries) but this was only seen in the ambiguous evidence trial. The researchers also highlighted the fact that there was no difference in ratings of the ambiguity of the evidence between prosecution- and defence-biased jurors, even though there was a difference in verdicts both pre- and post-deliberation. De La Fuente, De La Fuente and Garcia (2003) argued that this may be because individual differences may be mediating variables, rather than having direct effects.

Myers and Lecci (1998) subjected the original JBS scale to a factor analysis, which had not been done during the initial development. The results did not find two distinct structures, which was the theorised design as illustrated by Kassin and Wrightman (1983). Myers and Lecci then revised the scale so it had a six item, two factor pro-prosecution scale, and a six-item three-factor pro-defence scale. The original and revised scales were compared by Lecci and Myers (2002), who compared results from trials featuring murder, rape and armed robbery. The revised pro-prosecution scale, was split into the two items measuring "confidence in some aspect of the criminal justice system" and "cynicism toward some aspect of the justice system" (Myers & Lecci, 1998, pp. 246-247), was better able to predict verdict than the original JBS (Lecci & Myers, 2002). Smith (2011) points out that in neither trial were the samples entirely jury eligible, with the former study featuring participants who were too young, and the latter with participants who were too old. Lecci and Myers also recognise that conclusions of the 1998 study may be limited as the trial featured rape (prior to the murder), with previous studies showing inconsistent responses among jurors when sexual assault was featured in stimulus materials (2002).

Pre-trial Juror Attitude Questionnaire

Although they acknowledged that research using the JBS had contributed substantially to understanding the influence of pre-trial attitudes, as a follow-up, Lecci and Myers (2008) argued that the JBS could not capture bias that would lead to specific juror decision making. The pro-defence scale could be generalised to almost all cases, and items on the pro-prosecution subscale would be too case specific. Their previous research and discovery of the two factor nature of the pro-prosecution scale made the use of the JBS more complicated than was desirable.

The Pre-trial Juror Attitude Questionnaire (PJAQ) was designed to specifically measure attitudes relevant to jury decision making. The PJAQ contains 30 items, and in an initial study was presented to 601 college students alongside the 17 items from the JBS. Six factors emerged after confirmatory factor analysis was used: conviction proneness, system confidence, cynicism toward the defence, social justice, racial bias, and innate criminality. Item analyses led to the removal of problematic items, and the final PJAQ scale contained 29 items, which included 12 items from the JBS. A third study found that the PJAQ was a better predictor of verdict than the JBS and revised LAQ measures. The PJAQ was tested against different length video-taped trial simulations to investigate predictive ability during murder, rape and robbery trials (Lecci & Myers, 2009). Four out of six factors (all but racial bias and cynicism) were significantly associated with verdict choice except during the rape trial (where none of the factors were associated), and overall between 15% and 21% of variance in the verdict was accounted for. The researchers did note that the simulated trials may have limited the ability of the PJAQ to predict verdict change, as the overall conviction rate was quite high, at 0.73 for pre-deliberation verdicts. Attitudes and bias are expected to carry the greatest weight when verdicts are around 50%, allowing an individual's personal preferences to influence judgement (Devine et al., 2001; Kalven & Zeisel, 1966). Lecci and Myers also concluded by saying that "it is expected that indirect effects, particularly as mediated by the evaluation of evidence, would carry the greatest influence of pre-trial bias" (Lecci & Myers, 2009, p. 2029).

In her 2011 dissertation, Smith argued that the previous scales of juror attitudes have successfully been able to predict verdict preference, but only under some circumstances. The JBS can predict verdict preference but only when evidence is ambiguous, and deliberation appears to accentuate the bias. The PJAQ was created using the act-frequency approach to creating a measure, based on work by Smith and Kendall (1963) and Buss and Craik (1983). However, this method is based on examining the frequency of responses from a sample, in this case 42 students who were asked to generate three items that would demonstrate "a bias that was likely to affect verdicts" (Lecci & Myers, 2008, p. 2013). A second group of 110 students then rated how likely they were to represent pre-trial bias on a 5-point Likert scale. Although Lecci and Myers argued this was the best method of identifying lay perceptions of bias that would affect verdict choice, it does not build upon previous research in the jury-decision making field, nor does it rely on prior literature concerning cognitive biases. Smith argued that using literature on jury decision making may have generated more "robust" scale items (2011). Smith also argued that the lack of investigation of any dependent variable other than verdict choice means that a lot of other information about the perception of evidence remains unknown, and, that there was no explanation for why some subscales had more predictive ability than others. In summary there are questions about the validity of the scale.

The LAQ and JBS were developed and tested before the use of forensic science became commonplace, and the early studies developing the scales did not look at physical evidence (Smith, 2011). The PJAQ features a mixture of evidence including police reports, an alibi given by a friend of the defendant, a police line-up, and fingerprints from the scene[2] (Lecci & Myers, 2009), but did not investigate whether there were attitudes that could bias perception of forensic evidence. It is unclear whether any of the responses from the original

[2] As part of the argument for the creation of the FEEBS, Smith (2011) argued that "newer attitude measures (e.g. PJAQ) have used exclusively non-physical evidence in trial stimuli" which is not correct.

95 items in the PJAQ discussed forensic evidence. Finally, Smith argued that the limited focus on the dichotomous verdict outcome may limit results, as "it is possible that the impact of pre-trial juror bias may be operating at a lower level than ultimate verdict preference such as the evaluation of evidence, criteria for reasonable doubt or the determination of mitigating circumstances" (Smith, 2011, p. 54). This, combined with an interest in testing the CSI effect, led to an investigation of lay-person perceptions of forensic evidence.

Forensic Evidence Evaluation Bias Scale

The questions on two subscales of the FEEBS were initially developed by examining the CSI effect, and were designed to reflect potential beliefs regarding forensic evidence which could affect verdict choice. As reviewed in Chapter One (section 1.7.5), empirical studies have failed to find support for an effect on verdict choice (Cole, 2015; Cole & Dioso-Villa, 2011), and there are mounting claims that the CSI effect is anecdotal rather than a real influence on juror behaviour (Podlas, 2006a). But this is true if the CSI effect is operationalised by viewing habits – particularly, by frequency of viewing CSI or CSI type shows. High expectations that forensic evidence will appear in a trial will also place an unrealistic burden on the prosecution, and while there is little support for an influence on verdict choice, high expectations of forensic evidence have been found in multiple studies of forensic evidence (Kim et al., 2009; Shelton, Kim, & Barak, 2009), including in interviews with Australian jurors (Wheate, 2006). Secondly, the defence may be disadvantaged if jurors do not understand the capabilities of forensic evidence. They may see it as more reliable and trustworthy than it really is and subsequently be more prone to convict.

During the construction of the FEEBS, an initial pool of 21 items were developed that "reflected beliefs that jurors have anecdotally reported to hold (within the CSI effect literature), which are thought to be potentially problematic in cases where forensic evidence is presented" (Smith, 2011, p. 146). These items, along with four filler items reflecting more general attitudes towards being a juror, and six items from the JBS and LAQ (to disguise the purpose of the measure) were presented to 219 UK jury eligible participants. The final version of the scale, after items were removed based on inter-item and item-total correlation scores, includes 10 items, included in Appendix B, and has a Cronbach's α=.78.

After PCA, two clusters emerged, which were argued to be "conceptually similar to the two main types of bias described by the *CSI effect* literature" (Smith, 2011, p. 156). The two components explained 47.99% of variance in participant responses. The components were then separated into the two subscales, of five questions each. The first looks at a pro-prosecution bias (FEEBSpp), the second measures pro-defence bias (FEEBSpd). Items in the first subscale, FEEBSpp, tap into beliefs concerning the infallibility and conclusiveness of forensic science, e.g. "forensics always IDs the guilty person", whereas the second subscale gives a rating for the reliance on forensic evidence for conviction, and the expectation that it will appear in a court trial, e.g. "if no forensics are recovered the defendant is probably innocent".

The scale has been tested using exploratory factor analysis with undergraduate students (Smith, 2011; Smith & Bull, 2012), and Confirmatory Factor Analysis (CFA) on a community sample of jury eligible participants in the United Kingdom (Smith, 2011; Smith & Bull, 2014). In the earlier study, undergraduate students were presented with the FEEBS and a murder trial involving DNA evidence that was of low-moderate probative value (traces from on a cigarette found at the crime scene), which would allow juror biases to influence perceived strength (Smith & Bull, 2012). A mediation model was tested between bias (as measured by the FEEBSpp) and probability of guilt, and found a significant indirect effect of the strength of DNA evidence on guilt ratings. Interestingly, neither the JBS nor General Belief in a Just World Scale (GBJW, Dalbert, Montada, & Schmitt, 1987) correlated with the FEEBSpd subscale. And while there was a statistically significant correlation between the FEEBSpp and JBS, and a very small correlation with GBJW, the comparison indicates that it is measuring a more specific, and conceptually different, attitude. The FEEBpp subscale was the only predictor of the strength of the DNA evidence. Each version also contained an eyewitness.

In the final study, Smith used a 2 x 2 between-subjects design, presenting jurors with a robbery or sexual assault trial that either did or did not involve DNA evidence from the prosecution. Again, the CFA confirmed the two component nature of the FEEBS. Positive statistically significant correlations were found between FEEBSpp and the GBJW, and small negative correlations between the FEEBSpd and subscales of the JBS and GBJW. However, much like in the previous study, the correlations were all small. Perceived strength of the DNA evidence was the only statistically significant predictor of verdict in the robbery scenario, and perceived strength of DNA and of the eyewitness evidence were the only predictors in the sexual assault trial. The researchers argued that this later finding may be because the credibility of rape victims has been shown to rely upon having an eyewitness giving evidence that an attack occurred (Sleath & Woodhams, 2011, as cited in Smith & Bull, 2012). In the conditions where no DNA evidence was presented, the only significant predictor of probability of guilt ratings was testimony presented by the arresting officer. For both evidence types, there was a statistically significant negative relationship between the FEEBSpd and the perceived strength of this testimony, indicating that higher expectations that forensic evidence would appear in the trial led to lower ratings of the strength of the officer's testimony. This supported the researchers' predictions, which were based on the "anti-prosecution" version of the CSI Effect (Maeder & Corbett, 2015; Podlas, 2006a), as one of the concerns of the CSI Effect was that jurors would be unwilling to convict when no forensic evidence was present, and would find other types of evidence non-persuasive. A mediation model was tested between FEEBSpd, perceived strength, and probability of guilt, but the model was only significant in the robbery scenario. Overall, while the relationship between the scale and verdict choice may be influenced by case specific contexts, the measure shows good predictive validity across a variety of trial types, including sexual assault, which Smith (2011) points out had not been the case with the JBS or PJAQ (Kassin & Wrightsman, 1983; Lecci & Myers, 2008; Lecci & Myers, 2009).

The FEEBS is therefore a more appropriate inclusion in studies of juror responses of unreliable evidence than a more general jury attitude questionnaire, and will potentially reveal more about individual decision making than looking at general cognitive abilities such as Epistemology and Argument Skill. Pro-prosecution bias may capture an important aspect of juror decision making when forensic evidence is invalid. Jurors who are high on the FEEBpp scale may be inclined to ignore potential faults and threats to reliability, and trust the prosecution expert (therefore giving them a higher reliability rating, and therefore being more likely to give a guilty verdict). In contrast, that individuals who do not have such a large amount of trust in forensic science may be more accepting of the criticisms given by the opposing expert, and will therefore give lower ratings of the reliability of the evidence.

The FEEBS was only developed recently, and has had limited testing: the creators have looked at perceived strength of DNA in the context of a murder, sexual assault and robbery trials (Smith & Bull, 2012; 2014). It has not yet been used to measure perceptions of other types of forensic evidence, or when invalid science is used. Incorporating the FEEBS scale into the current study allows for an investigation of whether bias towards forensic evidence will directly affect ratings of methodological reliability (rather than evidence strength ratings), and whether this will indirectly or directly affect verdict choice. It was predicted that there would be direct effect of pro-prosecution bias on verdict, and indirect effects through methodological reliability, with higher scores on the FEEBSpp leading to higher methodological reliability ratings, and, in turn, a higher likelihood of giving a guilty verdict. Figure 3 shows the model to be tested.

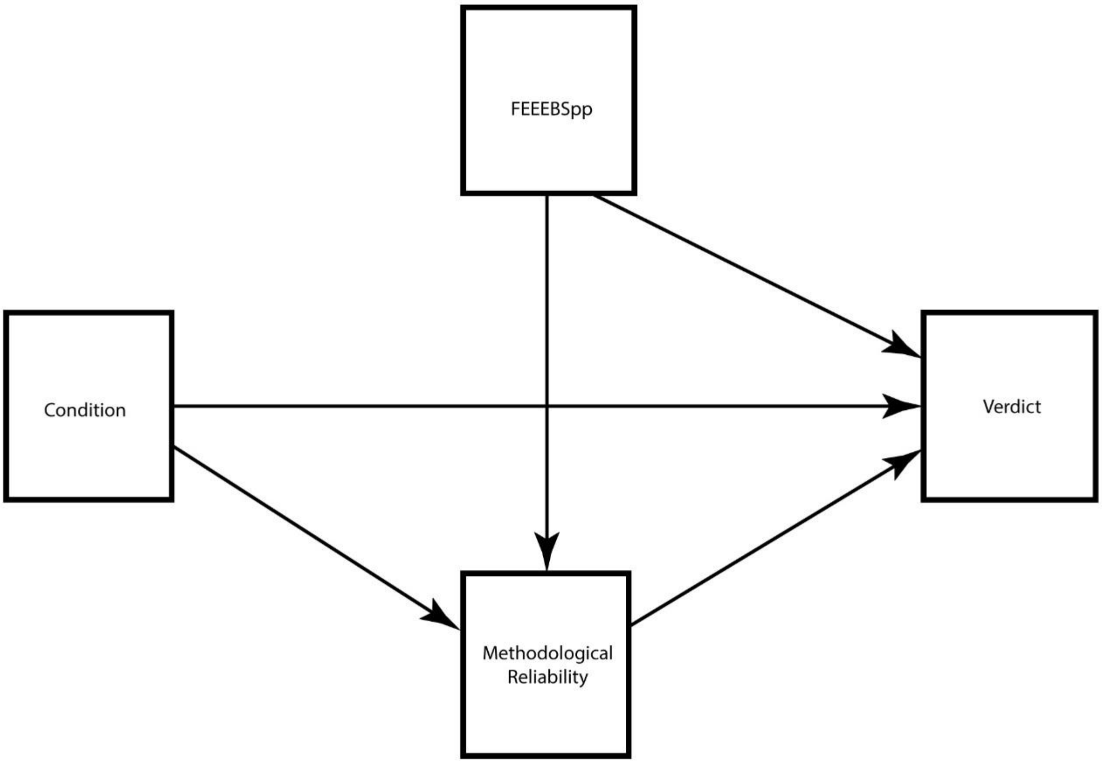


Figure 3. Path model showing predicted relationships between Condition, Methodological Reliability, scores on the FEEBSpp subscale and Verdict

3.1.2 Note Taking

In Australian courts, jurors' note taking is allowed if permission is given by a judge; however, a survey of 136 Australian and 49 New Zealand judges found that most permit it (Ogloff, Clough, Goodman-Delahunty, & Young, 2006). As some participants in the first study had expressed a desire to make notes while they had been reading this option was included for all participants. Prior studies have found that free style note taking may lead to increased juror accuracy in recall of trial details (Costabile & Klein, 2005; Forsterlee, Kent, & Horowitz, 2005; Horowitz & ForsterLee, 2001). More recently, Thorley, Baxter and Lorek (2016) found that note-takers recalled trial information more accurately even if they did not have access to the information at a retrieval stage, however note-takers were not significantly more likely to distinguish between accurate and false information than non-note-takers.

The use of an online platform in study two allowed easy note taking by participants. It was believed this would increase satisfaction and enjoyment of taking part in the study. Participants were also asked the same questions about their verdict choice as in the first, with the goal of conducting a content analysis on the responses. It was also hoped that allowing jurors to take notes would increase the quality of responses without affecting the persuasiveness of the evidence presented in the trial, as the note taking was free style (and presumably, jurors would only take notes if they were concerned about forgetting details or already considered the information important). However, to ensure there was no difference in ratings of methodological reliability or verdict, scores between note-takers and non-takers would be compared.

3.2 Method

The study was hosted on the online platform Qualtrics™. Participants were able to sign up for the study through two anonymous portals: one was designed for undergraduate psychology students who participated for course credit; the second was for members of the public who participated for the chance to win a $150 voucher. Each participant completed the study individually, and was warned at the beginning of the study that they needed to complete the study in one sitting (for Information Sheets, Consent Forms, and Instructions see Appendix B). They were also informed that Qualtrics™ would record the IP addresses of each participant to ensure that no one could do the study twice. The information and consent forms given to participants were otherwise the same as in the initial study.

The study began by asking participants about demographic details and asking them to complete the FEEBS. The trial transcript was then presented across several successive pages, and was the same as in the first study with one notable change. To be more in line with identification evidence reporting recommendations, at the end of the prosecution expert's testimony she now reported that "Given the number and quality of matches between facial features on the surveillance footage and pictures I was given of the accused, I would say they show a high amount of similarities". Qualtrics™ allowed for the recording of the time taken on each page although this was not shown to participants. After the trial participants were asked to give a verdict and to answer the methodological reliability measure, (which had a Cronbach's α of .91). For this study, participant responses from the two samples were initially kept separate.

Qualtrics™ allowed participants to make notes using text entry boxes placed at the bottom of each page. Any notes made appeared on the final pages where participants made verdict choices and answered short answer questions.

3.3 Results – Quantitative Analysis

In the public sample 150 participants began the online study, with 76 jury eligible and completing it. In the undergraduate sample, 91 undergraduate psychology students clicked on the link, with 71 eligible participants completing the trial.

Participants were asked to report their highest level of completed education, with 51.7% having completed high school, 29.9% having completed an undergraduate degree, 11.6% a postgraduate course, 5.4% a TAFE qualification, and 1.4% other trade qualifications. An "other" category was provided with a text entry option but no participant who selected this option completed the study. This is somewhat different to the rates of actual jurors: an exit survey of empanelled jurors in 2008 found that 26.1% had a university degree as their highest educational qualification, 14.5% had a trade certificate or equivalent, 15.0% had a diploma or equivalent, and 39.0% had their highest qualification as high school or less (O'Brien, Goodman-Delahunty, Clough, & Pratley, 2008). This study also has a large difference in age with only 18.6% of empanelled jurors from that study being between the ages of 18-29, whereas 82.3% of the participants in this study were within that age group.

3.3.1 Public and Undergraduate Sample Differences

Independent samples t-tests were conducted comparing the two FEEB subscales scores of the participants who signed up through the undergraduate participation system, and those who signed up through the public online portal to see if differences would emerge. Significant differences were found between the two groups for both the prosecution bias subscale of the FEEBS (public *M*= 12.29, *SD* = 3.18, undergraduate *M* = 13.47, *SD* = 3.21), *t* (145) = -2.23, *p* <. 05, [-1.18, .53], and the defence bias subscale (*M* = 11.50, *SD* = 2.85, undergraduate *M* = 13.21, *SD* = 3.07), with participants from the undergraduate sample having higher scores on both subscales, *t*(145) = -3.51, *p* <. 05, [-1.71, .49]. A chi-square analysis was also conducted to see if there was a difference in verdict choice; the proportion of participants choosing guilty did differ by participant type, χ^2 (1, *N* = 147) =11.78, *p* > .001, with the undergraduate sample making more guilty verdicts (47.9% of verdicts) than did the public sample (21.1%). However, the two samples also differed significantly in age (public *M* = 28.12, SD = 10.94, undergraduate *M* = 21.24, *SD* = 7.05), with participants from the public sample being older. A Mann-Whitney U Test showed significant differences between the two groups of participants in age, *U* = 979.00, *z = -*6.71*,* p > *.001, r*=-0.05. The effect size of the difference was very small.

It is possible that another unidentified difference between the samples could lead to differences in verdict choice; however, as age was subsequently shown to be correlated to the FEEBS subscales and methodological reliability scores; the two participant samples were combined when conducting the remaining analyses (see Correlation section below).

The combined age of participants who completed the study ranged from 18 to 62 (*M* = 24.8, *SD* = 9.86), and 61.9% were female. Just over half (51.7%) of participants were recruited through the advertisement. Combined scores on the Prosecution Bias subscale ranged from 5 to 21 (*M* =12.86, *SD* = 3.24); scores on the Defence Bias subscale ranged from 5 to 20 (*M* =12.33, *SD* =3.07).

3.3.2 Note-taking

Scores between note-takers and non-note-takers were compared to ensure there was no difference in ratings of reliability or verdict. There was no statistical difference in the number of participants taking notes between conditions, $\chi^2(2)= 1.9$, $p > .05$.

An independent samples t-test was run on participants who saw the anthropometric facial comparison expert (a combined sample of conditions two and three) comparing the methodological reliability score of note takers ($M = 23.19$, $SD = 7.59$), and non-note takers ($M = 22.54$, $SD = 7.84$). There was no significant difference in ratings, $t\,(93) = -.41$, $p > .05$, [-3.80, 2.50]. As a final comparison, a Chi square analysis was run to compare note taking and verdict choice. There was no significant association: participants who took notes were just as likely to convict as those who did not take notes in all three conditions, $\chi^2\,(1) = .73$, $p =.39$.

3.3.3 Verdict choice and guilt

There was a significant association between the condition and verdict choice, $\chi^2\,(2) = 13.25$, $p < .001$; with participants who saw only the prosecution expert having a higher proportion of guilty verdicts (54.1%), than the control (21.3%), and opposing expert conditions (with the opposing expert condition almost returning to the rate of the control condition, 26.9%). As was shown in Study One, the prosecution expert evidence had an effect on verdict choice, as did introduction of opposing expert testimony.

3.3.4 Normality of data for use in the path analysis

Assessment of the univariate and multivariate normality was conducted, using only the data from the expert conditions, as the data from the control condition was not included in the path analysis. The distribution of age was not normal, with significant skew and kurtosis (*z*-scores were 10.65 and 9.438 respectively); however this was not going to be included in the path analyses. All values were within the range of = +/- 3, with skew and kurtosis values < 2. When examining the multivariate normality of the variables to be used in the path analysis, Mardia's multivariate kurtosis was -4.04, with a critical ration of -2.91. Bentler and Wu (2002) suggest that a value of 3 and above indicates concern when interpreting the $\chi2$ value.

However, given that the bootstrapping produced bias corrected confidence levels, the analysis proceeded using maximum likelihood estimates.

3.3.5 Correlations

The relationships between age, FEESBs subscales, methodological reliability, probability, certainty and verdict choice were examined for participants in the expert conditions using a Pearson correlation analysis, presented in Table 7. There was a moderate correlation between age with all other variables except for the defence bias subscale of the FEEBS and certainty.

Table 5. Correlations between condition, age, FEEB subscales, methodological reliability, probability, certainty and verdict for expert conditions

	Condition	Age	FEEBSpp	FEEBSpd	Methodological Reliability	Probability	Certainty	Verdict
Condition	1	.12	-.04	-.12	-.25**	-.19	-.07	.28**
Age		1	-.30**	-.32**	-40**	-.25**	-.11	.35**
FEEBSpp			1	.55**	.38**	.16	.21	-.31**
FEEBSpd				1	.34*	.25*	.19	-.32**
Methodological Reliability					1	.54**	.31**	-.69**
Probability						1	.34**	-.64**
Certainty							1	-.39**
Verdict								1

3.3.6 Path Analysis

As for the first study, it was predicted that the relationship between verdicts and opposing expert testimony would be partially mediated by reliable ratings of the anthropometric facial comparison evidence. It was also predicted that pro-prosecution bias towards forensic evidence would be associated with ratings of the methodological reliability and with verdict choice. According to independent samples t-tests, there was no difference in the mean FEEBSpp scores between the participants in the prosecution expert condition (*M* = 13.08, *SD* = 3.24) and those who also saw the opposing expert (*M* = 12.75, *SD* = 3.53); $t(97) = .50$, $p = .62$, *95% CI* [-1.02, 1.69].

The model had four variables, none of which were latent, and no missing data. As with the first study, only participants who had read through the expert testimony were included in the path analyses. The sample size was still small, with 100 participants being included in the path analysis. Bootstrapping (maximum-likelihood) with 2000 samples was used to generate 95% confidence intervals. Figure 4 shows the model with standardized estimates.

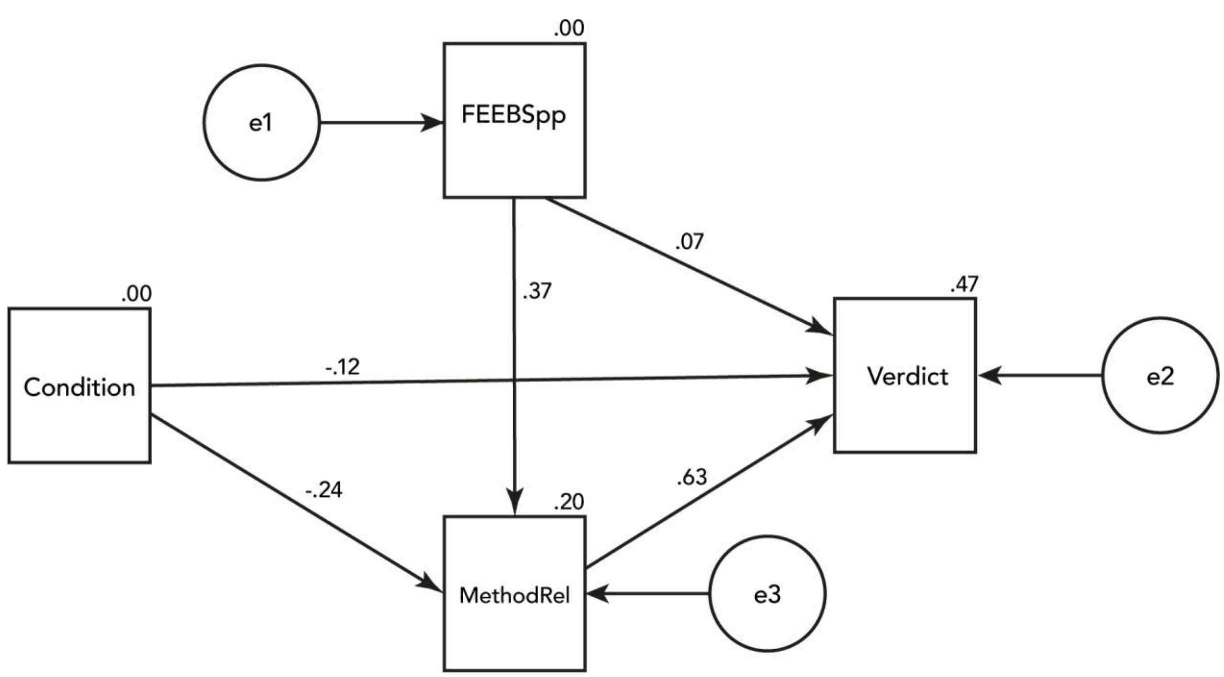


Figure 4. Path model for Study 2 showing standardised estimates.

As was predicted, the opposing expert had a moderate sized total effect on ratings of methodological reliability and on verdict choice, with the presence of an opposing expert leading to lower ratings of reliability and decreased likelihood of giving a guilty verdict, although the effect on verdict did not reach statistical significance (Table 4). Methodological reliability had a large total effect on verdict choice such that the lower the ratings of the forensic evidence's reliability, the less likely a participant was to give a guilty verdict. The FEEBSpp subscale scores had moderate effect on ratings of methodological reliability.

Table 6. Unstandardized and standardized regression weights, and total, direct and indirect effects for FEEBpp, methodological reliability, and verdict for expert conditions

	Original Estimate			**Bootstrapping**			
	b	S.E. b	b (95%CI)	S.E.b	β	β (95%CI)	S.E. β
Condition → Verdict	-.11	.07	.11 [-.03, .27]	2.17	-.12	.12 [-.03, .28]	.08
Total			.26 [.08, .44]			.27 [.08, .45]	
Direct			.11 [-.04, .26]			.12 [-.03, .27]	
Indirect			.12 [.01, .22]			.15 [.04, .27]	
Condition → Methodological Reliability	-3.56*	1.36	-3.56 [-6.15, -.78]	1.29	-.24*	-.24 [-.42, -.06]	.09
Total			-3.56 [-6.15, -.78]			-.24 [-.42, -.06]	
Direct			-3.56 [-6.15, -.78]			-.24 [-.42, -.06]	
Indirect			…			…	
Methodological Reliability → Verdict	.04**	.01	.04 [.05, .03]	.00	.63**	.60 [.75, .51]	.07
Total			.04 [.05, .03]			.60 [.75, .51]	

Direct			.04 [.05, .03]			.60 [.75, .51]	
Indirect			...			..	
FEEBSpp → Verdict	.01	.01	-.01 [-.03, .01]	.01	.07	-.04 [-.23, .10]	.08
Total			-.04 [-.07, -.02]			-.30 [-.07 -.02]	
Direct			-.01 [-.03, .01]			-.07 [-.22, .10]	
Indirect			-.03 [-.06, -.02]			-.24 [-.37, -.12]	
FEEBSpp → Methodological Reliability	.84**	.20	.84 [.43, 126]	.22	.37*	.27 [.20, .53]	.09
Total			.84 [.43, 1.26]			.28 [.08, .45]	
Direct			.84 [.43, 1.26]			.28 [.08, .45]	
Indirect			...			...	

3.4 Discussion

The results of the path analysis supported findings from the first study: the predicted relationship between opposing expert testimony, ratings of reliability, and verdict choice were found in a second sample, with the opposing expert's testimony leading to lower ratings of methodological reliability, with methodological reliability reducing the likelihood of giving a guilty verdict. Bias towards forensic evidence also showed the predicted relationships with the other dependent variables: higher scores on the FEEBSpp lead to higher ratings of reliability, and increased the chance of a guilty verdict.

3.4.1 Opposing Expert Testimony and Methodological Reliability

The replication of the indirect relationship between opposing expert testimony and verdict choice via ratings of methodological reliability supports the idea that opposing experts are capable of educating jurors about the reliability of forensic science. Perception of the methodological reliability of the evidence played a substantial part in explaining variance in verdict choice, with 49% of variance explained by the model. Lower ratings of the reliability of the anthropometric facial comparison evidence led to fewer guilty verdicts. The strength of the relationship between reliability and verdict is also very similar to the same relationship in the first study (.61 compared to .63). Rates of verdict choice were also similar, with just over half the participants who saw only the prosecution expert in each study giving a guilty verdict (56.2% in study 1 and 54.1% in study two).

3.4.2 Bias Towards Forensic Evidence

The second study has also shown that measuring specific individual differences in bias towards forensic science may have more predictive utility than general cognitive measures. Scores on the pro-prosecution subscale of the FEEBS had a small, not statistically significant, direct effect on verdict, but a moderate total effect. This operated through ratings of methodological reliability: the higher the score on the FEEBS, the higher the reliability score, and the more likely participants were to give a guilty verdict, with a moderate effect size. The evidence presented here is consistent with that of Smith and Bull (2012, 2014), showing that the initial pre-trial bias of a juror can have a large impact on perception of evidence presented during a trial, and that this can have an important effect on verdict choice.

In the initial validation study, Smith (2011) found that the pro-prosecution bias was working through an indirect effect, which was hypothesised and tested after seeing that the participant's scores on the FEEBSpp were the only significant predictor of strength ratings for the DNA evidence featured in the trial. Strength ratings had been given to each item of evidence (eyewitness, police officer, pathologist, and DNA analyst), and had been measured as responses to a question asking how useful the evidence was for determining guilt, given on a 1-5 Likert scale. While it is currently unclear how conceptually similar these scores are to agreeing with the statements written by Levett and Kovera (2008) that were used in the Methodological Reliability scale, both are implicitly asking participants to judge the probative value of the evidence.

Using the FEEBS shifts attention from measuring the frequency of viewing CSI-type shows, and instead measures statements reflecting the problematic attitudes discussed in the CSI Effect literature. This does not necessarily show support for a "CSI" Effect, as there is no way of knowing where or how participants were exposed to exaggerated portrayals of forensic scientists, or how much is based on lack of awareness of scientific principles. However, it does show that among many of the participants in this study, there is a high expectation that forensic science will appear in court, and that this evidence will be conclusive. Additionally, although the scores were not used in the path analyses (given that the only participants in the analyses had been exposed to forensic evidence, and the FEEBSpd is concerned with attitudes when such evidence is absent), it appears that many participants were also inclined to agree with the pro-defence statements of the FEEBSpd. Tyler (2006) had suggested that exposure to CSI type shows would have this type of contradictory effect: raising expectations so high that jurors were reluctant to convict in the absence of forensic evidence, but also raising trust when it is adduced to a court so that the strength is exaggerated.

So far, few studies have incorporated the FEEBS measure into their research, and those who have incorporated the FEEBS originated from the initial authors of the measure. While this study was looking at a similar crime type to one featured in previous studies (murder) the trial itself was completely different, as it featured a previously unused type of forensic evidence (as Smith had only used DNA and fingerprints as forensic evidence), and came from a different country. This demonstrates the broad applicability of the measure.

Both this study and those conducted by Smith and colleagues had large proportions of participants who were undergraduates or had completed post graduate studies (there were also very similar scores given on the FEEBSpp and FEEBSpd, with Smith's study having 13.5 and 11.41, respectively). For example, in the first study using the FEEBS, only 19% did *not* have an undergraduate or postgraduate degree (Smith, 2011). This differs substantially to those who become empanelled jurors, both in Australia and in the UK (Matthews, Hancock, & Briggs, 2004; O'Brien et al., 2008). Smith (2011), by using online sampling as well as snowballing methods, was able to get a large range in ages, with the average age being 34.77. Despite being online, this study had a slightly lower mean age of jurors, however inspecting participants' reported mean education levels showed that it is clear there was a good range of participants.

3.4.3 Age

There was a moderate correlation between age with all other variables except for the defence bias subscale of the FEEBS and certainty. No studies specifically examining the association between juror age and forensic science were able to be located through a literature search. Many studies of jurors rely on undergraduate students, where the mean age will likely be in the low twenties, and possibly contain a restricted range. Research on actual jurors in Australia has shown that 18.6% of empanelled jurors (compared to 20.6% of jury eligible citizens) are between the ages of 18-39 (O'Brien et al., 2008), whereas in this study, 82.3% were in that age bracket. An informal consensus seems to have been arrived at concerning the lack of relationship between age and verdict decisions. Because of this, few studies include age in their hypotheses, or in analyses, so important information about juror decision making may be being missed.

Additionally, age is likely to be related to education level (which was not able to be included in the statistical model). It is possible that older participants in this study had reached higher levels of education, as this has shown to be related to better understanding of some forensic evidence (Hans, Kaye, Michael Dann, Farley, & Albertson, 2011), although a meta-analysis of jury decision making studies more broadly has shown no support for an effect on guilt (Devine & Caughlin, 2014). Future studies should investigate whether there is a difference in the perception of forensic evidence between age groups and between education levels, or even between those who have had more exposure to scientific methodology and principles through formal and informal education or occupation.

3.5 Results - Qualitative Analysis

3.5.1 Aim

Content analysis of the short answer questions had several aims: primarily, to investigate how mock-jurors justified their chosen and alternate verdicts, and whether the forensic evidence was overwhelming, and if, and how, participants responded to arguments concerning the flawed methodology of anthropometric facial comparison evidence.

The questions also required jurors to think of reasons why their verdict was not the correct choice and to consider other hypothetical jurors' points of view, similar to the tasks required when part of a deliberating panel. By examining the answers it would be possible to see how mock jurors justified their verdict choice, and to see how the arguments made by the two experts were received.

3.5.2 Preparation for analysis

The analysis was conducted in several stages. While there were research questions that the analysis was designed to answer, the coding scheme was data driven, with categories based on common themes within the responses that were developed after reading the data multiple times. The analysis was conducted in several stages.

1st - Read through

This was to see how participants were responding, to see how detailed (or vague) answers to questions were, and to get a general feel for the themes. Responses were read through several times.

2nd - Grouping responses into arguments and excluding responses

Responses that didn't meet the inclusion criteria (as discussed in the general scoring rules in Appendix B) were removed. Twenty-two participants were excluded for having short responses. Fourteen were from the undergraduate sample, and 8 from the public sample. Six were from the first condition, seven from the second, and nine from the third. Each response was separated into individual arguments. This helped to reveal the types of responses participants were using, and the types of themes that were emerging. For the purposes of the analysis, we must assume the comments written in the short answer question responses reflect the understanding of the participants.

> *E.g. "Because, **[**outside of the match made through the film**]** **[**there wasn't enough convincing evidence to say that the third offender was the accused**]**. The **[**film quality was not good enough to make that accusation on its own**]**, and the accusation was only possible because of **[**fingerprints (which might have been placed several hours after the crime)**]** and a **[**possibly inaccurate method of identification**]**. Reasonable doubt exists. I call not guilty". (ID qCRo)*

3rd - Identifying themes

After the responses had been separated into arguments, themes were identified. The questions given to participants asked them to consider reasons why their selected verdict was justified, as well as to consider reasons why it could have been wrong, and why other participants (or hypothetical jurors they could meet in a deliberation room) would have preferred the alternate verdict. This meant that jurors had to consider both verdict choices, and led to jurors covering a larger amount of evidence in their responses than would have been mentioned if they only had to respond to: “why did you pick that verdict?” However, each participant could use a single piece of evidence in support of a guilty verdict in response to one question, while in the next question the participant could point out reasons why that evidence was satisfactorily accounted for by the defence, or vice versa. Distinct categories were therefore made to separate thematic arguments into those justifying the guilty verdict and those justifying the not guilty verdict. In the example above, all the arguments were made to justify a not guilty verdict.

4th – Coding

Arguments were assigned individual codes. During this process, several thematic categories were modified and merged together when it became apparent that responses blurred the lines between the two, or the categories needed to be more precise.

E.g. arguments from previous example **with final coding category**
[outside of the match made through the film**] Anthropometric facial comparison, unspecified**
[there wasn't enough convincing evidence to say that the third offender was the accused**] Doubts/not enough evidence**
[film quality was not good enough to make that accusation on its own**] footage quality**
[fingerprints (which might have been placed several hours after the crime)**] cigarettes (explanation of prints)**
[Possibly inaccurate method of identification**] Anthropometric facial comparison - unreliable**

The final categories of responses are listed below, in table 7. The following sections have also been split into arguments that were made containing the anthropometric facial comparison evidence, and the remaining evidence (which is discussed first).

Table 7. Argument coding categories

Guilty verdict responses	**Not-guilty verdict responses**
Cigarette prints	Cigarettes (explanation of prints)
Cigarette brand	Eyewitness couldn't ID
Machete	Co-accused refusing to ID
Gum	Previous customer of co-accused
Eyewitness ID	Machete (explanation)

Defendant knew co-accused	Machete wasn't found or tested
Defendant's character	Could be innocent
Defendant lack of alibi	Set up
Defendant's story	Consequences for defendant
Defendant is only suspect	Possible 3rd man
Strength "beyond reasonable doubt"	Story/explanation
Prosecution had forensic science	Doubts
Defence only had circumstantial evidence	No "hard evidence"
AFC (unspecified)	Bias could have affected comparison
Qualifications	Comparison was unreliable
Method	Footage quality
Confidence	No statistical tests
Match from co-accused	No peer review
Combination with other evidence	Confirmation bias
	No 2nd expert
	Confusing language
	Risk of error

Graphs have been included showing the proportion of participants who made an argument in each category when arguing for a particular verdict choice (e.g. of those in the control condition, 73.17% mentioned the prints found on the cigarettes when discussing why a guilty verdict could have been the correct choice whether they had selected it or not, and 63.41% mentioned the explanation of cigarettes while arguing that the not guilty verdict could have been correct). Examples and the coding guidelines (including any exclusion criteria can be found in Appendix B.

Note: when providing examples I have replicated the responses as they were written into Qualtrics unless further context explanation was required.

3.5.3 Non-facial comparison responses

Guilty verdict justification

Participants showed a wide range of arguments concerning the non-anthropometric facial comparison evidence. When discussing reasons why the defendant could have been guilty the most discussed aspects of the trial were the fingerprints found on the retrieved stolen cigarette packets, and the machete which a witness testified to lending the defendant over a period of time including at the time of the robbery. As can be seen in Figure 5, there was little difference between the conditions in discussion of the cigarette brand being a favourite of the defendant's, eyewitness ID, that the defendant knew the co-accused, his character, lack of alibi, or of the story provided by the defence to explain the defendant's behaviour. However, with the exception of the Eyewitness ID, the proportion of participants mentioning the non-anthropometric facial comparison evidence was less than the other conditions. Slightly more participants in the control condition mentioned the cigarette prints.

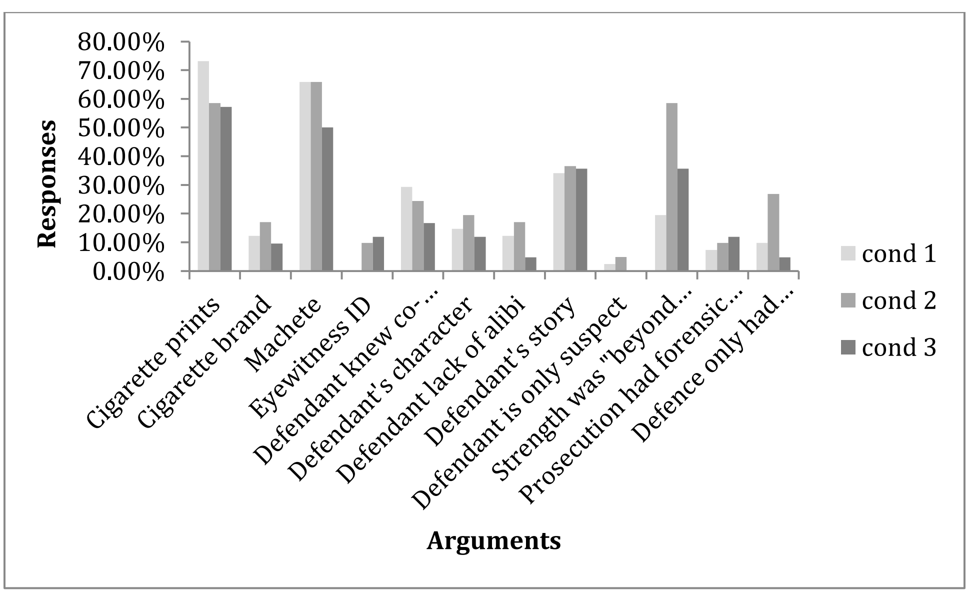


Figure 5. Non-facial comparison guilty verdict responses separated by condition

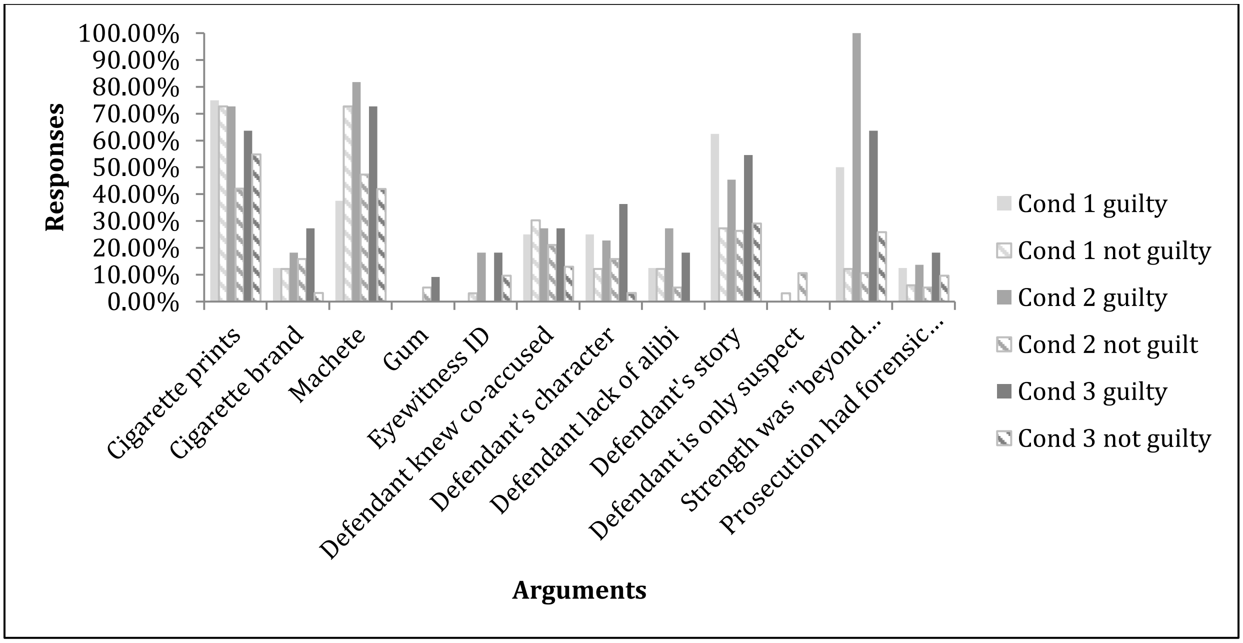


Figure 6. Non-facial comparison guilty verdict responses separated by condition and verdict

Responses show a lot of variation between conditions and verdict choice, although participants who gave guilty verdicts were more likely to make these references within each condition, which is not surprising. One pattern that does emerge is that participants in the second condition who gave a not guilty verdict consistently make fewer references to the other circumstantial evidence in the trial (indicating that their responses discussed other elements of the trial) compared to other condition-verdict combinations. Overall, however, it is clear that the anthropometric facial comparison evidence did not overwhelm them, as the majority of responses mentioned the other evidence in reply to one of the questions.

The fingerprints found on the cigarette received the most mentions. Most participants made comments calling the defence's justification for the location of the prints unlikely, whereas other jurors argued that the explanation was possible, or at least added to their doubt. Others extrapolated from the story given by the prosecution in the opening statement to reason why the cigarettes would have prints on them if Holt *did* commit the crime, as this was never explicitly mentioned in the trial transcript:

"*The accused prints were found on cigarettes in possession of the two co-accused*…" (ID sOnX)

"*The fingerprints on the packet could have been put there by Mr Holt while putting the packets in bags during the robbery*" (ID a15m)

"*if involved in this he would have been more careful – NOT to leave prints*" (ID WBXn)

This was closely followed by the possession of the machete at the time the crime took place. Many found that being in possession of a machete was suspicious, whereas others were disappointed that the prosecution were unable to match the specific machete to the one used in the crime, despite this being an unreasonable (and impossible) task:

"*It seems unlikely that he would just have happened to have borrowed the machete during the time that the crime occurred*" (ID J9A1)

"…he had previously used the machete for non-nefarious means" (ID wJDb)

"The [accused] had access to a machete, and a machete was used to perpetrate the robbery. The Crown would have had a better case if the defendant's machete was able to be connected to the crime. The fact that he had a machete that he borrowed regularly does not immediately prove it is the same machete used in the robbery. Even if it were, there needs to be conclusive proof that the defendant used the machete at the robbery in order to convict him" (ID EXPf)

"*Not proven it was the same machete used in the robbery"* (ID ekqR).

Those giving a guilty verdict were more likely to mention the story presented by the defence as being unlikely or suspicious, but there were fewer mentions of this kind in the prosecution-expert only condition.

"*Story was misleading and didn't add up*" (ID ykfS)

"*The time between the robbery and arrest was about 3 hours – which seems to be very fast when considering that the defendant just happened to call and '[ask]' the other men if they had cigarettes/marijuana to sell during this period, rather than the other way around."* (ID V28R)

Participants giving a guilty verdict were also more likely to mention that the combined weight of evidence was "beyond reasonable doubt". Substantially more participants in the prosecution-only condition justified the choice to pick a guilty verdict by saying that the combined strength of the evidence produced by the prosecution was very high or beyond reasonable doubt (as can be seen in Figure 2, of those who picked guilty, every participant made a comment about the strength of the prosecution's case or it being beyond doubt):

"*I didn't choose not guilty because I believe the evidence proved enough for a conviction" (*ID 625P).

Only a small number of participants mentioned that they felt the prosecution's case was strong(er) because it had forensic evidence, and there was no substantial difference between those giving a guilty verdict and those giving a not guilty verdict. However, participants were more likely to argue that the defence only had circumstantial evidence:

"*…largely due to the lack of evidence given in the defence of Mr Holt, the circumstantial defence and evidence of his fingerprints… the lack of video footage put forward to show Mr Holt attending and leaving the meeting with the co-accused in the alley way – [give] that is was in a busy commercial area and the presence of surveillance cameras EVERYWHERE"* (ID HAwM).

"*I felt the evidence from the prosecution strongly suggest that the accused was guilty and that the defence's case was weak. The defence's case was based solely on testimony from the accused wheres the prosecution provided sound scientific evidence supporting the guilt of the accused*" (ID Fee9)

Participants in the second condition made more comments about the relative "weakness" of the defence's case. Mainly, these involved criticising the lack of "hard evidence" supporting innocence, and the lack of alternate forensic tests or analyses conducted by the defence, despite many of these tests or forms of evidence being improbable. Although the defence case in the third condition didn't present any additional evidence or analyses this type of response was less likely to occur, with only 4.8% of participants making these comments as opposed to 26.8% in the second condition (and 9.8% with this type in the control condition). Many of these comments also suggested that the participant had gone in with the mind-set of "guilty until proven innocent", or that once the prosecution had produced their forensic evidence, it was up to the defence to provide alternative evidence or analyses that disproved it:

"*I don't think there's any way to really know if he's guilty or not. He could be telling the truth but there was no strong evidence to suggest that he was not guilty."* (ID oE4F)

"*…the experts only talked against the [prosecutor's] experts, and didn't have their own evidence."* (ID edHT).

"*While the defense was plausible it wasn't supported by evidence*" (ID k7XW).

Not-guilty verdict justification

There was more variety in the arguments concerning reasons why the defendant could be innocent. While there was not much difference between conditions mentioning how the prints on the cigarettes could be indicative of guilt, the number of participants mentioning the defence's explanation of the prints diminishes sharply from the first condition to the second and third (Figure 5). The pattern is similar with regards to mentions of the fact that the owner of the Quick-Mart couldn't single out the defendant in the line-up. Few participants mentioned the co-accused's refusal to ID the defendant as the third man, or that the defendant had a previous relationship with them, in arguments for why he could be innocent (substantially less than the proportion of arguments saying his previous relationship with them could justify a guilty verdict).

Fewer participants in the second condition mentioned the explanation for the machete, that no attempt was made to locate the machete, or that the idea of the defendant being innocent weighed on their decision. A very small number of total participants suggested that the defendant could have been set-up. Equal proportions of participants in the first and second condition discussed the "third man in the shadows" or the story told by the defendant, with fewer participants in the third condition making the same arguments. Almost equal proportions of participants in the first and third conditions mentioned doubts about the guilt of the defendant, and that there was no "hard evidence". Participants in the second condition made proportionally fewer of these comments.

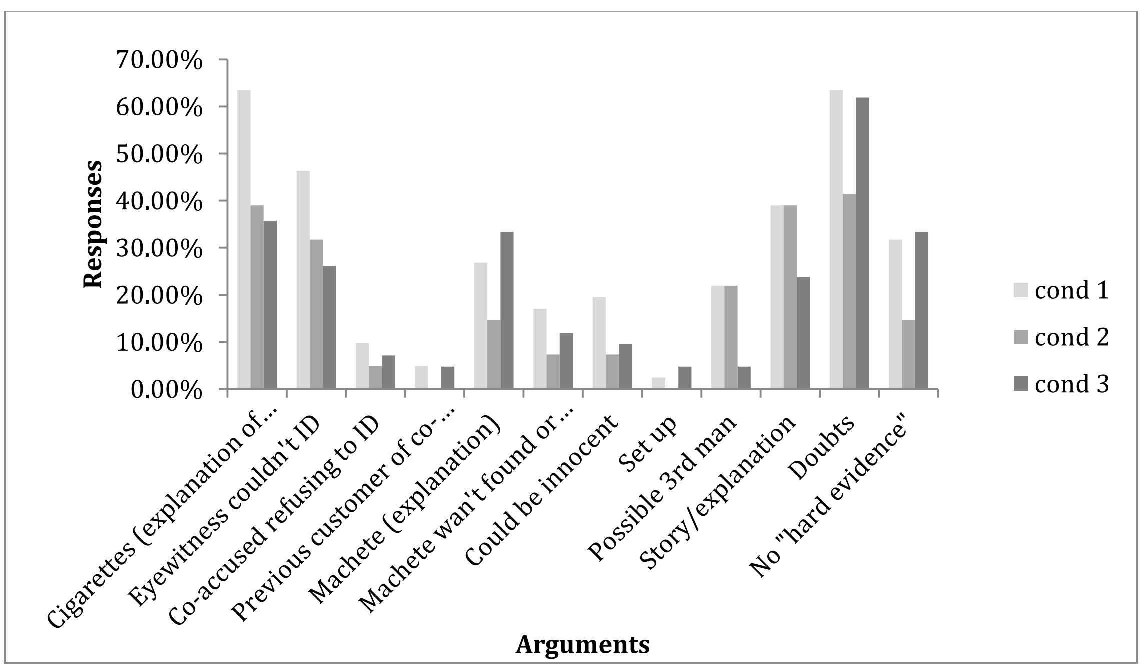


Figure 7. Non-facial comparison not guilty verdict responses separated by condition

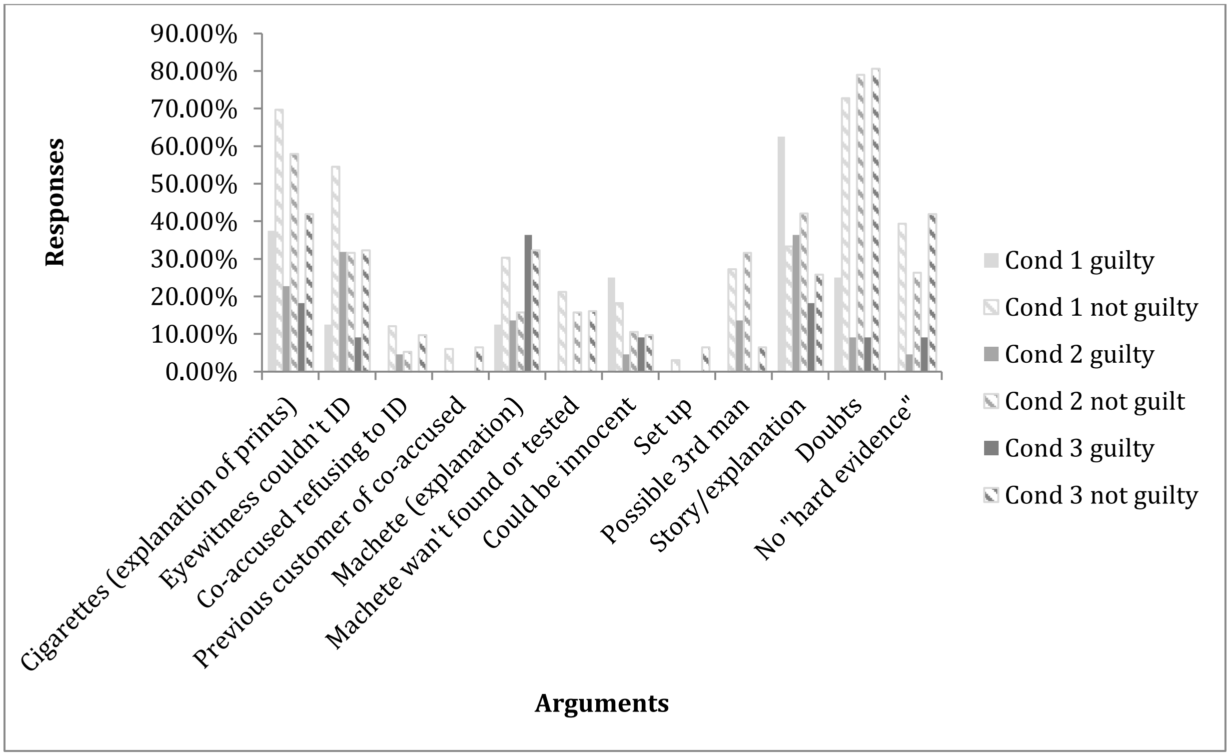


Figure 8. Non-facial comparison not guilty verdict responses separated by condition and by verdict

Participants were almost as likely to mention the defence's explanation for the prints, as they were to mention that the prints linked the defendant to the crime. Fewer jurors showed an understanding of the mobility of the cigarette packets as evidence:

"*The reason for Holts fingerprints being on the stolen goods does give a viable excuse for this evidence against him*" (ID qlGG).

"*The only substantial piece of evidence was the presence of Holt's fingerprints on the cigarette packets. The defense offered a plausible alternative explanation for this*." (ID tZ5f)

"*The finger prints on the cigarette cartons could plausibly have occurred in the car later that day*" (ID sctb)

While receiving less attention than the fingerprints, many participants in the first condition also found the failure of a certain eyewitness ID to be an argument against giving a guilty verdict.

"*…but he would have been IN THE SHOP being the ONE WHO WAS ROBBED, with the machete guy looking RIGHT AT HIM when he made the request… how would someone remember that but not the face of the accused*" (ID wugF)

"*Store keeper couldn't pick him out*" (ID ykfS)

However, in the conditions where participants had read through the prosecution expert testimony, the failure of the victim to ID the defendant from the line-up was rarely seen as a strong point in favour of innocence, and most mentions of the eyewitness were made in response to questions asking if there were other factors.

Just as participants were likely to mention evidence being beyond reasonable doubt as a reason they, or in a very few cases - others, would consider the defendant to be guilty, many participants mentioned that they still had doubts.

"*The accused didn't deny being with the other two people, but rather gave reason as to why he was there, and whether or not he is telling the truth I don't know but he could be and so there is reasonable doubt in my mind*" (ID 0UTf)

While there were fewer overall arguments about the lack of "hard" evidence decisively linking the defendant to the crime, for many participants the lack of a conclusive piece of evidence formed the basis of their reason to give a not guilty verdict.

"…*nothing conclusive to suggest that Holt was any where near the crime scene" (*ID RYud)
Anthropometric facial comparison responses

Guilty verdict justification

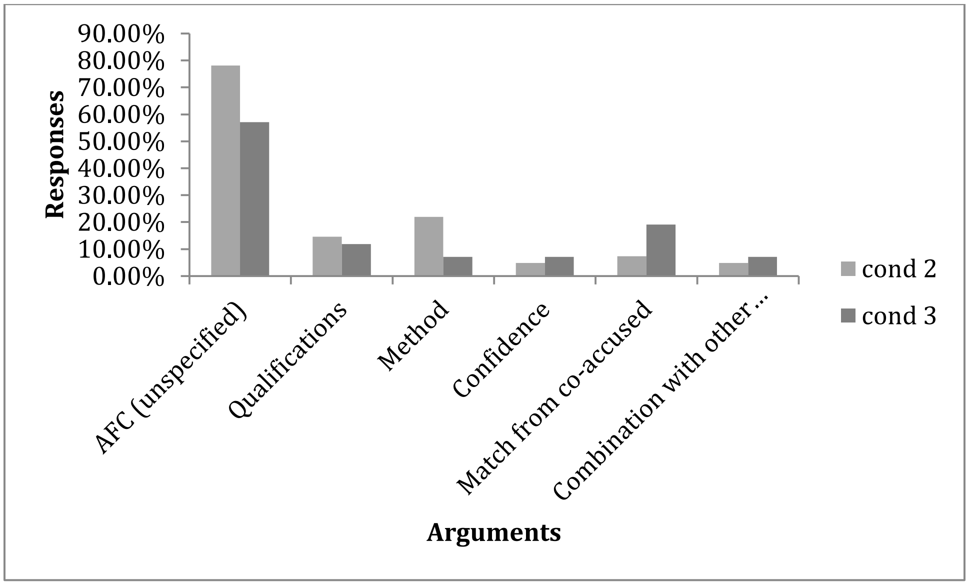


Figure 9. Anthropometric facial comparison guilty verdict responses separated by condition

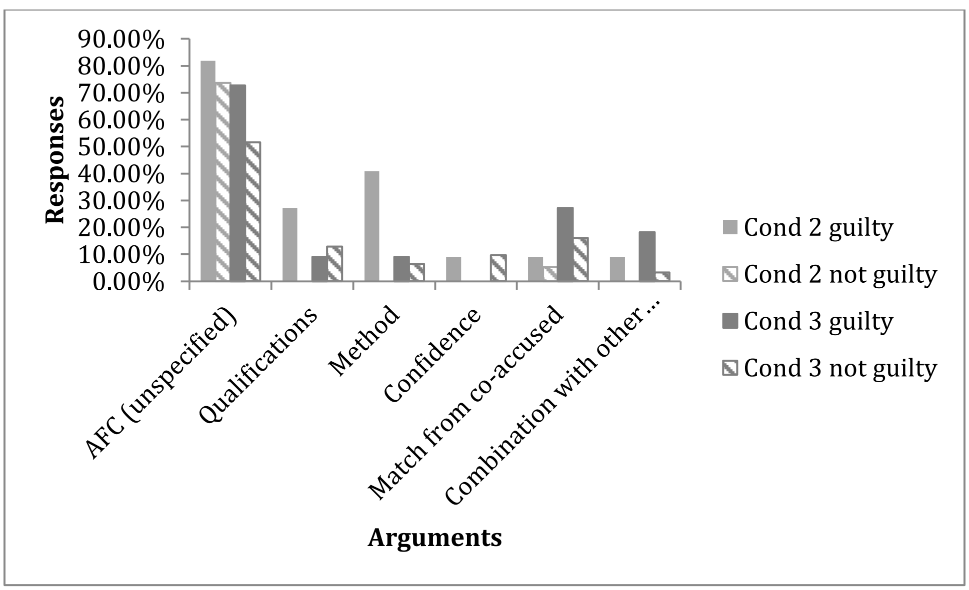


Figure 10. Anthropometric facial comparison guilty verdict responses separated by condition and verdict

AFC (unspecified)

A reference to the anthropometric facial comparison evidence was deemed “unspecified” when there was no discussion of any particular aspect of the evidence (method, reliability, etc.). Unsurprisingly, these were very common. As can be seen in Graph 5, these type of responses occurred more frequently in the second condition than in the third condition.

"The forensic expert matched photos of the accused with the cctv." (ID EKEr)

"Likeness in CCTV footage." (ID BWej).

Substantially fewer participants made references to the reliability of the method, confidence in findings, or the fact that the expert mentioned they had used the same techniques to match the co-accused to their images from the footage. Qualifications only received a small number of mentions in the responses, but the pattern of references was different between the two conditions. While the third condition showed little difference between the different verdicts, no respondents in the second condition made a reference to the qualifications of the expert in any answer: all the references were made by those who chose guilty.

“*… the reliability of Dr. Hamilton as she was an academic”* (ID rK8h)

“*Dr. Asley [Hamilton] completed biological science at University of Melbourne and completed a PhD and had focused on anthropometry. She seems confident with the methods used and was very thorough – seemed to know what she was doing*” (ID 7UcC)

Several participants also inferred from the testimony that a computer program had been used to take the measurements between facial features.

“…expert using a compter program found similarities to support for positive identification”

Overall, while the majority of participants in the prosecution-only and prosecution and opposing expert conditions made some mention of the anthropometric facial comparison evidence, very few provided a reason as to *why* this evidence was reliable or trustworthy.

Not-guilty verdict justification

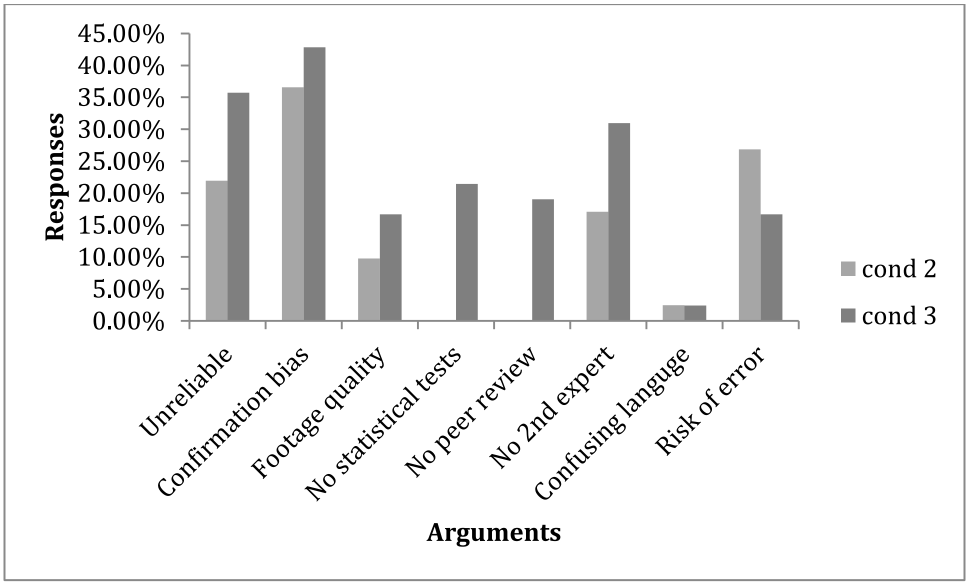


Figure 11. Anthropometric facial comparison not guilty-verdict responses separated by condition

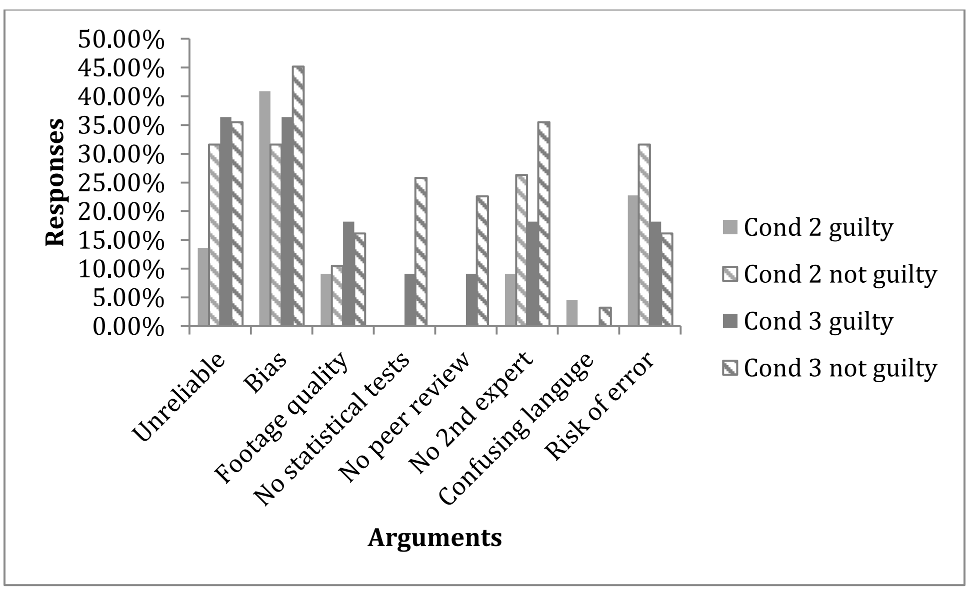


Figure 12: Anthropometric facial comparison not guilty verdict responses separated by condition and verdict

Comparison between conditions and verdicts shows more difference in arguments made supporting a not-guilty verdict. Participants who saw the opposing expert in addition to the prosecution expert were more likely to mention that the anthropometric facial comparison evidence was unreliable, with very few participants who gave a guilty verdict acknowledging there had been arguments calling the evidence untrustworthy. However, the pattern is very different when looking at mentions of the evidence being biased: more participants who gave a guilty verdict and who saw only the guilty verdict mentioned bias than those giving a not guilty verdict.

During the cross examination, the defence lawyer briefly asks about the number of "set comparable features" used in the analyses, mentions the low quality of CCTV footage and asks how they overcome difficulties relating to poor quality footage, and also enquires if there was a second expert (the main transcript is in Appendix A, with the amended sections used in Study 2 in Appendix B). The answers given by the expert are brief and refer back to their experience and the protocol they have developed for making comparisons. The anthropometric facial comparison expert also admits that there is "always some risk of error, of a false positive when there is no true match". The opposing expert goes into a lot more detail regarding these arguments, as well as discussing the lack of statistical tests and peer review. Despite the brevity of the arguments made during cross examination, it is clear that participants in the prosecution-expert only condition found them compelling.

"*…because the expert stated that she made up her process of photo ID so she can't say that she isn't biased*" (JFAt)

"… more than one expert should have been consulted, and they of course should have made independent examinations" (T5uV)

Both the lack of statistical tests and peer review was directly mentioned by around 1 in 5 participants, although it would likely have played a part in the arguments that the method was unreliable or potentially biased. However, it is clear that for a substantial number of participants in the third condition, the absence of validation studies or a standardised method that had been approved by experts in the field was seen as concerning.

"I would have explained that as the expert witness didn't provide any sort of empirical evidence to show that she could accurately identify the defendant to any reasonable degree… it would be unreasonable to say that we could find him guilty…" (Participant ID XkxW).

"T*he case seemed to rely fairly heavily on the facial mapping techniques. Once these were undermined by a second expert, that only really left the more circumstantial evidence to be taken into account…literally no studies have investigated how reliable they are, then how can we be convinced that they're in any way useful*?" (ID aYeF)

A number of participants, like the one quoted above, also seemed to believe that if you accepted the arguments made by the opposing expert it should "cancel out" the identification evidence (even if it is not possible for researchers to know how effectively jurors are able to disregard evidence).

The examination of the written responses did reveal some other causes for concern regarding juror comprehension of the evidence. The anthropometric facial comparison expert admits that there is a chance an error could occur (as is true of all forensic evidence), and the opposing expert argues that the lack of an error rate means the technique is not established as reliable. A small, but still prominently sized, number of references were made that indicated the potential for error had led to the jurors disregarding the evidence as having no probative value.

"*The psychologist stated that photo comparison is not necessarily as accurate as DNA evidence is*." (5AhO)

"*Couldn't prove 100% he was guilty. Doctor gave some compelling evidence but couldn't be 100% accurate*." (ID ykfs)

Finally, several participants who gave guilty and not-guilty verdicts indicated disappointment with the lack of conclusive forensic evidence either establishing that the defendant had been at the crime, or supporting his exculpatory version of events. High expectations and exaggerated beliefs in the abilities of forensic evidence are among the concerns represented by the CSI effect (Cole & Dioso-Villa, 2006; Maeder & Corbett, 2015). For some participants, the responses indicate that in the absence of physical forensic evidence they were unwilling to convict (and in one case, unwilling to acquit).

"*There was no forensic evidence at the crime scene suggesting that the defendant was there*" (ID 4CNH)

"*…there is no forensic evidence for the 3rd man the accused claims was present at the transaction he made*" (LOSg).

The final question asked jurors to consider how another juror would consider the evidence. The researchers who developed the initial questionnaire included this question because being able to take another's point of view is an important skill when considering arguments (Kuhn et al., 1994; Weinstock, 2011). A recurring theme in the answers to this question among participants who gave a not guilty verdict, was that participants who accepted the facial comparison evidence would consider this enough to give a guilty verdict.

"*They would most probably be influenced by the "legitimacy" of an expert opinion and forensic evidence"* (ID 2Nkb).

3.6 Discussion

Analysis of responses to the short answer questions confirm previous findings that mock jurors made substantial errors in their reasoning concerning forensic evidence. Nevertheless, responses also showed that they were not overwhelmed by the identification evidence, and many participants understood and engaged with the criticisms mentioned during the cross examination as well as those put forward by the opposing expert.

Due to the particular questions asked, it is difficult to assess how participants who viewed the anthropometric facial comparison evidence positively (even after hearing the criticisms) came to those estimations. Most of the guilty-verdict mentions of the facial comparison evidence were unspecified, and seemed to rely on the reader also having an awareness and understanding that the evidence was reliable. Future qualitative analyses should consider asking direct questions regarding the forensic evidence encouraging participants to talk about their perception of the evidence, how it was analysed, and the expert. This may reveal more clearly how participants are using the criticisms, how this ties in with their previous knowledge and perceptions, as well as whether they had difficulty understanding the critiques (or considered them products of biased defence expert, a “gun for hire”).

Many participants voiced their dissatisfaction with the limited forensic tests mentioned in the trial, which seemed to indicate an unrealistic faith or belief in the powers of forensic science, or what is capable of being adduced in a criminal trial. These arguments were all assimilated within the coding scheme rather than forming a separate category. The main repeated complaints were that no attempt was made to “match” the machete owned by the defendant’s friend to the machete seen in the footage and that no attempt was made to locate a blue shirt owned by the defendant. Others mentioned that they found it disappointing that there were no fingerprints or DNA evidence found in the burgled store to show that the defendant had been in there, and that security camera footage was not used to confirm the meeting of the defendant with the other two accused of the burglary to support his alibi. Some of these ideas were written in the notes but did not make it into written responses to the questions.

Overall, the responses show that many participants understood and appreciated the criticisms offered by the opposing expert or during cross examination, even if they still gave a guilty verdict. However, a substantial proportion of participants presented attitudes consistent with those mentioned in the CSI effect, commenting on disappointment with forensic tests deemed “missing”, asking for impossible analyses, or considering that because there was mention of an error rate, the analyses must be completely untrustworthy and should be disregarded.

3.7 Summary

This was the second empirical study to investigate **RQ1***, can opposing expert testimony inform jurors about unreliable forensic science?,* confirming that opposing expert testimony is an effective way of helping jurors to more accurately assess unreliable forensic identification evidence. Using a different sample and an online format to encourage a more representative sample strengthened support for the indirect effect of opposing expert testimony on verdict, through ratings of methodological reliability. Analysis of the written responses to questions asking participants to defend their verdicts shows a range of understanding and engagement with the criticisms given by the opposing expert. Although the responses were brief, content analysis found that most participants were not overwhelmed by the expert testimony but their exposure to the opposing expert encouraged consideration of other aspects of the trial when forming their verdicts. While opposing expert testimony was a successful educational measure, the arguments made by that expert could potentially be given by the lawyer for the defence. Analysis of the responses shows that even participants in the prosecution expert only condition made references to some of the main flaws with anthropometric facial comparison, which were given by the defence lawyer during cross-examination. This suggests that it is not only the presentation of an opposing expert that can cause more effective evaluation of testimony – but also the use of an effective cross-examination of an expert that may trigger more in-depth processing of claims. Therefore, the next study further investigated mock juror perception of unreliable forensic evidence, and whether cross examination would be an effective educational method in direct comparison to an opposing expert.

In investigating **RQ4,** study two has found support for the role of pre-trial bias towards forensic evidence in influencing ratings of methodological reliability, and through those ratings, verdict choice. Previous scales of juror characteristics have focussed on more general attitudes about legal decision making. The FEEBS has received limited previous testing but was selected because it was conceptually the most appropriate measure of layperson perception of forensic evidence and was based on the concerns captured by literature on the CSI Effect. This study found a moderate size relationship between scores on the pro-prosecution subscale and ratings of reliability, showing that participants who hold inappropriate levels of faith and exaggerated expectations of the value of forensic evidence gave higher ratings to the unreliable anthropometric facial comparison evidence. The next step will be to investigate this relationship using a different sample, and using a different type of crime and/or field of forensic evidence.

4 STUDY 3

Study Three was the final empirical study conducted as part of this thesis, and was designed to investigate whether the criticisms given by the opposing expert would have a similar educational effect on jurors if they were presented by a defence lawyer as part of cross-examination.

A review of previous research looking into the effectiveness of opposing expert testimony found that there were areas that had been under-researched. However, the results of Study One and Two found that an expert appearing for the defence was capable of raising awareness of the limitations of an unreliable science. Ratings of the reliability of the method used by experts who presented unreliable evidence were substantially lower when participants were exposed to opposing expert testimony compared to participants who were not, showing that opposing experts may help minimise errors in decision making concerning complex forensic scientific evidence.

Comparatively more studies have looked at the two alternative methods of educating jurors: judicial instructions, and cross-examination. Judicial instructions have consistently shown to be ineffective at helping jurors to more accurately assess evidence (Cutler et al., 1989; Dartnall & Goodman-Delahunty, 2006; Eastwood & Caldwell, 2015). As discussed in Chapter One, research involving real lawyers and other legal personnel has shown that many struggle with more sophisticated scientific research, and may lack the awareness of methodology needed to present a critique of unreliable forensic evidence. Legal commentators have pointed out that this frequently leads to a reliance on criticisms of an expert's qualifications, educational history, and prior appearances in court. However, with better education in scientific principles, there is the possibility that the difficulty in presenting arguments that could help jurors would be resolved.

This could allow more effective cross-examination benefiting the decision making of jurors. The goal of Study Three was to investigate this possibility. The chapter opens with a discussion of why cross-examination would be advantageous above opposing expert testimony, before presenting a review of previous studies looking at the effect on juror decision making. As such, the goal was to answer **RQ2:** *Can cross-examination that focuses on the methodology behind identification evidence, rather than the qualifications and expert of the expert, sensitize jurors to unreliable forensic science?* A comparison of the results from the previous two studies and Study 3 will also allow for an examination of **RQ3**: *Is the effect of cross-examination on juror decisions similar to, or different, from that of opposing experts?*

Bias towards forensic evidence remained in the study as the focal individual difference. Additionally, as the first two studies had examined the same forensic science which limited their generalisability, the final study contained three versions of the crime scenario, featuring either anthropometric facial comparison, fingerprint or voice identification evidence allegedly linking the defendant to the crime.

4.1 Cross-Examination and Jury Decision Making

The first two studies show support for using opposing expert testimony in criminal trials to help educate jurors, but there are several barriers for including opposing experts in practice. Legal commentators have noted that it is harder to locate willing or experienced forensic experts who will appear for the defence (Cordner, 2015), and that most defendants will struggle to hire experts to appear for them (Roberts, 2015). Odgers (2015) argues that limited funding reduces the incentive for lawyers to spend time preparing effective challenges, and in some cases may restrict access to defence experts who could prepare reports concerning evidence. A recent report by the House of Commons in the United Kingdom found that the abilities and resources of prosecution and defence experts were inconsistent, with defence experts rarely having access to samples in their original condition, and having much less time to perform an analysis (2013). Forensic experts may have better awareness of the general acceptance of techniques and the criticisms that are made of different methodologies, but for many defendants they may be unavailable. In practice, defence teams rarely call rebuttal experts (Goodman-Delahunty & Wakabayashi, 2012). Even finding qualified defence experts may be difficult (Giannelli, 2007).

While it is hoped that the first two studies included in this thesis will encourage more research into the abilities of opposing experts to educate jurors, and that this will lead to more trials involving opposing experts, there is also a need to examine the usefulness of alternative gatekeeping methods to ensure all jurors are provided with the information that will allow them to accurately assess the reliability of scientific evidence.

4.1.1 Forensic science evaluation by lawyers

Defence lawyers may have difficulty presenting a defence that adequately represents the problems surrounding identification evidence (Kelly, 1998). A survey of 203 Australian judges and magistrates found that just over a third of responders in each category reported they “often” encountered a failure of the lawyer performing the examination-in-chief to pose appropriate questions, and two thirds said they saw it “occasionally” (Freckelton, Reddy, & Selby, 2001). When asked about the most serious problems with expert evidence, the most frequently given response was expert bias (29.59% of responses), however poor cross-examination was the next most serious response (tied with difficult language, both received 19.39%).

Most lawyers avoid arguing with experts about technical aspects of scientific issues (Bowers, 2014; Edmond et al., 2009), are unlikely to be familiar with sociological scientific theory or to present such in court, and most focus on attacking a particular piece of evidence (Edmond & Mercer, 1999). Many defence lawyers focus on chain of custody and conflicts of interest (Findley, 2008; Garrett & Neufeld, 2009), or personal trial experience of the expert (Edmond & Roberts, 2011). Even when cross-examination focuses on reducing the credibility of the expert based on methodology, strategy typically focuses on the mishandling of evidence and not on inherent technical problems with identification sciences (Edmond, 2011a; Edmond et al., 2009). Edmond and Roberts have called these attacks “low hanging fruit”, as they are not good substitutes for criticisms that focus on a lack of empirical investigation or reliability, and do not address concerns that commentators voice about junk science, even though they may be more easily understood by jurors.

As cross-examination is an established part of a trial, more research has focused on the ability to use it as a gatekeeping method. Jurors will judge lawyers on the strength of their cross-examination (Diamond, Casper, Heiert, & Marshall, 1996). Even the *Daubert* Justices argued that "vigorous cross-examination… [was a] traditional and appropriate means of attacking shaky but admissible evidence (*Daubert v. Merrell Dow Pharmaceuticals,* 1993, p. 595). Strong cross-examination that addresses flaws in expert evidence may alert jurors to indefensible claims, although this will not necessarily lead to a change in jurors' verdicts (Diamond et al., 1996).

4.2 Empirical Studies of Cross-Examination

A considerable amount of empirical research has looked at the impact of cross-examination on juror decision making. However, similar to the research on opposing expert testimony, disparate use of stimulus materials and research focus has led to mixed findings. However, while many legal commentators argue that cross-examination is ineffective at educating jurors, this is primarily because most defence lawyers lack the background, training, or resources to present adequate arguments against forensic evidence (as introduced in Chapter One). The following section will review the previous, recent, research looking at cross-examination, with a focus on studies looking at unreliable evidence.

Research by Kraus and Sales (2001) compared responses to actuarial (which use empirically verified risk factors) and clinical assessments of dangerousness, which are frequently inaccurate (Monahan & Steadman, 1994; Otto, 1992) and prone to misuse by the jury (see Marquart, Ekland-Olson, & Sorensen, 1989). They tested four different adversarial procedures: ineffective cross-examination (attacking the credibility of the expert (by arguing the expert was not a medical doctor and was getting paid); effective cross-examination that focused on the content (e.g. that dangerousness predictions are often wrong, a lack of training in dangerousness predictions, and that actuarial techniques have shown more accuracy than clinical instruments); an opposing expert who disagreed with the initial expert; or an opposing expert who disagreed and focused on the content. Mock jurors' ratings of dangerousness of the defendant, expert credibility, influence, and level of scientific knowledge in all four conditions were negatively affected after cross-examination, with the ineffective cross-examination having the least effect. However, cross-examination had a comparatively reduced negative impact on the less scientific clinical testimony, when the opposite effect would have been expected if cross-examination had educated the mock jurors about the problems with dangerousness research.

Kraus and Lee (2003) sought to replicate the findings of the previous study but added a deliberations section, as it was theorised to offer a "forum for justification of viewpoints, multiple and divergent comprehension abilities, and additional time" (p. 117). The effect of clinical testimony on mock jurors' dangerousness ratings was still significant even after cross-examination and deliberation, although it was highest after ineffective cross-examination. However, in this study the negative effect of cross-examination was strongest when effective arguments had been given – as would be expected if cross-examination had been able to educate jurors about the unreliability of the testimony.

In an attempt to better understand the contrasting findings of the previous two studies, and why jurors could be more persuaded by less scientific testimony, Krauss et al. (2004), used Cognitive-Experiential Self-Theory (CEST; Epstein, 1994; Epstein & Pacini, 1999), to induce participants to use rational, rather than experiential decision making. Gawronski and Creighton (2013) argue that as a generalized dual process theory, CEST is more commonly used for interpreting the results of experiments (as it was used by Krauss et al. (2004), rather than making testable predictions, as it does not concern specific behaviours such as persuasion, as is the case with the ELM (Cacioppo & Petty, 1983), although the distinctions between the two modes of processing are similar. In CEST, rational decision making is more effortful, deliberative, analytic, and intentional. During experiential decision making, however, decisions are made automatically, based on emotions, or "gut instinct", as well as heuristics (Epstein, 1994; Epstein & Pacini, 1999). Epstein and Pacini (1999) argue that the experiential mode is the default way of processing, as it is efficient and less effortful, but if the rational mode of processing is used, reliance on using the limited information immediately available and heuristics, is reduced. The researchers argued that

CEST "would predict that unless participants are induced to adopt the more effortful rational mode that requires logic and evidence to justify conclusions, clinical expert testimony should be particularly persuasive" (p. 808). The predictions were supported: when a rational mode of processing was induced in participants the effect lessened, and rational processing had more impact on dangerousness ratings. However, this effect only occurred *after* cross-examination (modelled on the "effective" conditions from previous studies), indicating that that "rational" mock jurors were better able to utilise the information and criticisms provided by the cross-examination. Overall, the series of studies shows that effective cross-examination may help jurors to recognise flaws with unreliable evidence if jurors are motivated to think effortfully about the arguments, but, that without help, many may rely on heuristics to make judgments about unreliable expert evidence. However, dangerousness ratings, while unreliable, are used differently from identification evidence, and may be perceived differently to identification evidence.

In the first of a series of studies by Lieberman et al. (2008), which were introduced in Chapter One, undergraduates read through descriptions of a trial featuring DNA evidence, fingerprint, hair fibre, eyewitness, or victim testimony. The evidence was either inculpatory or exculpatory, and the crime was either a murder or a rape. As expected, DNA was highly valued by participants as both inculpatory and exculpatory evidence. Fingerprint and hair comparison were also highly valued when it supported guilt, but was less powerful as evidence of innocence, with eyewitness testimony the least valued type of evidence. Physical evidence has been found to be more highly valued than other types of evidence in previous studies: Golding, Stewart, Yozwiak, Djadali, and Sanchez (2000) found that DNA evidence lead to more guilty verdicts than victim testimony in a case of child sexual abuse, and Skolnick and Shaw (2001) found that comparisons of physical evidence from stabbing wounds (from a unique knife shape) were more influential than eyewitness evidence. There was a stronger, positive effect on evaluations of the level of responsibility of the defendant when strong physical evidence was presented in comparison to strong eyewitness testimony, and evaluations of the defendant (trustworthiness, intelligence, warmth, attractiveness) depended on the physical evidence, but not eyewitness evidence.

In the third study of their series, Lieberman et al. (2008) included a more complete discussion of the potential weakness of DNA evidence via cross-examination, and the evidence either came from a laboratory where proficiency testing was used to test the reliability of the findings, or from an unaccredited police department. Cross-examination of the expert was either expert-focused and discussed the credibility of the expert (his experience, academic record from college, and whether he was paid) but did not introduce any arguments about the DNA testing, or was evidence-focused (raising specific criticisms of the DNA evidence, including the potential for contamination, prior wrongful convictions, and the potential for subjective judgements using the match window in DNA analysis). The participants expressed less confidence in the unreliable laboratory after hearing evidence-focused cross-examination than after hearing expert-focussed examination. There was no support for a main effect of laboratory condition on probability of guilt, and a guilt and confidence composite measure, however there was an interaction effect between cross-examination style and the two levels of laboratory reliability on the two dependent variables. Mock jurors who were in the unreliable laboratory condition, and who heard the evidence-focussed cross-examination, were less likely to convict. The authors argued that overall this supported the idea of using cross-examination to make jurors aware of scientific flaws in expert scientific testimony.

Kovera et al. (1999) investigated whether jurors would use heuristic cues, such as general acceptance and ecological validity, when evaluating the quality of an experiment. The simulation presented to the undergraduate participants concerned a hostile work environment trial, based loosely on *Robinson v. Jacksonville Shipyards* (1991). Participants either viewed expert testimony which featured a single study concerning gender stereotyping and male behaviour, or no expert testimony. In the conditions with expert testimony, the researchers manipulated four elements of the trial: general acceptance of the evidence, ecological validity of the research, construct validity of dependent measures or if the cross-examination was scientifically informed. The cross-examination criticised the sample chosen by the expert and their dissimilarity to the co-workers who had been accused, and the operationalization of sexual harassment used in the research. Mock jurors found the expert's research to be more trustworthy if it had been generally accepted (published in a peer reviewed journal and cited in major psychological textbooks), and ratings of the expert and plaintiff's credibility also increased if the expert had studied participants who resembled the co-workers of the plaintiff (ecological validity). However, the scientifically informed cross-examination did not make jurors more sensitive to the validity of the research.

Salerno and McCauley (2009) also examined the effectiveness of cross-examination in a civil trial with opposing experts who differed in evidence quality. Undergraduate participants were presented with high-quality evidence from an expert appearing for a plaintiff, and a defence expert with low quality evidence (that lacked a control group, inadequate sample size, and had not been peer reviewed or published) with attorneys who either cross-examined the experts or not. However, when cross examination did occur, the defence attorney focussed on the credibility of the plaintiff's expert's (whether it was in his self-interest to find results consistent with previous research), but when the plaintiff's attorney performed cross-examination on the defence expert, they focussed on the scientific (in)validity of the defence expert. Participants who heard cross-examination gave lower credibility ratings to the defence expert, and higher culpability ratings. There was also no evidence of a scepticism effect as ratings of the plaintiff's expert were not significantly affected. Individual differences were also found to matter during a deliberation section, with lower NC jurors most affected by deliberation, although this relationship was not as strong when they had seen cross-examination, which the authors suggested indicates that effective cross-examination could help jurors who are less likely to thoroughly process the information seen in a trial. They contrast their finding with previous work, that was unable to find support for differences in juror response to weak, strong and strong cross-examination with opposing experts (Diamond et al., 1996).

In the most recent tests of cross-examination, Austin and Kovera (2015) presented jurors with an armed robbery trial containing scientifically naïve or informed cross-examination that was conducted by a lawyer or a judge, in addition to having the evidence be either valid or invalid (missing control group). As with Levett and Kovera's other studies mentioned above, and in

Chapter 1, the expert presented the jurors with results from a single trial, but in this study the scientifically informed cross-examination asked about experimentation, why random assignment is important, and why control groups are important. Undergraduates who saw expert testimony were less likely to give a guilty verdict, and those who saw the scientifically-informed cross-examination gave lower ratings of reliability to invalid science, but the mock jurors who heard scientifically naïve cross-examination did not differ in ratings of validity. The judge was seen as more trustworthy than an attorney, but this did not lead to jurors being better able to distinguish valid from invalid science, and the authors argued may have indicated an overall scepticism effect. A second study used members of the public as participants, but this time scientifically informed cross-examination led to more guilty verdicts. However, as there was no interaction between validity and verdict, the authors argued this was more in line with a scepticism effect. As previous research had not found support for an effect, in these two studies Levett and Kovera made their cross-examination more thorough and extensive in explaining scientific methodology, and explained to jurors why control groups were important. The specific testimony explained in detail why a missing control group would affect the results of the study used and essentially invalidate the conclusions.

The studies described above show that although there is some evidence cross-examination may be capable of helping jurors evaluate expert evidence, there is a need for further research. The studies above found that scientifically-informed cross-examination was capable of raising awareness of the unreliability of several different types of expert testimony. However, the only study using forensic evidence was the Lieberman et al. (2008) study, which used DNA profiling. DNA evidence, while not immune to threats to validity (Lieberman et al., 2008; Thompson, 1995), is not representative of the methodological flaws of other forensic fields. Study 3 will therefore be a valuable contribution to the field by examining the use of scientifically-informed cross-examination as a method of educating jurors about the flaws with identification evidence.

The previous studies looking at experimental design have presented mock jurors with a variety of threats to the validity of an expert's testimony, but have found mixed results in terms of the effects on verdicts (or judgments of culpability). Austin and Kovera (2015), in their most recent study, provided participants with thorough and detailed cross-examination, but only found a change in verdict choice when the participants were members of the public, and no relationship between validity and verdict, whereas Lieberman and Sales (1999) found an interaction effect. Other studies indicate that jurors may rely on heuristics when making decisions, unless they are given adequate material or prompting to use more systematic processing (Cooper et al., 1996; Krauss et al., 2004; Krauss & Sales, 2001). One such heuristic is the experience or qualifications of an expert.

4.2.1 Expert Experience and Qualifications

Without scientifically informed questions or statements to educate them, fact finders are left to examine the qualifications, experience, credibility, and even demeanour, of an expert to assess whether evidence is valid (Edmond, 2014). Mentioning expertise, which is often the first thing a juror will hear about an expert (Ivkovic & Hans, 2003), has been found to increase perceptions of credibility (Hurwitz, Miron, & Johnson, 1992). Vidmar and Diamond argue that jury instructions even frequently advise jurors to pay attention to an expert's "qualifications and experience" (2000, p. 1131). Questions on personal experience and authority do not address concerns that commentators voice about junk science, even though qualifications may be a widely accepted marker of scientific reliability.

As Koehler, Schweitzer, Saks, and McQuiston (2016) point out, using these as a measure of expertise or accuracy is "problematic", because more case experience or more impressive credentials are not suitable indicators of performance. They cite Armstrong's (1980) research in saying that expertise and accuracy are likely to be only minimally related. Forensic experts rarely get appropriate, unambiguous feedback, which is necessary for improvement (Einhorn & Hogarth, 1981). Camerer and Johnson (1991) have argued during their research on training and experience, that these are frequently not enough to produce accuracy or genuine proficiency in a field.

Freckelton et al.'s (2001) study of judges and magistrates found that only 9% of respondents considered educational qualifications to be the most persuasive factor in viewing opinion evidence. Another 28.32% said that the most persuasive factor was prior field experience. The number of publications was found relevant by less than 1%. In a review of two Australian studies and six international studies examining judges' ideas about expert evidence, Martire and Bali (2016) found that the second most concerning element of admitting expert evidence to surveyed judges was poor-cross examination of experts and their evidence.

In a rare study specifically examining expert qualifications and jurors, Cooper et al. (1996) investigated the effects of message complexity and juror decision making in a civil trial. An expert's credentials had no effect when the message was of low complexity. However, when the expert testimony had high complexity, the credentials influenced mock jurors, with higher credentials associated with a more persuasive message. However, Cooper and Neuhaus (2000b) found that jurors may be suspicious of experts who testify too often if they hear that the expert will be paid a substantial amount for their appearance.

In a more recent study, Koehler et al. (2016) argued that while a "rational juror might focus on the extent of scientific testing and validation (or lack thereof) of the technique used to generate the forensic evidence, an actual juror might rely on more peripheral cues" (p. 410), such as expert experience or the technological level of the technique used to compare evidence samples. Their research found a strong main effect for examiner experience on a composite measure of evidence strength (Cronbach's α = .94), where the evidence (either fingerprint or bite-mark) given by a more highly experienced examiner was judged as stronger as and more persuasive than, a less experienced examiner (when giving the exact same evidence). Jurors were sensitive to another manipulated variable, whether the evidence had been scientifically tested or not, but this did not affect verdict choice. Koehler et al. (2016) suggested that jurors may not be able to understand or recognise the link between scientific validation and accuracy. They suggested that jurors may trust an individual (experienced) examiner, believing that the examination was accurate when they performed it, perhaps, because as an experienced examiner, they would not use a technique they did not consider to give them an accurate result.

Previous studies on source credibility have also found that sources considered to have more experience are highly persuasive (Chaiken & Maheswaran, 1994a; Hovland & Weiss, 1951; Hovland, Janis, & Kelley, 1953). In their studies using the Heuristic-Systematic Model to investigate the persuasiveness of source cues, Ratneshwar and Chaiken (1991) argue that higher message comprehension leads to more systematic processing of information, which attenuates the impact of source expertise. When comprehension is lower, the reliance on the heuristic of expertise is increased. Scientifically-informed cross-examination, by educating jurors about the methodology behind identification evidence, should increase knowledge about that evidence, and a juror's ability to assess it, therefore decreasing reliance on heuristics, including the qualifications and experience of the expert.

4.3 Cross-Examination and Opposing Expert Testimony

The previous study showed that an opposing expert was able to sensitize jurors to problems with one type of unvalidated science. Compared to those who only heard the testimony of the prosecution expert, those who heard the opposing expert testify were more likely to give lower ratings to the reliability of the facial comparison evidence and the methods used to collect it, and, were more likely to give a not-guilty verdict. This provides support for the idea of opposing experts adding value to criminal trials.

The third study was designed to test for an indirect relationship between cross-examination and verdict choice via ratings of methodological reliability, with scientifically-informed cross-examination leading to lower scores. This would correspond to the relationship between opposing expert testimony and reliability, as same criticisms of identification evidence were made during cross-examination. Based on results from the previous studies, it was predicted that jurors who experienced cross-examination focussing on the lack of valid scientific methodology behind the forensic evidence, as opposed to cross-examination focussing on the qualifications and experience of the forensic expert, would give lower ratings to the methodological reliability of the forensic science featured in the trial. It was also predicted that this would lead to fewer guilty verdicts.

Jurors who experienced scientifically informed cross-examination, as opposed to qualification based cross-examination, would be better informed about the reliability and validity of the science behind the evidence featured in their trial. They should therefore give lower ratings of methodological reliability to the forensic evidence represented in the trial.

4.4 Differences between types of forensic sciences

Surprisingly few studies have measured the reliability ratings of different types of forensic evidence. It is probable that for many laypersons, their knowledge of different forensic methodologies will come from media reports or fictional depictions, which are likely to give inaccurate representations (Cole, 2015; Hallgrimsdottir et al., 2006).

As was discussed in Chapter 1, Lieberman et al. (2008) asked participants to rate the trustworthiness of evidence. This showed that, unsurprisingly, DNA evidence had the highest ratings, with students giving a mean rating of 93.7% and jury samples 94.9%. Fingerprint evidence was also rated very highly (90.1% and 91.4%), as was video tape surveillance pictures (88.1% and 87.3%). A generic "expert testimony by scientists" category was rated as 81.5% and 76.7%, and showed a statistically significant difference between groups of students and jury members.

In their investigation of how crime show viewing affected judgements of forensic evidence, Baskin and Sommers (2010) interviewed 1,201 residents of California over the phone. Participants were asked to rate the reliability of several types of criminal evidence. The researchers found that DNA was considered to be the most reliable evidence (89.5%), followed by fingerprints (78.8%), medical expert testimony (30.3%), police (23.3%) victim (21.2%), and eyewitness (21.2%). When asked whether they would convict during a case involving rape or murder in the absence of scientific evidence, slightly more than half of respondents said they would convict.

Smith, Bull and Holliday (2011) found that participants considered both the mobility and relevance of DNA, fingerprint and shoeprint evidence when the evidence was presented to them without contextual information. Participants did not make strength judgements based on the type of evidence alone, but were able to use information about how the evidence was found (mobility and investigative relevance, e.g. finding DNA on a cigarette at the crime scene, vs. finding DNA on a victim's body).

In a second study looking at elements of the Story Model, participants were given case information in the form of prosecution and defence summaries. When the contextual information was provided there was an increase in probative value given to the evidence, but this effect was the strongest for DNA evidence that had a weak or moderate value. The authors argued that this showed support for the interactionist model (Kassin & Wrightsman, 1988), where juror behaviour is seen to be influenced by the situation (like the specific trial and evidence) and the characteristics of the individual (beliefs, attitudes). However, increases in strength ratings did not lead to an increase in guilty verdicts, and the authors point out that it is possible mock jurors were able to appreciate differences between the strength of moderate and strong evidence, even if this was not shown in the other dependent variables of evidence strength and probability of guilt, but only in the final verdict decision.

While commentators point out that no field of forensic evidence other than DNA have undergone the empirical testing necessary to establish validity, it is clear that in the eyes of laypersons, some types of identification evidence are seen as more trustworthy, and more inculpatory, than others. Studies 1 and 2 looked at anthropomorphic facial comparison evidence, but the research cited above shows that mock jurors may respond differently to different types of forensic evidence. Additionally, while the criticisms given in Chapter 1 apply all types of identification evidence, with the exception of DNA profiling, it is important to understand how jurors will differ in their reactions to hearing those criticisms.

As a comparison to anthropometric facial comparison, Study 3 presented participants with either fingerprint or voice identification comparison evidence. The following sections will give a brief description of the fields as they currently stand and the more exclusive criticisms that have been made of them from legal commentators regarding their reliability and use in court.

4.4.1 Fingerprint Identification

Fingerprint analysis has been used as evidence for nearly a century, leading to an almost indisputable reputation. For many, this is evidence enough that the technique works (Mnookin, 2003), despite even the NAS (2009) pointing out that it suffers from the same lack of empirical analysis as other identification sciences. Thompson and Cole have gone as far as to say that the science behind print identifications is "weak”, with no studies showing that they can correctly identify sources (Hardy, 2007), and no studies showing the frequency of characteristics in a population.

The growing numbers of prints held in databases mean that it is more likely false positives generated by computers will occur (Cole, 2005; Dror & Mnookin, 2010; Dror, Péron, Hind, & Charlton, 2005), which makes it even more important to raise awareness of biases that can occur. Tangen, Thompson, and McCarthy (2011) point out that while CSI gives the appearance of it being an automated system, fingerprint matching is still conducted by a human examiner. There is still a large amount of human judgement involved in latent print analysis.

Research programs investigating the performance of fingerprint experts have only begun recently: (Tangen et al., 2011; Thompson, Tangen, & McCarthy, 2013, 2014) have investigated the proficiency of qualified, court-appearing experts using signal detection paradigms. Experts were not “infallible”, but performances were highly accurate, and errors were more likely to occur when non-matches were declared for two prints that did come from the same source. Similar recent experiments using print examiners from the FBI showed similar findings (Ulery et al., 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2012). A series of studies by Thompson and Tangen (2014), however, found that non-analytic processing could account for a large proportion of variance in decisions, which goes against the explanations given during most expert testimony. It is clear that while these studies are important steps in establishing the accuracy of fingerprint examiners, more controlled studies need to be conducted to establish the limitations of expert judgments.

Two different individuals can have similar ridge characteristics. However, no studies have examined the probability of two people sharing a particular number of characteristics, undermining the probative value of any analysis. There has been no valid statistical model proposed to show variation in prints in the population (Mnookin, 2003). When making a print, the elasticity of the skin, application of pressure, and differences in materials all combine to make sure that even two fingerprints from the same source will not be perfectly similar (Dror & Cole, 2010; Ulery et al., 2011). Ironically, this means that no prints are, in fact, unique, even if they come from the same finger (Hardy, 2007; Vanderkolk, 2009). This means it is common for substantial differences to exist between latent prints that are declared matches (Thompson & Cole, 2007), and this is even before considering prints with smudges, overlapping, or partial prints that are commonly the sources of analysis (Mnookin, 2003; Ulery et al., 2011). Given the common use of fingerprints and the many examiners engaged in the field, it is then perhaps surprising that there have been few studies into error rates (Mnookin, 2003). Despite this, latent fingerprint examiners frequently testify that they have identified the print as belonging to the defendant, or say that they are "100% sure" (Thompson & Cole, 2003).

Experts differ in which features they feel should be compared (Ashbaugh, 1999). Some examiners use point counting but there is no set amount of points that are required, whereas others will use a more holistic approach. There are differences in how examiners from different locations determine matches, in some places an arbitrary number of ridge characteristics has been defined. Australia requires 12 points of comparison, Sweden 7, and Brazil 30, whereas many examiners in the USA do not use a point system (Barkan & Bryjak, 2011). But no matter what standard is set for declaring similarity, there is no empirical basis for suggesting these points are enough (Thompson, Ford, et al., 2003). Other places just leave the decision about when there is sufficient evidence to determine individualisation or not up to the examiner (Ashbaugh, 1999). ACE-V process for fingerprint identification is used in most common law jurisdictions , however Justice Funt, (R v Bornyk 2013 BCSC 1927, para 36), pointed out that the use of process does not mean the results are “accurate or reproducible”, and even the NRC 2009 report said they do not qualify as a validated method.

Worryingly, many experts “just know” when a match has occurred (Mnookin, 2003). Fingerprints only meet general admissibility standards; there is also a lack of any objective standard for gauging whether there is a match (Mnookin, 2003). Despite the long history and reputation of fingerprint comparison, no data has been compiled to show how (un) common ridge characteristics - alone or in combination - are in a population (Zabell, 2005), meaning that analysts have no scientific basis for their testimony (Hardy, 2007). As Thompson and Cole (2007) argue, this also means that there is no empirically set threshold for the number of ridge characteristics needing to match before individualisation can be confidently declared, and there have been no studies examining how many, or what type of, ridge details are needed to convince examiners of a match. Mnookin (2003) has pointed out that it is ironic that DNA evidence was historically challenged by the types of questions that were rarely asked of fingerprints.

4.4.2 Voice Identification

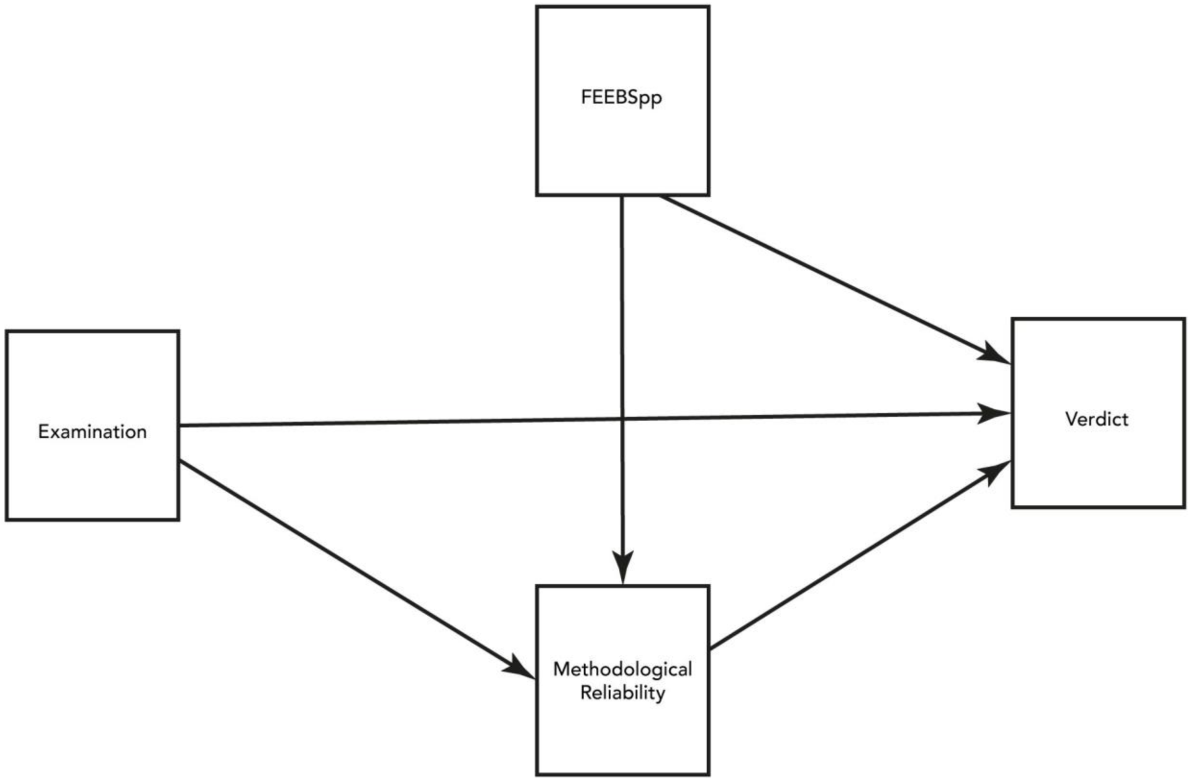
The voice is the product of two mechanisms: speech organs and language, and both have considerable flexibility (Nolan, 2001). The speech organs are capable of making a diverse range of sounds. Voices are also capable of communicating in a wide range of tones, with pronunciation and style that can change from context to context (Bogan, 2011). As Bogan argues, "the result of this 'plasticity'… is that no two utterances from the same individual are ever, strictly speaking, identical in a physical sense" (p 382, 2011). Nolan (2001) points out that this plasticity can be manipulated to disguise the voice, during stress, fatigue, and illness, as well as when it is modified through loudness, pitch, emphasis, and rate. Even style and pronunciation can change in different circumstances. Examiners also have to contend with the transfer from natural spoken word to electronic recordings (Morrison, 2009), older recordings on tape degrade quickly, (Bogan, 2011), and there is the likelihood of interference from background noise (Bogan & Roberts, 2011; Morrison, 2014).

Voice experts are used to analyse recordings taken during an investigation with recordings taken of a known suspect. There are few regulations placed on the use of voice identification in Australian courts, and it has only been recently noted that the dangers associated with photographic comparison might be applied to voice analysis (Edmond, 2011b). The "experts" who give voice identification testimony are frequently phoneticians, translators or linguists, with skills acquired through training in acoustic measurement and visualisation.

Voice identification experts use two main approaches: auditory and acoustic. In the auditory approach, experts listen for the “accent” of speakers, and attempt to locate distinctive or unusual features, which are largely caused by vocal anatomy and idiosyncratic ways of making sounds. Acoustic analysis is conducted on a computer, normally by producing waveforms to plot amplitude, and spectrograms to show a visual display of the acoustics of speech, including fundamental frequency (pitch), and resonant frequencies. While the use of both types of technique can give experts a range of points of comparison between samples of recordings, the ultimate decision must be made by the expert based on their overall impression.

Concern over voice identification and the use of spectrograms was raised as early as 1979, in a report made by the National Research Council (Committee on Evaluation of Sound Spectograms & Sciences). Despite this, there has been little validation research undertaken to improve the use of such evidence. No population statistics about the frequency of features exists (Bogan & Roberts, 2011), and a recent survey of expert voice analysis in INTERPOL shows that there are large variations in the qualifications, techniques, approaches, and reporting frameworks used worldwide (Morrison et al., 2016). In that survey, less than half of the respondents indicated they had more than a single recording of the same speaker for use in comparisons. While voice comparison is an “identification” science, much like anthropometric facial comparison there is no physical evidence to compare.

Study 3 therefore would test three types of forensic science, and two types of cross-examination. Comparing models across fingerprint, anthropometric facial comparison, and voice identification would allow for investigation of differences between the perception of different evidence types. However, based on the results from Study 2, it was predicted that bias towards forensic evidence would also affect the ratings of reliability and, indirectly, of verdict. The final hypotheses was that for each group, there would be a direct effect of pro-prosecution bias on verdict and an indirect effect through methodological reliability. Figure 5 shows the model to be tested



.

Figure 13. Path model showing predicted relationships between Examination type, Methodological Reliability, scores on the FEEBSpp subscale, and Verdict

4.4.3 Forensic odontology: Sensitivity manipulation check

A final activity was given to participants as a way to test for sensitivity to unreliable expert evidence. Participants read through a newspaper article describing the work of an expert in forensic odontology. It was presented as an extra task that they were doing to help a future study that wanted to know if "journalistic articles could be used to present necessary trial information". Participants read through a mock newspaper article that was reporting on a murder where a forensic odontologist had examined a bite mark found on the victim's body. If scientifically informed cross-examination had helped to educate participants about the flaws embedded within fingerprint, anthropometric facial comparison, or voice identification, then they should more easily recognise the same criticisms when presented in an article, and give the evidence a lower reliability rating, than participants who had read through the experience and qualification based cross examination.

Forensic odontology involves the identification of a person by comparing a record of their set of teeth with a bite mark that has been left on a victim. Bite mark comparison is considered by Thornton and Peterson (2007) to have “very high” subjectivity, and even argue that other pattern comparison specialists would argue that forensic odontologists suffer from “The Emperor’s New Clothes Syndrome” (p. 48). Forensic odontology, or dentistry, is mainly used as an identification tool, whether it concerns identifying remains when injury or post-mortem changes means typical techniques cannot be used, or in the identification of suspects from bite-marks found on victims (Avon, 2004). Information used in the article regarding the field was taken from Kennedy (2011), Avon (2004), Aksu and Gobetti (1996), and Bowers (2010).

Arguments against the reliability of forensic odontology are similar to those for the previously discussed fields in that the quality of the material can differ wildly (as skin malleability makes it hard to accurately record enough information) and experts will differ in the methods they use to counter this. There is a lack of certification needed to work in forensic areas, as opposed to being a dentist, and no wide scale studies have examined the frequency of dental characteristic distribution either as a single feature or in combinations (Aksu & Gobetti, 1996; Avon, 2004; Bowers, 2010; Kennedy, 2011; Page, Taylor, & Blenkin, 2011). Some experts argue that dental evidence can be used as exclusion evidence (Aksu & Gobetti, 1996), but most commentators agree that it should not be used as identification evidence.

The article (included in Appendix D), described the testimony of the expert, including the method used during the comparison, and also some of the arguments used against him during cross-examination. There are several criticisms of the field: no statistical studies to back up claims, there was no second expert, and there was the possibility of confirmation bias (as the examiner knew the suspect was the only one being considered by the police). These were similar to those used in the criminal trials that participants had just read, although paraphrased differently so that they were not too familiar. To ensure participants had carefully read the article a couple of filler questions were asked. Participants were then asked to rate the reliability of the forensic evidence presented in the article. It was predicted that if participants had been educated by the scientific cross-examination then the odontology evidence should receive a low reliability rating. Forensic odontology was also one of the forensic fields asked about in the beginning of the study, and ratings could be used as a pre- and post- comparison, and it was predicted that groups who had seen the scientifically-informed cross-examination would have greater reductions in the ratings of reliability.

4.5 Method

Procedure

Study 3 was also hosted on Qualtrics™. The method was very similar to the second study: participants accessed the link individually through an anonymous portal which recorded IP addresses so that no participants could make two attempts. Upon starting the study, participants were given information sheets, gave consent, and completed a task to ensure they were reading instructions thoroughly. Participants could not proceed past the instructions unless they completed the task correctly. As in study 2, responses were initially kept separate, with undergraduate students participating for course credit and those completing the study for a chance to win one of three JB Hifi vouchers using separate links. Participant responses were also timed. Given that there had been no effect from note taking, participants were not provided with text entry boxes at the bottom of each page during Study 3.

The study began by asking for demographic details. Participants were also told the study was looking at how people get information about science and technology, and were asked to indicate via check boxes where they got most of their information about science and technology. Options were: newspaper (including online), magazine, TED talks, You Tube, Internet (other), Books (fiction), books (nonfiction), TV (non-fiction), TV (fiction), Radio, Government Agency, or Family/friends/colleagues. There was also an optional text entry "other" box. Participants then completed the Scientific Literacy questionnaire. On the next page, participants were asked to give reliability ratings for the following types of evidence (using a slider scale from 0 – 10, with 0 being not at all reliable): hair comparison, bloodstain pattern, handwriting comparison, forensic dentistry, DNA, voice identification, eyewitness testimony, facial comparison, fingerprint, and firearm identification. Finally, they completed both subscales of the FEEBs. After reading through the transcript participants gave a verdict, rated the case strength, and completed the methodological reliability questionnaire.

Trial Summary

The trial used in Study 3 was entirely fictional, as a search looking for cases that could easily accommodate fingerprint, facial mapping, and voice identification evidence were unable to be located. The trial involved a defendant charged with the murder of a female he was previously not acquainted with (for the complete transcript see Appendix C). The defendant was accused of attempting to burgle her home, but when the victim came home unexpectedly there was a struggle and she was pushed over the balcony, being killed upon impact with the ground. In the scenario the defendant has come to the attention of the police several days later when attempting to sell the victim's Macbook to a Cash Converters (a second hand goods retailer). The police subsequently investigate the defendant including obtaining a warrant to search his house, where they come across other items from the victim's house. The non-forensic identification evidence, given in the form of summaries, includes testimony presented by a paramedic and police officer who attended the scene, an arresting officer, medical examiner, manager of Cash Converters, the property manager of the building, and the former employer of the defendant.

Three versions of the trial were used: in the first version finger prints are found on a patent leather belt found on the victim as well as on the railing of the balcony, in the second version CCTV footage from a security camera showed a male walking into the apartment building, and in the third version the victim is described as having made a phone call to emergency services while her attacker is in the apartment (which was recorded). In each version the expert describes how they conducted the analyses. The final pages showed the same jury instructions on the burden of proof and the meaning of beyond reasonable doubt that were used in study 1 and 2.

Sensitivity Check

The sensitivity check was presented as an extra task that they were doing to help a future study. After completing the verdict and reliability ratings, the next page presented a message asking participants to help rate data for an Honours project that would take place next year. Taking part was voluntary (and some participants did elect to skip it). If participants agreed to take part, they were directed to a page with a description of a fake study that wanted to know if "journalistic articles could be used to present necessary trial information". Participants then read through a mock newspaper article that was reporting on a murder where a forensic odontologist had examined a bite mark located on the body. After reading the article, participants were asked to rate the reliability of the evidence presented by the examiner using a sliding scale, ranging from 0 to 10, where 10 was "very reliable".

4.6 Results

4.6.1 Public and Undergraduate Sample Differences

Participants were either undergraduate students in a first year psychology course who participated for course credit, or members of the public who responded to hard-copy posters, online versions shared over social media, or were recruited through word of mouth.

Undergraduate students sample

One hundred and sixty-three students began the study, with 144 completing it (*M* = 20.89, *SD* = 5.49). Of those who completed, 97 identified as female, 45 as male, and 2 as "other" (with one using an optional text entry box to identify themselves as trans-masculine). Of those who didn't finish, 7 did not get past the instructions and attention check, 4 did not get past the demographics section, and 4 exited after starting to read through the trial. The random assortment into a trial takes place when participants start reading the forensic expert testimony.

Of those who started the trial, two exited before being sorted into a condition, one was in condition who saw the fingerprint evidence with qualification-focussed testimony, and one in voice identification with scientifically-informed. All participants who completed the study were citizens and were jury eligible.

Public sample

One hundred and fifteen participants accessed the study through the public link, with 67 completing it. Of those who didn't finish, 31 did not get past the instructions, 14 did not get past the demographics section (with 1 person answering that they worked for the law courts or justice system, which automatically directed them to a message saying they are not eligible to do the survey) and 3 did not complete but exited after starting to read the study.

Differences between samples

There were several demographic differences between the two groups. There was a statistically significant difference in age between the student sample, M = 20.94, SD = 5.277, and the members of the public M = 37.30, SD = 13.55; U = 8916.00, z = 610.03, p > .001, r=-.05, with a very small effect size. However there was no statistical difference in gender between the samples, χ^2 (1, n = 185) = 213, p = .62, *phi* = -.05, using Yates continuity correction. As in Study 2, the two samples were combined after examination of correlations between variables and of differences in highest education level achieved (see below).

A chi-square test for independence indicated an overall statistical difference between the samples in highest level of education achieved. The numbers at each level are shown in Table 1.

Table 8. Highest education level achieved split by sample and total (with percentage of total sample)

	Undergraduate	Public	Total
Year 11 or below	2	3	5 (2.4%)
Year 12	110	5	115 (54.5%

	Undergraduate	Public	Total
Certificate I to IV	10	5	15 (7.1%)
Tafe Diploma/Advanced Diploma	11	11	22 (10.4%)
Bachelor Degree	10	23	33 (15.6%)
Postgraduate Degree	1	20	21 (10%)

Highest Education Level Achieved

Independent samples Kruskal-Wallis tests were used to compare scores on the FEEBS subscales, age, scientific literacy score, and the reliability ratings given to each type of forensic science. Statistical differences were found for age and anthropometric facial comparison reliability ratings.

Table 9. Kruskal-Wallis Test statistics

	n[3]	X^2	*p*
Age	211	113.75	<.001
FEEBSpp	211	1.90	.86
FEEBSpd	211	4.15	.53
Hair	208	5.55	.35
Bloodstain	207	5.09	.41
Handwriting	207	7.33	.19
Forensic Dentistry	209	6.07	.30
DNA	209	1.83	.87
Voice ID	209	2.49	.78
Eyewitness	210	8.08	.15

[3] Not all participants chose to complete every question

	n^3	X^2	*p*
Facial Comparison	210	11.82	.04
Fingerprint	209	6.05	.30
Firearm Identification	209	3.32	.65

df = 5 for all comparisons

As the only difference was in age and anthropometric facial comparison, the two samples were combined for the other statistical analyses. The combined age of participants who completed the study ranged from 18 to 75 (M = 25.79, SD = 11.34); and 67% identified as female.

4.6.2 Sources of information

Participants were asked: "Where do you get most of your information about science and technology?" The responses are shown in Figure 2, below.

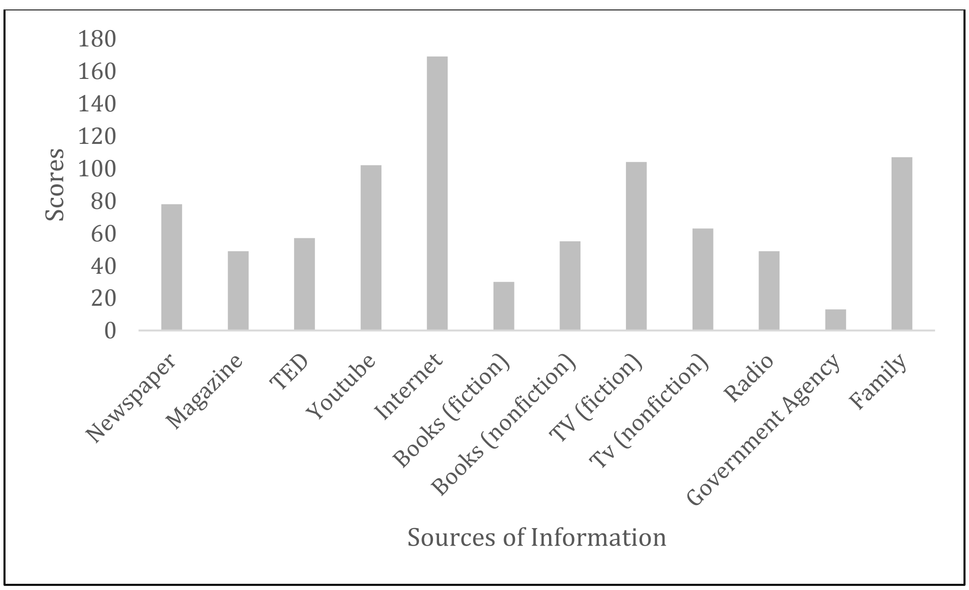


Figure 14. Frequency of responses to participants' sources of scientific information

The internet was unsurprisingly the most frequently selected source of information, with 80.2% of participants saying they learnt about science online, with YouTube and TED talks also popular. The responses show that participants use a range of sources to get information (*M* = 4.22, *SD* = 2.46), and while fictional depictions of science through television or books were also popular choices (selected by 29.9% and 14.2% of respondents), they were never the only selected source.

4.6.3 Normality of data for use in the path analysis

As the path analyses were conducted at a multiple-group level (defined by the type of forensic science presented in the trial), multivariate normality of the variables to be used in the path analysis was calculated for each. Mardia's multivariate kurtosis was -1.37 with a critical ration of -.69 for the fingerprint group, -3.96 and -2.01 for facial mapping, and -1.42 and -.69 for voice identification. Although the facial mapping kurtosis was just over Bentler and Wu's (2002) recommended level, maximum likelihood was used on all three groups.

4.6.4 Reliability of forensic sciences

The reliability of forensic science was measured on a 10 point Likert scale, with 0 indicating "not very reliable", and 10 indicating "very reliable". The ratings appear below in Table 3.

Table 10. Reliability of types of evidence

	M	**SD**
Hair	6.14	2.65
Bloodstain	6.46	2.41
Handwriting	4.80	2.30
Forensic Odontology	8.32	2.00
DNA	9.14	1.40
Voice Identification	5.40	2.19

Eyewitness	4.61	2.17
Facial Comparison	5.31	2.19
Fingerprint	8.58	1.49
Firearm	6.17	2.30

Unsurprisingly, DNA and fingerprint examination had the highest reliability ratings, with eyewitness identification receiving the lowest score. Voice identification and facial comparison were very close, in both ratings and in standard deviation scores.

FEEBS

Ratings on the FEEBS subscales were similar to Study 2, (FEEBSpp M = 13.95, SD = 3.29; FEEBSpd M = 12.79, SD = 2.98).

Verdict choice

Verdict choices for each condition are shown in Table 11, below.

Table 11. Verdict Choice by Condition

	Fingerprints			Facial Mapping			Voice Identification		
	Not guilty	Guilty	Total	Not guilty	Guilty	Total	Not guilty	Guilty	Total
Scientifically Valid	22 (59.5%)	15 (40.5%)	37	22 (61.1%)	14 (38.9%)	36	14 (43.8%)	18 (56.3%)	32
Experience Focused	17 (48.6%)	18 (51.4%)	35	21 (58.3%)	15 (41.7%)	36	19 (54.3%)	16 (44.8%)	35
	39	33	72	42	29	72	34	34	67

In the Fingerprint conditions, there were more guilty verdicts in the scientifically valid condition than in the experience focussed condition. In the facial mapping condition, there was little difference between verdict choices for the two types of cross examination. For voice ID the pattern is reversed, more guilty verdicts were given after hearing the scientifically valid cross-examination questions, but only barely.

Multinomial logistic regression was also conducted to see if there was a difference in verdict choice due to either examination type or the forensic science presented. The Pearson chi-square statistic was low, and the test did not reach significance, indicating that there was a good fit to the data, χ^2 (2, N = 211) =1.58, p = .66. However, the variance explained was small (Cox & Snell R^2 = .24, Nagelkerke R^2= .33), and neither science (χ^2 (2, N = 211) =1.534, p = .46), nor examination (χ^2 (1, N = 211) =0.04, p = .85) were significant predictors.

4.6.5 Methodological Reliability

The multi-group comparisons shown above compared relationships between variables, but did not compare mean scores across groups. Factorial ANOVA was used to compare the main effect of different evidence type and examination type, as well as the interaction, on methodological reliability.

The main effect for type of forensic science was not statistically significant, yielding an F ratio of F(2, 205) = .11, p= .90, η^2 = .00, indicating that ratings of reliability were not different for the types of science shown. However, there was a statistically significant effect for examination type, with an F ratio of F(1, 205) = 5.51, p > .05, η^2 = .03. The interaction was not statistically significant, F(2, 205) = 1.00, p= .37, η^2 = .01.

Table 12. Means and Standard Deviations for Methodological Reliability Ratings by Condition

	Fingerprint	Facial Mapping	Voice Identification
Scientifically informed	24.54 (6.04)	23.75 (6.22)	22.97 (4.37)
Experience Based	25.31 (6.25)	25.19 (5.42)	26.49 (6.74)

Correlations

The relationships between age, FEESBs subscales, and the pre-trial ratings of the different types of forensic identification sciences (and eyewitness evidence), are presented in Table 6.

Age in this study only had a significant correlation with scores of pre-trial ratings of facial comparison reliability. The pro-prosecution subscale had low to moderate relationships with almost all of the identification science ratings, with only firearm and voice ID not achieving statistical significance, and having the smallest effect sizes.

Table 6: Correlations for Age, scores on the FEEBSpp and FEEBSpd, Reliability of forensic science fields, and Methodological Reliability

	Age	FEEBS pp	FEEBS pd	Hair	Blood -stain	Hand	Dentistry	DNA	Voice ID	Eye-witness	Facial	Finger-print	Firearm	Method. Rel.
Age	1	.04	-.11	.10	-.10	.09	.05	-.01	.07	-.07	-.16*	-.03	.11	-.08
FEEBSpp		1	.39**	.19**	.29**	.14*	.20**	.19**	.04	.26**	.24**	.28**	.07	.19**
FEEBSpd			1	.04	.17**	.10	13*	.04	.02	.10	.16**	.13*	.01	-.02
Method. Reliability				1		.17**	.34**	.27**	.14*	.08	.08	.21**	.12*	.06
Hair					1	.37**	.36**	32**	.20*	.24**	.31**	.37**	.18**	.08
Bloodstain						1	.41**	.24**	.53**	.39**	.46**	.28**	.28**	.23**
Handwriting							1	.59**	.28**	.22**	.25**	.40**	.28**	.13*
Forensic Odontology								1	.26**	.18**	.17**	.41**	.24	.11
DNA									1	.43**	.52**	.21**	.29**	.22**
Voice										1	.67**	.34**	.25**	.23**

Eyewitness	1	.41**	.21**	.20**
Facial Comparison		1	.25**	.16*
Fingerprint			1	.22**
Firearm				1

4.6.6 Age

In Study 2, age had a statistically significant negative relationship with methodological reliability ratings. The table above shows that there was no overall significant relationship, however the data was split into different types of science. While there was no statistically significant relationship between age and methodological reliability ratings for fingerprint comparison (r = -.15, p = .22), there was a statistically significant negative relationship with anthropometric facial comparison (r = -.35, p > .05), and a positive relationship with voice identification (r = .26, p > .05).

4.6.7 Path analyses

The sample size was small, and so bootstrapping (maximum likelihood) was used with 2000 samples, and 95% bias corrected confidence intervals. There were two categorical independent variables (three types of forensic identification, and two types of cross-examination). Grouping participants via forensic evidence type meant that path analyses remained the best method of testing hypotheses. A multi-group analyses was used through *IBM®* SPSS® *Amos* 24. The multi-group analysis will be discussed first, before the individual models.

Multi-group Analysis

Multi-group path analysis was used to measure invariance and to compare the three models to one another. This type of analysis would allow the comparison of the association between variables (e.g. methodological reliability and verdict) based on the coefficients, even if both were statistically significant in separate models. This is done by examining whether the coefficients are significantly different across the types of forensic evidence. It would also calculate model fit statistics.

Model fit statistics

Calculation of the fit indices of the unconstrained baseline comparison model showed good, but not perfect fit, with a CFI of .997, Hu and Bentler (1999) consider 0.95 to indicate good fit. The RMSEA was 0.02, which is under Hu and Bentler's (1999) recommended value of 0.06, and using the criteria set out by MacCallum, Browne, and Sugawara (1996), indicates good fit. However, given that there was a low number of degrees of freedom, there may be sampling error (Kenny, Kaniskan, & McCoach, 2014).

Differences between the types of forensic evidence

The Chi-square difference test shows that there was no difference between groups, indicating that there was no overall difference in the fit of the models, χ^2 (3, N = 211) =3.30, p > .05.

Constraining the weights between two variables in a multi-group analysis would allow testing of the hypotheses that the relationship between those two variables would differ across groups. This calculates a Chi-square difference test, where the three models were freely estimated except for the path between two selected variables, which was constrained to be equal across groups.

To test the hypotheses that cross examination type would affect the three types of forensic science differently, the relationships between examination and methodological reliability, and then examination and verdict were constrained in turn.

Constraining the relationship from examination type to methodological reliability found no significant difference, indicating that there was no difference in the relationship across the three types of forensic science, χ^2 (3, N = 211) =3.30, p = .48. Nor were differences found between scores on the FEEBSpp and ratings of methodological reliability, χ^2 (5, N = 211) =3.54, p = .62. Constraining examination type to verdict, however, indicates a statistically significant difference, indicating that there was variation in the relationship across the three types of forensic science, χ^2 (5, N = 211) = 12.69, p = .03.

Fingerprint

For participants in the Fingerprint evidence group, the model was able to predict just under half the variance in verdict choice (please note: the evidence based cross examination was coded as 1, and qualification based cross examination is coded as 2).

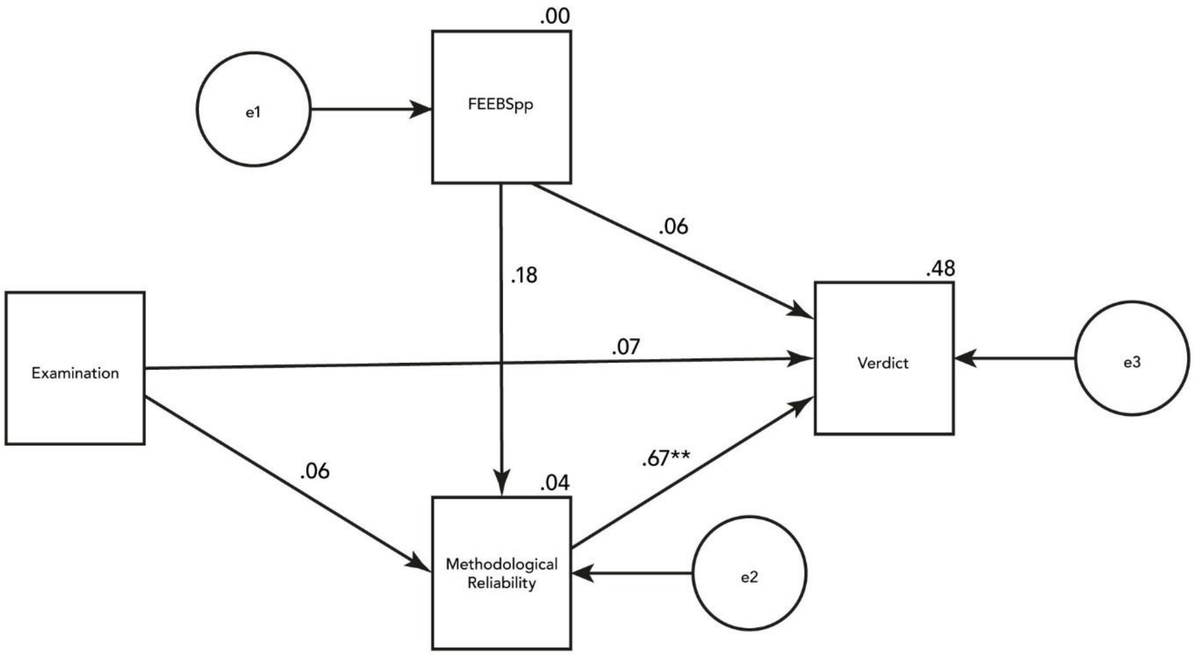


Figure 15. Path model showing examination, Methodological Reliability, scores on the FEEBSpp subscale and Verdict for Fingerprint Identification

Methodological Reliability had a large, statistically significant effect on verdict, such that those who rated the fingerprint evidence as more reliable were more likely to give a guilty verdict. FEEBSpp scores had a moderate relationship with reliability scores, although this did not reach significance, and examination of the total, direct and indirect effects (in Table 7, below), show that there was a small indirect effect on verdict, although this again did not reach significance.

Table 13. Unstandardized and standardized regression weights, and total, direct and indirect effects for FEEBpp subscale, methodological reliability, probability, certainty and verdict for fingerprint identification conditions

	Unstandardised				**Standardised**		
	Original Estimate		Bootstrapping		Original Estimate	Bootstrapping	
	b	S.E. b	b (95%CI)	S.E.b	β	β (95%CI)	S.E. β
Examination → Verdict	.07	.09	.07 [-.11, .24]	.09	.06	.07 [-.11, .24]	.09
Total			.12 [-.12, .33]			.12 [-.12, .33]	
Direct			.07 [-.11, .24]			.07 [-.11, .24]	
Indirect			.05 [-.11, .20]			.05 [-11, .20]	
Examination → Methodological Reliability	.75	1.40	.92 [-1.96, 3.58]	1.40	.06	.07 [-.16, .30]	.12
Total			.90 [-1.96, 3.58]			.07 [-.16, .30]	
Direct			.90 [-1.96, 3.58]			.07 [-.16, .30]	
Indirect			...			...	
Methodological Reliability → Verdict	.06**	.01	.06 [.04, .07]	.01	.67**	.67 [.53, .78]	.07
Total			.05 [.04, .07]			.67 [.53, .78]	
Direct			.05 [04, .07]			.67 [.53, .78]	
Indirect			...			...	
FEEBSpp → Verdict	.01	.01	.01 [-.02, .04]	.02	.06	.06 [-.13, .26]	.10

Total			.03 [-.01, .06]			.18 [-.05, .40]	
Direct			.01 [-.02, .04]			.06 [-.13, .27]	
Indirect			.12 [-.01, .04]			.12 [	
FEEBSpp→Methodological Reliability	.34	.22	.34 [-.16, .76]	.22	.18	.18 [-.09, .37]	.11
Total			.34 [-.15, .74]			.18 [-.09, .37]	
Direct			.34 [-.15, .74]			.18 [-.09, .37]	
Indirect			...			...	

Anthropomorphic Facial Comparison

Similar to the fingerprint condition, there was a large, statistically significant relationship between methodological reliability and verdict. Examination of the total, direct and indirect effects shows that the strength of the relationship was very similar to that shown between fingerprints and reliability ratings. A smaller percentage of the variance in verdict was explained, indicating that this model had a lesser fit.

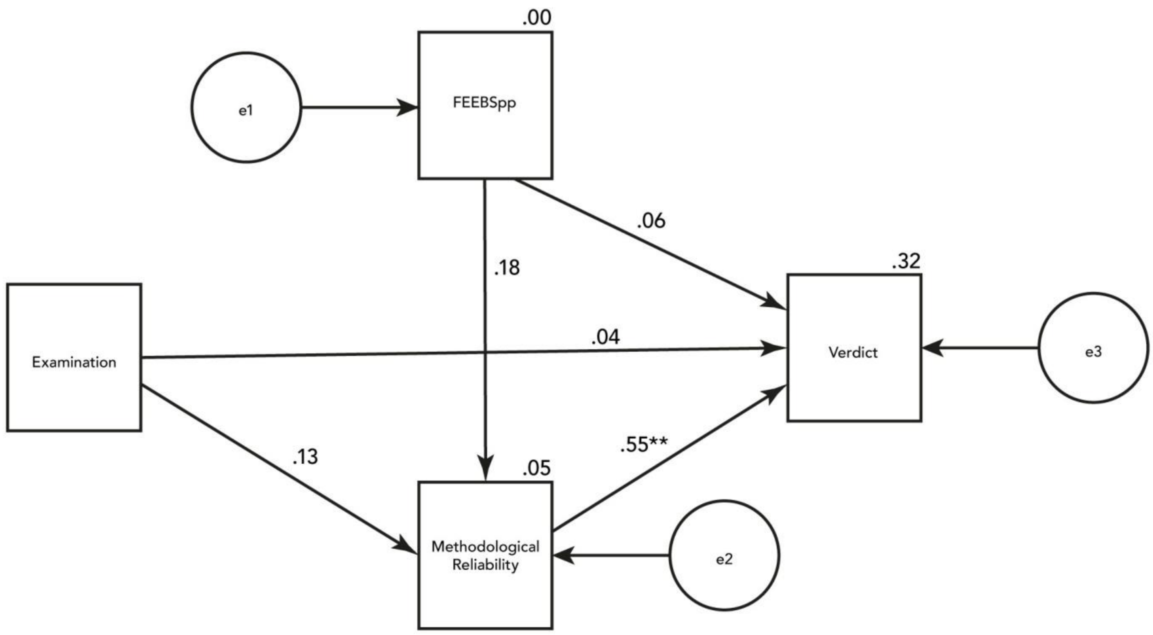


Figure 16. Path model showing examination, Methodological Reliability, scores on the FEEBSpp subscale, and Verdict for Anthropomorphic Facial Comparison Identification

Table 14. Unstandardized and standardized regression weights, and total, direct and indirect effects for FEEBpp subscale, methodological reliability, probability, certainty and verdict for anthropomeorphic facial comparison conditions

	Unstandardised				**Standardised**		
	Original Estimate		Bootstrapping		Original Estimate	Bootstrapping	
	b	S.E. *b*	*b* (95%CI)	S.E.*b*	β	β (95%CI)	S.E. β
Examination → Verdict	-.04	.10	-.04 [-.24, .16]	.10	-.04	-.04 [-.25, .16]	.10
Total			.05 [-.17, .27]			.05 [-.17, .27]	
Direct			-.02 [-.21, .17]			-.02 [-.22, .18]	
Indirect			.07 [-.04, .17]			.07 [-.20, .25]	
Examination → Methodological Reliability	1.47	1.32	1.75 [-.1.22, 3.85]	1.26	.13	.13 [-.11, .33]	.11
Total			1.69 [-.1.22, 3.85]			.15 [-.11, .33]	
Direct			1.69 [.1.22, 3.85]			.15 [-.11, .33]	
Indirect			...			...	
Methodological Reliability → Verdict	.05**	.01	.05 [.03, .06]	.01	.55**	.55 [.32, 65]	.08
Total			.04 [.03, .06]			.04 [.32, .65]	
Direct			.04 [.03, .06]			.50 [.32, .65]	
Indirect			...			...	
FEEBSpp → Verdict	.01	.02	.00 [-.04, .05]	.02	.06	.06 [-.16, .29]	.11
Total			.01 [-.04, .06]			.06 [-.08, .39]	

Direct			.00 [-.04, .05]			.02 [-.16, .28]	
Indirect			.01 [-.01, .04]			.04 [-.03, .22]	
FEEBSpp → Methodological Reliability	.38	.24	.40 [-.13, .83]	.23	.08	.19 [-.06, .37]	.10
Total			.18 [-.13, .83]			.18 [-.06, .37]	
Direct			.18 [-.13, .83]			.18 [-.06, .37]	
Indirect			...			...	

Voice Identification

Methodological reliability also had a strong effect on verdict. Participants who rated the voice identification evidence as more reliable were more likely to give a guilty verdict. Unlike the previous two types of forensic science, the type of examination had a statistically significant effect on both verdict and reliability ratings. Participants who read through the qualification based examination type gave higher ratings of reliability. However, participants reading through the qualification based cross-examination were more likely to give a not-guilty verdict, which was the opposite of the predicted interaction.

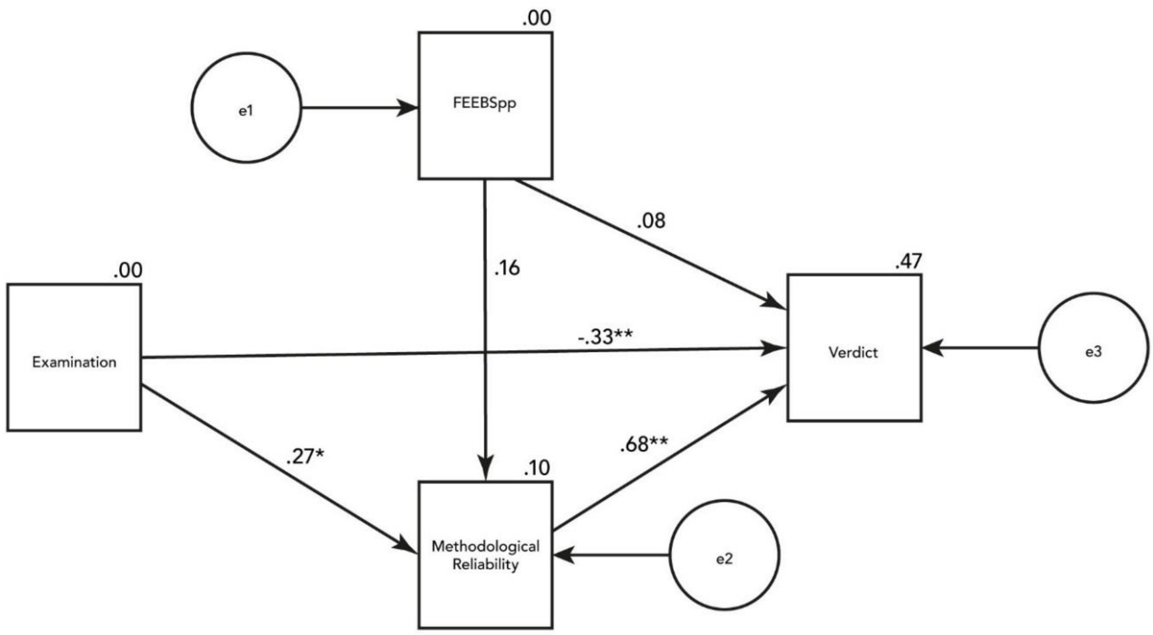


Figure 17. Path model showing examination, Methodological Reliability, scores on the FEEBSpp subscale, and Verdict for Anthropomorphic Facial Comparison Identification.

Table 15. Unstandardized and standardized regression weights, and total, direct and indirect effects for FEEBpp subscale, methodological reliability, probability, certainty and verdict for voice identification conditions

	Unstandardised			**Standardised**			
	Original Estimate		Bootstrapping	Original Estimate		Bootstrapping	
	b	S.E. b	b (95%CI)	S.E.b	β	β (95%CI)	S.E. β
Examination → Verdict	-.33**	.09	-.33 [-.51, -11]	.10	-.33**	-.32 [-.50, -.13]	.09
Total			-.15 [-.39, .11]			-.15 [-.38, .11]	
Direct			-.33 [-.51, -.13]			-.33 [-.50, -.13]	
Indirect			.18 [.04, .36]			.18 [.03, .36]	
Examination → Methodological Reliability	3.11*	1.37	3.11 [.56, 5.93]	1.41	.27*	.25 [.05, .47]	.11
Total			3.12 [.56, 9.05]			.27 [.05, .47]	
Direct			3.12 [.56, 6.05]			.27 [.05, .47]	
Indirect			…			…	
Methodological Reliability → Verdict	.06**	.01	.06 [.05, .07]	.01	.68**	.67 [.51, .83]	.08
Total			.06 [.05, .07]			.68 [.51, .83]	
Direct			.06 [.05, .07]			.08 [.51, .83]	
Indirect			…			…	
FEEBSpp → Verdict	.01	.01	.01 [-.01, .03]	.01	.08	.08 [-.06, .22]	.08
Total			.02 [-.01, .06]			.14 [-.11, .39]	

Direct			.01 [-.01, .03]			.08 [-.06, .22]	
Indirect			.01 [-.01, .04]			.09 [-.09, .27]	
FEEBSpp→Methodological Reliability	.15	.19	.24 [-.26, .55]	.21	.16	.16 [-.17, .35]	.13
Total			.15 [-.26, .55]			.10 [-.17, .35]	
Direct			.15 [-.26, .55]			.13 [-.17, .35]	
Indirect			…			…	

4.6.8 Sensitivity Check

The final analyses concerned the sensitivity check. Not all participants opted to take part, but 204 participants gave ratings after reading the article. Factorial ANOVA was used to compare the main effect of different evidence type and examination type, as well as the interaction, on ratings of reliability. Scores on the FEEBSpp were entered as a covariate. The ratings are shown below.

Table 16. Means and Standard Deviations for Odontology Reliability Ratings by Condition

	Fingerprint	**Facial Mapping**	**Voice Identification**
Scientifically informed	58.08 (27.91)	68.74 (22.24)	59.84 (20.68)
Experience Based	64.72 (23.00)	70.38 (19.47)	62.97 (22.45)
Total	61.16 (25.78)	69.56 (20.76)	61.48 (21.52)

The main effect for type of examination was not statistically significant, yielding an *F* ratio of $F(1, 197) = .1.06$, $p = .31$, $\eta^2 = .01$, indicating that ratings of reliability were not affected differently by whether participants read through the qualification-based scientifically-informed cross examination. However, there was a statistically significant effect for type of science type, with an *F* ratio of $F(2, 197) = 3.51$, $p > .05$, $\eta^2 = .03$. The interaction was not statistically significant, $F(2, 197) = .25$, $p = .78$, $\eta^2 = .00$. The covariate, scores on the FEEBSpp, was also statistically significant, $F(1, 197) = 3.93$ $p > .05$, $\eta^2 = .02$.

4.7 Discussion

This study focussed on **RQ3**: *Is the effect of cross-examination on juror decisions similar to or different from that of opposing experts?* And, in addition, also looked at whether the relationship between the individual characteristic of bias towards forensic evidence (**RQ4**) to perception of reliability would be similar to that shown when opposing expert testimony was used as an educational method. The specific hypotheses were that scientifically-informed cross-examination would lead to lower ratings of methodological reliability than experience-based cross-examination, and that this would then lead to fewer guilty verdicts. It was also predicted that bias towards forensic evidence, as shown by higher scores on the FEEBSpp subscale, would lead to higher scores on the methodological reliability scale. This would then lead to more guilty verdicts. Overall, the results showed partial support for the first prediction, and showed some support for the second.

4.7.1 Methodological Reliability

Across all three types of forensic science, there was a large, positive relationship between methodological reliability and verdict choice, indicating that participants' perceptions of the reliability of either fingerprint comparison, anthropomorphic facial comparison, or voice identification, explained a large proportion of the variance in verdict choice. This is consistent with the results from Study 1 and 2.

If scientifically-informed cross-examination had been successful at explaining why the three types of forensic identification sciences were unreliable, then participants in those conditions should have had lower methodological reliability scores than those seeing experience-focussed cross-examination. The results from the factorial ANOVA show that this was the case overall, with those in the scientifically-informed conditions having lower ratings, although the effect size was small. Examination of the individual path analyses shows that for facial comparison and voice identification, there was a significant effect of the examination type on ratings of methodological reliability. However, this was not the case for fingerprints. The question is why is this so?

Fingerprints are highly regarded, and have a long history as one of the renowned types of identification sciences, even though legal commentators point out that the empirical evidence behind the science is as in need of validation studies as the other fields, with the exception of

DNA profiling (Mnookin, 2003, 2008; Thompson & Cole, 2007). At the beginning of the study, participants rated fingerprint evidence as the most reliable type of identification evidence second only to DNA, which is consistent with the few previous studies that have asked laypersons to rate the different types of evidence (Baskin & Sommers, 2010; Lieberman et al., 2008). It is possible that a layperson's perception of fingerprint identification as a field is so robust, that the criticisms against the expert's testimony had less influence in reducing reliability ratings. Anthropometric facial comparison and voice identification had highly similar initial ratings of reliability in the current study, although this was much lower than the ratings for fingerprint comparison. The criticisms may have had more of an impact due to those types of evidence being initially considered not as infallible.

Koehler et al. (2016) pointed out that in their study, manipulation checks showed that mock jurors were able to recognise whether a fingerprint matching method had been subject to scientific validation or not, but this did not affect guilty verdicts, showing no sensitivity. They argued that this was similar to McQuiston-Surrett and Saks (2009), who found that participants were not impacted by hearing that a hair analyst's technique had *not* undergone scientific testing. Similarly, in their second study, Austin and Kovera (2015), found no evidence of an interaction between validity and verdict, even though scientifically-informed cross-examination lead to more guilty verdicts.

4.7.2 Examination Type

Comparison of the relationships across groups using the Chi-square difference test revealed that there was a difference in the relationships between examination type and verdict. The only time this relationship reached statistical significance was in the voice identification conditions, where there was a moderate to large effect. However, this relationship was in the opposite direction to what was predicted: participants who read through the qualification based cross-examination gave *fewer* guilty verdicts than those in the cross-examination condition. In the other two types of forensic evidence this pattern, although not reaching significance, was the opposite. So why were participants who saw the voice identification expert testimony less likely to give a guilty verdict when hearing experience based cross-examination?

None of the located previous studies asking for reliability, or trustworthiness, ratings of different types of forensic sciences have asked participants about voice identification, nor has it been used as an evidence type in any jury decision making studies (although research has looked at the accuracy and perceived usefulness of layperson "earwitnesses", these are all studying non-expert voice recognition, see Olsson, Juslin, & Winman, 1998; Philippon, Cherryman, Bull, & Vrij, 2007; van Wallendael, Surace, Parsons, & Brown, 1994). Therefore, it is not possible to know what laypersons know or think about voice comparison. The reliability ratings for the beginning of the study indicate that jurors considered it to be slightly *more* reliable than the anthropometric facial comparison evidence.

A possible explanation is that participants may have found the voice identification testimony to be harder to evaluate or comprehend in comparison to the fingerprint and anthropometric facial comparison testimony. Although it is unlikely many laypersons are familiar with the specifics of fingerprint matching or with the features looked at during facial matching, the measurements and pattern matching are intuitive concepts (and prominently featured in fictional depictions of forensic investigations). The expert voice identification testimony used in the current study, however, may have been more difficult to envision. The testimony, which describes the analysis of phonetic sounds, how vocal fold activity can affect sound and quality, and spectrographic analysis may have been more complex than the written testimony describing facial mapping and fingerprint comparison. A large amount of research on dual processing suggests that higher complexity can lead to an increase in reliance on heuristics (Chaiken & Maheswaran, 1994b), such as source expertise, with more experienced experts being seen as more persuasive (Cooper et al., 1996; Greenberg & Wursten, 1988; Ratneshwar & Chaiken, 1991), with the effect being especially powerful when information is new (Kumkale et al., 2010). Additionally, the qualification-focussed cross-examination is unlikely to have assisted jurors in processing the reliability of the voice identification evidence.

A limitation is that Study 3 did not directly measure whether participants thought that the expert as described in the experienced-based conditions was “highly” experienced, or whether he would be seen as inexperienced. If the qualifications of the expert gave jurors the impression of an inexperienced expert with inadequate knowledge of the field, the emphasis on attacking qualifications may have decreased source credibility of the expert. In turn, this may have led participants to place less value in whichever type of identification evidence they experienced, leading to fewer guilty verdicts and lower ratings of methodological reliability. An increased reliance on heuristics in the voice identification condition may have led to more participants *not* being persuaded by the voice identification expert, and therefore being less likely to give a guilty verdict. They would, however, have been less likely to give informed answers to the questions on the methodological reliability. This is what is shown in the path analyses: a moderate statistically significant relationship between examination type and reliability indicating that those who read through the qualification-based cross examination gave higher ratings than those in the evidence-focused condition.

4.7.3 FEEBS

Contrary to predictions, there was no support for a relationship between the FEEBSpp subscale and either methodological reliability or verdict choice. In the correlational analysis conducted prior to the path analyses, FEEBSpp had a small positive relationship with ratings of methodological reliability. The direction of the relationship is the same as in study two, indicating that, again, participants with bias towards forensic science gave higher ratings to all three of the types of identification evidence. Interestingly, the size of the relationship was similar across the conditions (.16 for voice identification, and .18 for facial mapping and fingerprints), but close to half the strength of the relationship from study two. The mean and standard deviation scores were similar to that of study two, which suggests that the difference is not due to the different sample in the current study. Additionally, FEEBSpp had moderate positive relationship with pre-trial ratings of the trustworthiness of most of the different types of forensic evidence.

Smith and Bull (2012) created the scale to explain bias that could affect verdict when evidence was weak or ambiguous. By changing the case material, including the crime and location of the evidence, the current study may have reduced the ambiguity in the evidence. However, this would mean that participants considered the fingerprint evidence (taken from the victim's belt) to be as weak or ambiguous as the facial mapping evidence (taken from security footage) and the voice identification (from a phone call). The scale was also previously tested on a murder trial (Smith & Bull, 2012), and sexual assault and robbery (Smith & Bull, 2014), and found to predict perceived strength of evidence in each case. However, there were slight differences in the questions posed to the initial prosecution expert in all three trials. For example, in study two, when asking about validation studies, the transcript, as part of the defence expert's testimony, is as follows:

> *"Similarly, with DNA evidence there is established evidence to enable an expert to say that there is a likelihood of a result occurring in a particular community. Here there is no such statistical justification for the technique nor its application. We do not know how many people have the same shape of chin, or thickness of lips. And we therefore cannot estimate how likely it would be that two people would have the same combination of features. Without that information it is not possible to reliably say that only one person would have that combination of features."*

However, in the current study the discussion about the number of similarities is more straightforward:

> ***Defence:*** *Doctor, can you direct us to any validation studies that support your analysis?*
> ***Expert:*** *Research is always being conducted.*
>
> ***Defence:*** *But do you know of any large scale population studies that actually show the statistical likelihood of combinations of ridge characteristics? How can you*

conclude that there are "enough" similarities? There is no empirical evidence allowing us to calculate the combination of features in a population.

Expert: *I don't know of any large scale studies, but I do trust in the research that has been conducted so far, and in the history of the field.*

Defence: *But do you think that there is sufficient statistical justification for these types of findings?*

Expert: *I think that there is justification through my experience, my experience being asked to testify in court, and through the analysts like me who are employed by the police to do this type of analysis, that is justification. Besides, I am not giving statistical evidence, I just said that there was a high number of similarities. When large scale studies are done I will be happy to present statistical evidence.*

The criticisms given in the scientifically-informed cross-examination section that had previously been part of the defence expert's testimony, as well as the initial questions asked by the cross examining defence lawyer, were rewritten based on feedback to make it more in line with standard questioning. Having a defence expert in study 2 also meant that a proportion of participants saw two experts and had two cross examination sections. This means that the current study transcript, by being shorter, and not doubling the number of experts and cross examination sections for some of the participants may have had the unforeseen outcome of making the criticisms easier to understand and to apply, as well as reducing complexity. Previous studies looking at ELM or HSM have also shown that more systematic processing leads to less reliance on biases (Chaiken & Maheswaran, 1994a). If participants were better able to understand the arguments made by the cross-examining lawyer and were better able to assess the arguments, or, if in study two participants had found balancing two expert opinions more cognitively demanding, the current study may have reduced some of the complexity or ambiguity of the criticisms. This may explain why the FEEBSpp had an effect in study two but not in the current study: bias towards forensic evidence had less of an effect on reliability ratings because there was more systematic processing.

4.7.4 Opposing expert testimony and cross-examination

The current study was also interested in **RQ3:** *Is the effect of cross-examination on juror decisions similar to or different from that of opposing experts?* Although the study did not include a condition where the arguments were presented by an opposing expert instead of by a defence lawyer during cross examination, it is possible to look at the responses given by participants between the current study and the previous two which incorporated opposing expert testimony.

The only study that could be located that measured the effect of cross-examination and opposing expert testimony was conducted by Krauss and Sales (2001), described above. However, the opposing expert in that study conducted their own analyses, either the same type as the initial expert (being clinical or actuarial) or the opposite type, so that the researchers could compare the *types* of testimony. No studies could be found that measured the effect of having lawyers and opposing experts give the same criticisms. However, as the criticisms made by the cross-examining lawyer in the current study were the same as the criticisms given by the opposing expert in study one and two, the pattern of results was predicted to be the same. Whether it was scientifically-informed or qualification-based cross-examination, or opposing expert testimony, the aim is to reduce the credibility of an expert's testimony (Stone, 1995). As such, it should ideally not matter whether the criticisms came from a lawyer or a second expert, only that they educate jurors about unreliable forensic evidence and assist them in making more accurate decisions. In all three studies ratings of methodological reliability had the largest impact on verdict choice, and in only one instance (voice identification conditions) did the condition have a direct impact on verdict. Opposing expert testimony directly affected ratings of methodological reliability in studies one and two, with bias towards forensic evidence affecting reliability ratings in study two. The lack of an effect of examination type or bias towards forensic evidence in the current was therefore unexpected, but may be due to several causes.

First, the difference may be because jurors respond to opposing experts differently to cross-examining lawyers. Participants may have engaged or trusted an opposing expert more than a defence lawyer, and have given more consideration to their arguments than when they came from a lawyer. Complex information can lead to shallower, heuristic processing when presented by an expert (Cooper et al., 1996), and without a second expert balancing the impact of the first, some participants may have trusted in the opinion and reliability of the only expert in the trial.

Second, the qualification-based cross-examination reduced credibility similar to the scientifically-informed cross examination, and participants responded to the methodological reliability questions by giving low ratings to the expert. Although the questions written by Levett and Kovera (2008) ask the reliability of results, participants who read through the qualification-based cross-examination may have considered the information given about the expert's training and qualifications to indicate the methods were not trustworthy.

Third, participants across all three of the different types of testimony recognised that the evidence was vulnerable to bias and was unreliable, and the criticisms presented in the scientifically-informed conditions were unnecessary. This option seems unlikely given that the majority of participants in the previous two studies were unable to recognise that anthropometric facial comparison was unreliable without the opposing expert's testimony, and it is also unlikely that participants would have given fingerprint evidence such a high rating in reliability prior to the trial, but reduced ratings after reading through the expert testimony.

Fourth, there was a reduction in informativeness as well as complexity of the arguments when creating the stimulus materials for the current study. The testimony presented by the opposing expert in the transcript in study one and two was taken from real testimony that featured in *R v Alrekabi* (2007), whereas the questions written for the scientifically-informed cross-examination were based on suggestions taken from a recent guide written by Edmond, Martire, et al. (2014). The guide was written to help lawyers ask questions about the validity and reliability of identification sciences. As discussed in the introduction to this chapter, studies and surveys have suggested that many lawyers will struggle with questioning expert witnesses and typically avoid addressing issues relating to scientific methodology or reliability (Bowers, 2014; Edmond et al., 2009; Freckelton et al., 2001). The questions were chosen because they were relevant to the criticisms identified in chapter one and easy to understand. The responses in the stimulus materials were written so that they could apply to all three types of forensic science. This may have led to a reduction in the complexity of the arguments presented compared to what participants would have experienced in studies one and two, leading to a reduction in impact from bias towards forensic evidence. However, it may also have reduced the overall informativeness of the criticisms. With less information about the method used by the expert being given to the participants, compared to the information given by the opposing expert, there was only a small effect between the cross-examination type conditions.

4.7.5 Sensitivity check

Scores on the FEEBSpp influenced ratings of the odontology evidence presented as part of the sensitivity check. Participants who had more bias towards forensic evidence were more likely to give higher ratings of reliability. While the predicted effect of examination type on sensitivity was not present as there was no support for participants who read through the scientifically-informed cross examination being better able to recognise the flaws in the expert evidence than qualification-based cross-examination, the type of evidence previously seen by the participants did have an effect. Participants who read through the anthropometric facial comparison testimony gave higher ratings to the reliability of the evidence.

Previous studies looking at multiple types of forensic evidence have not been additionally focussed on looking at unreliable evidence, whereas studies looking at ways to educate jurors about unreliable evidence have used single types of unreliable evidence. No predictions were made regarding the impact of having read through different types of forensic evidence. One possible explanation is that the fingerprint analysis, being seen as the “gold standard” of identification techniques (Lieberman et al., 2008), and voice identification, which was potentially seen as a more complicated and technical type of analysis, made odontology seem less impressive or less complex. The effect size was very small, however, and indicates the effect was not very robust.

Scores on the FEEBSpp were also used as a covariate, and produced a significant result. This is not surprising, as FEEBSpp scores also had a positive, statistically significant relationship with reliability ratings given by participants prior to reading through the trial. This intuitively makes sense, as only a brief description of the science was given and it did not give much technical detail about the process, biases towards forensic evidence would be able to influence perception of the reliability of the science used.

The brevity of the article provided to jurors in this study is a limitation, as participants may have been less motivated to attend to the details in it given that it was presented as a disconnected, additional task. However, given that participants were presented with the option of skipping through the task and only 7 out of 204 chose to do so, it does suggest that participants had some motivation or interest in reading the article. Previous studies such as Goodman-Delahunty and Hewson (2010) have presented mock jurors with questionnaires about DNA profiling to measure learning which are capable of testing learning when applied to novel scenarios. Although the effect sizes were small in the current study, future studies should consider using additional tasks to test for sensitivity as this may help researchers understand specific issues in learning or which arguments jurors need assistance in applying.

4.8 Summary

The current study investigated whether cross-examination that focuses on the methodology behind forensic evidence, rather than the qualifications and expert of the expert can inform jurors to unreliable forensic science. The results in this study have been mixed, with few predictions being supported by the data. Using an online format and increasing use of snowballing lead to a more representative sample, although still primarily coming from undergraduate students.

Few previous studies have compared different types of identification evidence, or have looked at different types of cross-examination. Changing the type of evidence from anthropometric facial comparison to fingerprint comparison or voice identification had a large impact on the models produced. However, the ratings given to the identification continued to produce a strong effect across conditions. This indicates that perception of the reliability of forensic evidence plays an important part in how the participants respond to the testimony and develop verdicts.

The type of cross-examination did not have a strong direct effect on ratings of reliability in the main analyses, and only a small effect on the factorial ANOVA, suggesting that criticisms given to the participants may not have been able to educate jurors adequately about how the various threats to validity affect the probative value of the evidence. This is further supported by the lack of an effect on the sensitivity measure. However, comparisons with the previous two studies, as well as conclusions about the effectiveness of cross-examination, are made difficult because of the changes made to the stimulus materials. A potential decrease in the complexity of criticisms, combined with the possibility that voice identification testimony was seen as more complicated than the other two types of identification science, means that future research is needed to untangle the effects of level of informativeness and complexity, and the impact of bias, on perceptions of evidence. Future studies looking at any method of educating mock jurors about identification testimony should consider obtaining ratings of the complexity of testimony as well as ratings of the "experience" of the experts featured in stimulus material.

The introduction of a separate sensitivity measure shows promise for future research, although it will need refining. The article used in the current study was brief, but the descriptions of the flaws with the odontological evidence should have been clear if the reader had paid attention to the reasons behind the arguments given by the cross-examining lawyer in the scientifically informed conditions.

Finally, in continuing to look at effect of cross-examination, and investigating whether bias towards forensic evidence would affect ratings of the reliability of evidence, this study found no support for an effect on verdict or reliability, as participants who gave higher ratings on the

FEEBSpp sale did not have higher ratings of forensic evidence in the path analyses for any of the three types of identification science. Again, this may indicate a difference in levels of complexity of the testimony. The next step in research in this area should consider stimulus materials with different levels of complexity to see if bias towards forensic evidence affects high complexity material, or if the result in the previous study (with a moderate effect size), was due to the particular case or evidence used in the stimulus materials.

5 Conclusion

5.1 Introduction

Jurors are not chosen because of knowledge, skills or personal views, and most laypersons are not equipped to evaluate the complex statistical (Wheate, 2010; Fong et al. 1986; Kahneman & Tversky, 1973), or methodological concepts being presented to them (McAuliff et al., 2009). Jurors are frequently asked to evaluate material they are unfamiliar with, except from fictional depictions, and made to listen to competing, contradictory interpretations (Saks & Spellman, 2016). Judges (Ackland, 2011; Keller, 2011), lawyers (Maricopa County Attorney's Office, 2005), and legal commentators (Goodman-Delahunty & Tait, 2005), have voiced concern that jurors are unable to accurately assess complex scientific testimony.

Yet juries are an integral and valued part of the Australian (and American and British) justice systems (Kirby, 1998), used so that “the law will not be applied in a way that affronts the conscience of the common man” (Devlin, 1979, p. 127). Jurors are rarely given the information needed to “rationally” evaluate expert evidence given in a trial (Edmond, 2015a), and previous research has not thoroughly explored how juror performance would improve if they were given more information about the evidence they are asked to evaluate. The studies presented in this thesis investigated ways to help educate jurors about unreliable forensic evidence, with the aim of seeing which methods worked best, and what differences there were between jurors that would affect their opinions of unreliable evidence.

The overarching goal of this research was to investigate whether scientifically informed opposing expert testimony or cross-examination would educate jurors about unreliable forensic science, and whether there were individual differences that would affect the perception of forensic evidence. Four research questions were considered:

RQ1 Can opposing expert testimony inform jurors about unreliable forensic science?

RQ2 Can cross-examination that focuses on the methodology behind forensic evidence, rather than the qualifications and expert of the expert, inform jurors about unreliable forensic science?

RQ3 Is the effect of cross-examination on juror decisions similar to or different from that of opposing experts?

RQ4 Are there individual differences that will affect jury decision making concerning unreliable forensic evidence?

Chapter one presented an overview of the state of identification sciences and the main criticisms of their reliability, as well as discussing previous psychological literature that applies to jury decision making and forensic science to identify the research gaps. Study one, discussed in Chapter two, established that the perception of the reliability of the forensic science in a trial, anthropometric facial comparison, was the leading variable explaining variance in verdict choice. Chapter three replicated the main finding of study one, as well as identifying that bias towards forensic evidence affected reliability ratings of the facial comparison evidence. The potential for cross-examination to show similar effects was examined in study three, and discussed in chapter four. Study three also compared two different types of forensic science in addition to facial comparison: fingerprint comparison and voice identification.

The final chapter summarises the key findings of the three studies and discusses the limitations of the research. It will also include recommendations for future studies looking at expert evidence. The chapter will end with a discussion of recent legal challenges in Australia regarding the area of identification expert evidence, recommendations for the legal community, and what this research can contribute to future discussion and actions.

5.2 Rating unreliable evidence

One of the primary concerns with identification evidence, whether it comes from CSI Effect literature (Baskin & Sommers, 2010; Maeder & Corbett, 2015; Podlas, 2006b), or from papers on unreliable sciences (Edmond et al., 2009; Porter, 2009), has been that jurors will not be able to evaluate forensic science accurately, and have inflated beliefs in the probative value of this kind of evidence. A primary concern arising from the literature review of previous studies was a focus on verdict choice, culpability ratings, or probability of guilt ratings throughout studies of expert evidence, which not only removed focus from the primary concern of many studies (whether forensic evidence itself had inflated probative value, was too complicated, or overwhelming to jurors), but limited the search for trial or individual factors that might affect the perception of evidence. The use of questions written by Levett and Kovera (2008) allowed for a clear rating of the reliability of research presented during expert testimony, and allowed for more precise hypotheses throughout the studies. Cronbach's α was consistently high between the studies presented in this thesis, as in previous studies using the measure (Jones & Kovera, 2015; Levett & Kovera, 2008; McAuliff et al., 2009), indicating the value of this measure to the research field.

Using a measure of the reliability of forensic evidence was more informative than using verdict alone as a measure of effectiveness. Scores on the measure were the highest predictor of verdict choice, with large effects across all three studies. Asking participants to indicate agreement with the statements provided more information about their perception of the reliability of evidence than looking at verdicts in isolation, or asking for more simple ratings of evidence strength. Although evidence of only one level of validity was presented to jurors throughout the studies, it is clear that using the questions developed by Levett and Kovera (2008) was an efficient and useful way of measuring perception of forensic evidence. Use of the measure's scores in correlational and path analyses was more informative than focusing on verdict or probability of guilt assessments. Also informative were content analyses of the short answer questions asking jurors to explain their reasoning.

5.2.1 Qualitative Analysis

One of the key areas of concern that emerged during the literature review was the dependency of basing conclusions on verdict choice (or culpability decisions) without using other methods of analysis. While this was partially addressed through the methodological reliability measure, the qualitative analysis conducted during the second study shows the value of asking participants to explain their verdict choice and other beliefs about the trial and trial process. Through responding to only a small number of short answer questions, participants revealed that while many were assisted by the opposing expert testimony and/or did not feel overwhelmed by the expert evidence, many others struggled with evaluating the anthropometric facial comparison testimony and made substantial errors in reasoning.

Analysis of the short answer responses confirmed that most participants were not overwhelmed by the expert(s) but engaged with the remaining evidence in the trial. Many participants who read through the opposing expert testimony did understand and engage with the criticisms; however most focussed on one or two criticisms, and many participants across conditions made substantial errors in their reasoning.

The frequency with which jurors mentioned their dissatisfaction with the (perceived lack of) forensic analyses, or the way they were communicated, was concerning. Legal commentators and researchers have been calling for more forensic experts to include error rates in testimony, which includes calling for the types of validation studies necessary to produce those statistics (Koehler, 2017; Martire et al., 2014; Thompson, Taroni, et al., 2003), but some participants responded to questions about error rates and the fact that the expert couldn't declare a match but talked about "similarities" as indications that the analysis was altogether untrustworthy because it could not be definite. This appears to indicate a "backfire" effect – where more information did not enhance the status of the information – but cause doubt about its validity. Other participants wanted impossible or impractical tests done, such as comparing the footage including the machete to one mentioned in the trial, or accused the prosecution of not trying hard enough because there was not enough forensic evidence. Problematic ideas about the powers of forensics or disappointment with the amount, or content, of real expert testimony is a commonly alleged symptom of the CSI Effect. As argued in Chapter One (section 1.6.5), forensic evidence within fictional depictions of forensic or police procedure is typically presented as certain and definitive (DiFonzo & Stern, 2007), capable of individualization (Saks & Koehler, 2008), and not in need of interpretation (Kruse, 2010). In her interviews, Wheate (2006) also noted the frequency with which “missing” forensic evidence sparked conjecture from jurors, and Robbers (2008) found, during the interviews with members of the legal community, a belief that a lack of forensic evidence was seen by jurors as “sloppy police work”. In response to interviewing jurors who were disappointed with a perceived lack of conclusive DNA evidence, Wheate (2006) suggested that more information should be given to the jury panel explaining the small number of forensic samples, and the reasons for limited conclusions being given. The responses made by participants in this trial would also support more information being given to jurors.

As participants were not interviewed or asked directly about the specific criticisms made concerning the evidence, it is not possible at this point to say whether some were understood more than others. McAuliff et al. (2009) suggested that experimenter bias or confounds were harder to recognise than missing control groups, and in the study by Koehler et al. (2016), participants recognised that some evidence had been scientifically tested, but this did not affect verdict. In this study, the varying frequency with which the criticisms were mentioned could indicate that some were more impactful than others. However, this may be a factor of the particular questions asked in the studies. Having the opposing expert present multiple criticisms, although representative of the field as it exists, does not allow us to see whether some of those criticisms needed more explanation or were more confusing to participants. Qualitative analysis, and/or interviewing may help reveal why certain criticisms are noticed or not, and if they are, why they don’t effect verdict.

Many jurors showed they had carefully considered the expert testimony, whether they had seen only one expert or two, but were not overwhelmed by it. Many of the arguments presented in the short answer questions concerned the other evidence in the trial. This supports the findings of Goodman-Delahunty and Wakabayashi (2012), who looked at the spoken content of deliberating groups. The researchers concluded that giving mock jurors more information about scientific methodology led to better understanding by jurors and did not lead to discussion about the differences between the experts dominating the discussion, but facilitated better evaluation of the evidence and the experts' theories.

Howes (2015) has suggested that more interviews or studies involving real-life jurors and other members of the judicial system will be needed to understand the limitations, strengths, and best ways to help utilise forensic evidence. The second study supports the use of qualitative methods of analyses to reveal more about juror perception of forensic evidence, including juror expectations and understanding of methodology. This can be used to inform both expert testing and cross-examination methods of educating jurors.

Recommendation 1*: Investigations of juror decision making should include qualitative analyses as to support quantitative analyses and reveal more about perceptions of trial procedure, expectations of forensic evidence, and interpretation of expert testimony*

5.3 Opposing expert testimony

Goodman-Delahunty and Wakabayashi (2012) argued that while many studies had examined single experts, remarkably few had looked at opposing experts with the goal of understanding how jurors reconciled differences between two experts. The first two studies in this thesis aimed to continue research on the effectiveness of opposing experts educating jurors. The results of studies one and two demonstrate that opposing expert testimony was able to educate jurors about unreliable identification evidence. Participants who read through the opposing expert testimony, which was taken from real expert testimony in *R v Alrekabi* (2007), gave lower ratings to unreliable forensic science evidence than those who only heard the initial expert, indicating they recognised that the method did not protect the examiner from the influence of bias, was not trustworthy, and could not reliably identify suspects. There was no support for a direct effect on verdict. Rather, reduced reliability ratings lead to fewer guilty verdicts. Opposing expert testimony led to reduced ratings on a measure of the reliability of anthropometric facial comparison, and unreliable identification science. This supports the use of an opposing expert to educate jurors about the limitations of forensic identification sciences.

The initial research question asked if opposing expert testimony was able to educate jurors about unreliable forensic evidence. The scepticism effect had been found in previous studies that had featured multiple types of opposing expert testimony: eyewitness research (Penrod & Cutler, 1989) and civil employment discrimination cases (Greene, Downey, & Goodman-Delahunty, 1999), as well as psychological testimony with research based on single studies concerning gender stereotyping and work place harassment (Levett & Kovera, 2008, 2009).

However, one of the limitations of the studies has been that there was only one "level" of scientific validity presented in the stimulus materials. As highlighted in the NAS report, with the exception of DNA profiling, all types of identification sciences lack the appropriate validation studies, proficiency testing, and protection from bias that would ensure the forensic evidence produced is reliable (National Academy of Sciences, 2009). For many disciplines and agencies, work is being done to establish procedures and standards (Haber & Haber, 2008; Peterson et al., 2003), and the recent PCAST report recommended that, in the US, the FBI laboratory should undertake a large scale research program to improve forensic science (President's Council of Advisors on Science and Technology, 2016). However, as the science behind identification evidence currently stands, the six criticisms outlined in chapter one will still apply both to traditional forms of identification evidence, and to newer, emerging disciplines. As jurors place a large amount of faith in forensic evidence (Garrett & Mitchell, 2016; Lieberman et al., 2008), the goal of the studies in this thesis was to research ways of more accurately presenting the fields of forensic science so as to reduce unrealistic faith and mistaken beliefs in the validity of the science, which should result in more accurate assessments of probative weight.

But by not presenting some participants with a more "valid" version of the identification evidence, it was not possible to test, or argue, for sensitivity. According to the categorization of the effects of eyewitness testimony set out by Cutler et al. (1989), to argue that a method of educating jurors had led to sensitivity, researchers would need to show that there were different responses to different levels of validity. This would indicate that, across conditions, the educational intervention had helped jurors recognise when eyewitness testimony was valid and when it was not. Future studies should consider providing jurors with versions of expert testimony where the evidence has greater validity. Potential differences in testimony could include having a second expert examine the results, or having blind testing of multiple "suspect" samples, as these criticisms can apply to individual examiners, rather than the field as a whole. Alternatively, future studies could follow Koehler et al. (2016), who in some conditions, since jurors are likely to have little knowledge about which techniques have been tested or not, simply told participants the techniques used had been "subjected to a great deal of scientific testing" (p. 4). The potential for opposing expert testimony or cross-examination to sensitize jurors, can then be directly assessed.

Recommendation 2*: investigate whether opposing expert testimony can sensitize jurors to the validity of forensic evidence by comparing ratings of methodological reliability of forensic evidence, and verdict, with stimulus materials of differing validity*

Experts who appear as "opposing" experts within a trial typically come from the same or similar field to an initial expert. They are therefore likely to be well educated about the research, techniques, and general acceptance of methodology. However, many trials are not able to use opposing experts (Cordner, 2015; Giannelli, 2007). The effectiveness of cross-examination in delivering the same criticisms was examined in the final study.

5.4 Cross-examination

Legal commentators such as Findley (2008), Garrett and Neufeld (2009), and Edmond, Martire, et al. (2014), have argued that lawyers typically avoid presenting arguments about forensic methodology, and instead focus on qualifications, experience in the field, and conflicts of interest of an expert. These arguments fail to educate jurors about the limitations of forensic evidence, and therefore leave jurors vulnerable to overestimating the scientific validity of identification evidence.

Study three of this thesis compared qualification-based cross-examination with scientifically-informed cross-examination, which presented jurors with the same criticisms against typical identification evidence methodology that had been argued by the opposing expert in Studies One and Two. It was hypothesised that the same relationships between the scientifically-informed cross-examination, methodological reliability measure, and verdict would be found across the three types of identification science, but this was not supported by the results: scientifically-informed cross-examination did not lead to differences in reliability measures compared to qualification-focussed cross-examination.

Three path analyses and multi-group comparisons were used to examine the effect of the two types of cross-examination on three types of forensic identification evidence. The multi-group analyses showed no support for a variance in the relationship between scores on the FEEBSpp and methodological reliability, nor between examination type and ratings of methodological reliability. There was a variation between groups in the relationship between examination type and verdict. Investigation of the individual path analyses showed that in the voice identification condition, the qualification-focussed cross-examination led to lower ratings of methodological reliability than did the scientifically-informed cross-examination. This was not predicted at the outset of the study.

A limitation of Study Three was there was no measurement of whether jurors considered the expert to be highly qualified and experienced, or whether they appeared inexperienced, especially after the cross-examination. During their interview with jurors who had served on civil trials, Ivkovic and Hans (2003) found that many jurors were highly attentive to the qualifications and credentials of the experts who had testified before them, and could recall specialisations, research, and professional history. Although previous studies had found that decision-makers faced with complex evidence were likely to use experience as a heuristic in judging the credibility of an argument (Chaiken & Maheswaran, 1994a; Cooper et al., 1996; Hovland et al., 1953), it was not predicted that the qualification-focussed cross-examination would affect either ratings of reliability or verdict. The results suggest that in the voice identification condition participants used qualifications or experience as a heuristic for deciding verdict, but not reliability. The cross-examination that focussed on reducing the credibility of the expert by suggesting that they did not have enough training or experience testifying about forensic evidence and led participants to give fewer guilty verdicts. However, the next step would be to test to see if this pattern appears for other types of forensic sciences that laypersons have little knowledge of, and to compare different levels of qualifications and different levels of validity.

Recommendation 3*: Follow up studies looking at cross-examination should compare different levels of qualifications or experience with different levels of validity of evidence*

From the point of view of lawyers wishing to argue against forensic evidence, particularly its inculpatory power, the use of qualification-based cross-examination to reduce the impact of forensic science could be seen as a positive finding. However, from a wider based view of the legal system it is a negative: jurors should be basing their verdict decisions on weighing the evidence, not the experts. Qualification-focussed cross-examination is not informing jurors about the limitations of forensic evidence, and therefore cannot lead to accurate assessments.

5.4.1 Comparisons between Opposing Expert Testimony and Cross-Examination

The relative effectiveness of the methods presented can be compared by using weighted scores. In Studies One and Two, the total standardised weight between the opposing expert presence and verdict was .22 and .27. In Study Three the total weight of the relationship between the cross-examination conditions and verdict choice was only -.08. However, the total standardised score between methodological reliability and verdict was -.60 for studies one and two, and -.67 for the third study. In all scenarios, the identification evidence was the strongest inculpatory evidence in the trial, so it is reasonable that ratings of the reliability of the evidence should explain a large portion of the variance in verdict choice. However, the differences between the identical scores of the relationship between opposing expert and methodological reliability in Studies One and Two, and the much lower score between cross-examination type and reliability in Study Three, suggests that in the later study participants based their reliability scores on other, unmeasured, sources of information. Participants may have not considered the arguments as systematically when they came from a cross-examining lawyer, and based their ratings on other information sources (such as prior knowledge or ideas). The next step in this area of research will be to directly compare the two methods.

One limitation of the studies is that the path models did not account for prior ideas about the reliability of the different types of forensic sciences. The initial question in study was only brief, and therefore does not capture participants' perception of the reliability of evidence. Although the questions written by Levett and Kovera (2008) ask about specific evidence related to a case, a corresponding series of questions asking participants about methodology more generally could inform researchers about general perception of evidence, as well as be used as prior measures, e.g. "The methods used by prosecution experts to identify suspects are based on good scientific principles". Additionally, there is a need to understand how much reception of the arguments is affected by jurors' belief in the (im)partiality of the opposing expert, versus the partiality of a cross-examining lawyer.

Recommendation 4*: In order to investigate whether participants place more value, or respond differently to criticisms of forensic evidence when they come from an expert compared to a lawyer, future studies should directly compare opposing experts and scientifically-informed cross-examination.*

As discussed in Chapter Four, differences between the impact of opposing expert testimony and scientifically-informed cross-examination may have been affected by differing levels of complexity in the stimulus materials – for example, in Studies One and Two the discussion around the number of points required to confirm similarities is condensed and uses more technical language: "Not in that sense because it's not a match of quantity, it's a match of quality. Whether you have a lot of points but yet none of them uniquely or distinctly characterizes that person it would weigh less than if you only have a small number of points but those are strong points, if they were unique identifiers." In the third study the discussion goes back and forth between the expert and defence lawyer, making it easier to read and making the language less formal:

Defence: But if we had a different expert from your field in that chair right now, they would probably be using a different technique?

Expert: It is possible.

Defence: Would they also have made the same number of matches as you or made the same measurements?

Expert: I would hope so, if their technique was sound.

Defence: But you can't say for sure?

Expert: No. But while the technique might differ in some aspects, the fundamentals of what we do are always the same: isolating small samples for comparison, and checking them.

This underscores the need to do further investigation, but it also emphasizes the importance of carefully considering the trial materials given to participants. It has been a common complaint against simulated jury studies that participants are not given materials that approach the length or complexity of a criminal trial (Bornstein et al., 2017), and even in Studies One and Two where testimony was taken from that given by an expert in a real case, it still had to be edited to make it shorter and to remove filler words, hedge words, or moments where the expert and lawyers or judges clarified the meanings or details given in the testimony, which could influence the way the testimony influenced the participants. However, the differences between the length and format between the studies in this thesis were very small in this case, with the exception of who was providing the arguments (lawyer or expert), and the complexity, with the opposing expert potentially having more difficult arguments, and the voice identification expert having more technical descriptions. Again, this should be investigated in future studies of forensic evidence because if differences can emerge from the way the stimulus materials are written, then the size of effects of previous studies also needs to be replicated using different types of expert evidence, and different descriptions of methodology (which requires more researchers to make the stimulus materials used in research more readily available). Additional attention should also be paid to what other heuristics or biases may be affecting participants who find the material complicated. The materials used in this study were based upon real cases – however, the complexity of the process, environment and procedures used in real trials is very difficult to replicate.

Recommendation 5: *Previous findings within the area of jury decision making research need to be replicated. Studies on forensic evidence need to make stimulus materials available.*

5.4.2 Differences between types of forensic sciences

Despite the importance of forensic evidence to criminal trials, and the number of papers expressing concern over the CSI Effect and jurors' use of forensic evidence, very few studies have compared different types of forensic evidence. Initial concern about the validity of forensic science left none of the disciplines free from criticism – yet there are clear differences in the perceptions of different forensic disciplines within the community. Study Three investigated differences between three different types of forensic sciences to see if the strength or relationships would differ. Previous studies had indicated that physical identification evidence was seen as more persuasive and more trustworthy than other types of evidence, such as victim testimony or eyewitness identification (Lieberman et al., 2008), although there may be some small discrepancies between the beliefs of lawyers and laypersons, and what lawyers think laypersons will believe (Garrett & Mitchell, 2016).

Study Three used fingerprint, anthropometric facial comparison and voice identification evidence. Comparisons between these models shows that although there was no statistically significant difference in the fit between the models, the actual paths within the models were quite different. However, the weight between methodological reliability and verdict remained moderate-to-large across all three. Changing the evidence type from fingerprints, to facial mapping, to voice identification, lead to large differences in relationships between variables in the path analyses.

In their studies of eyewitness expert evidence when attempting to elucidate when and why sensitivity would occur, Cutler et al. (1989) and Cutler, Penrod, and Dexter (1990) began by testing multiple invalid witnessing and identification conditions (including masking cues to hair and hair line, weapon focus, and retention intervals). They investigated wh`ether this would desensitize jurors to witness confidence. In follow up studies, the focus was on sensitizing jurors to two factors affecting line-up suggestiveness (foil bias and instruction bias, Devenport & Cutler, 2004; Devenport, Stinson, Cutler, & Kravitz, 2002). Sensitization occurred only in the earlier study (Devenport et al., 2002), with Devenport and Cutler (2004) admitting to having no suggestions for why the discrepancy occurred as the same materials and manipulations were used. While this shows the importance of replicating studies, it also shows that laypeople may have very different reactions to different criticisms of the credibility of testimony.

The comparison between studies looking at the use of experts to explain eyewitness evidence reliability and the use of experts to explain forensic evidence reliability also reveals another sharp distinction: a large amount of literature has investigated what laypersons know about eyewitness factors. Desmarais and Read (2011a) were able to conduct a meta-analyses of 23 surveys looking at lay knowledge, comprising of 4,669 respondents. Through this meta-analysis it is possible to see that, of the types of factors Cutler, Devenport and Penrod were studying, laypersons had varying degrees of agreement with experts about whether it would affect eyewitnesses (e.g. weighted agreement for weapon focus was 52%, line-up instructions was 75%). If surveys were conducted to examine lay knowledge of factors influencing forensic evidence comparisons, researchers would be better able to see which topics needed the most attention, so that educational methods could target those misconceptions. The National Science Foundation has argued that “The amount of information now available can be overwhelming and seems to be increasing exponentially. This has led to “information pollution,”(2000), which includes the presentation of fiction as fact. Thus, being able to distinguish fact from fiction has become just as important as knowing what is true and what is not.” It is clear that many jurors have exaggerated perceptions of the forensic scientific process across all fields and types of comparisons. Better understanding of what constitutes reliable science by jurors is one of the best defences against unreliable evidence having an undue influence on verdicts.

5.5 Individual Differences and the Interpretation of Forensic Evidence

Reviewing the previous literature found that while many studies had looked at stable individual characteristics that could affect jury decision making, there have been largely inconsistent results (Ask, Reinhard, Marksteiner, & Granhag, 2010; Smith & Bull, 2012). Devine and Caughlin (2014), in the conclusion of their meta-analytic review of individual characteristics, called for more studies to investigate juror attitudes towards the legal system as well as measures of psychological characteristics. These studies have contributed towards that end by measuring two general cognitive characteristics, epistemological sophistication and argument skill, and a measure of bias towards evidence.

In Study One, although it was predicted that higher argument skill and higher scores on a measure of epistemological sophistication would lead to lower ratings of methodological reliability, and to fewer guilty verdicts, there was no support for a direct or indirect relationship between the individual differences and either dependent variable. Previous studies had found that argument skill had been associated with evidence coverage and comprehension (Weinstock & Flaton, 2004), and better reasoning skills (Kuhn, 1991; Kuhn et al., 1994). Epistemology, which influences the way an individual argues, weighs and judges evidence, and makes decisions (Hofer, 2002), was also predicted to influence the way mock jurors would respond to competing argument claims. As argued in Study One, the lack of support may be because the highly educated sample or the coding scheme used to evaluate the arguments made by participants and the measure used to present a score of epistemological sophistication may have limited the chance of showing a significant relationship. There is no support for a relationship between argument skill and epistemology on ratings of the reliability of identification evidence or on verdict.

The second study turned to measuring individual differences that would specifically affect perception of evidence, rather than jury decision making more generally. Tyler (2006) has alleged that the "The CSI effect has become an accepted reality by virtue of its repeated invocation by the media" (p. 1083), dominating discussions of juror capabilities. Empirical studies, however, have found very mixed results that depend on study design, questions asked, and differing definitions of what makes a "CSI viewer". There is limited evidence of an effect of watching forensic-type TV shows that can be directly measured by viewing habits, as the wider "tech effect" and immense corpus of forensic material (including more general crime TV shows, documentaries, books etc.) make measuring viewing habits and the definition of a heavy/frequent viewer almost redundant. There is support for the idea that some jurors have high expectations that forensic evidence will appear (Kim et al., 2009; Shelton et al., 2006), and some believe themselves to be capable and accurate assessors of that evidence, which will affect reception of expert evidence (Hayes & Levett, 2013). However, given that there is limited effect of frequency of viewing on verdict choice (Maeder & Corbett, 2015), attention should turn towards operationalising the “CSI Effect” in different ways. The studies presented in this thesis did not study the “CSI Effect” directly: participants were not asked about viewing habits, but took the approach of understanding expectations about forensic science through a scale measuring problematic attitudes.

The FEEBS scale was created to measure unrealistic expectations of forensic evidence that are captured in the CSI Effect (Smith & Bull, 2012, 2014; Smith, 2011). In Study Two, scores on the pro-prosecution scale of the FEEBS had a moderate effect on ratings of methodological reliability, and examination of the standardised total effects (Table 4, section 1.3.6) showed an overall moderate effect on verdict. However, in Study Three, the relationship between scores on the FEEBSpp and methodological reliability did not reach statistical significance for any of the three types of forensic science, and had small overall effects on verdict (Tables 7-9, section 1.6.7). Bias towards forensic evidence influenced the perception of the reliability of evidence with a moderate effect size in Study Two, but had no statistically significant effect in Study Three for any forensic evidence type, indicating that there may be substantial limitations on when bias does and does not affect perception of evidence.

Prior to the studies in this thesis, the scale had been tested using stimulus materials featuring murder, sexual assault and robbery. By measuring scores on the FEEBS in a different sample (in a different country), and finding a relationship with the perception of different types of evidence, this thesis has provided further support for the use of the FEEBS as a way of measuring the problematic juror beliefs. It also indicates that bias can affect perception of unreliable evidence in some circumstances, although further research is needed to explore if there are types of forensic identification sciences that are more susceptible.

There were several differences between Studies Two and Three: the trial material changed, moving from a robbery to a murder, and the details and non-forensic evidence changed as well; cross-examination was used to educate jurors rather and opposing expert testimony; and the amount of detail, and format, of the descriptions of forensic evidence changed. There was no direct measure of how complicated participants considered the evidence to be, nor was there a comparison of the stimulus materials between studies. The Elaboration Likelihood and Heuristic-Systematic models predict that when motivation or ability to process a message was reduced, processing would be shallower, leading to reliance on heuristics and increased susceptibility to biases (Chaiken & Trope, 1999; Petty et al., 1983). Opposing expert testimony, which may have been more complicated, compared to cross-examination, may have led to increased reliance on heuristics or increased effect of biases. In light of such a distinct difference between the results in this thesis, future studies should consider the role of the complexity of the material they are presenting to participants. However, it will also be important to replicate this study using different case materials where the forensic evidence is more or less the focus of the trial, and to look at different crime types.

Recommendation 6*: Future studies should include the Forensic Evidence Evaluation Bias Scale in future studies looking at expert evidence but consider manipulating materials so participants are presented with different levels of complexity*

In interviews with South Australian jurors, along with the chance to experience skilled cross-examination and the atmosphere and drama of a trial, the chance to "hear scientific/forensic evidence.… the opportunity to participate in the trial and see the myths of TV dispelled, and having the opportunity to gain a better understanding of the legal system and the trial process” were highlighted as potentially being the most enjoyable part of a trial (Goodman-Delahunty et al., 2008). While empirical literature shows that many laypersons may struggle with accurately assessing forensic testimony, and may be influenced by biases, these interviews confirm that mock-jurors are concerned with making accurate decisions.

5.6 Limitations of the research

While the research in this thesis has taken steps towards answering the initial research questions, there are several limitations to be considered when drawing conclusions from the results.

5.6.1 Complexity of stimulus materials

As was discussed in Chapter Four, there may have been discrepancies in the level of complexity within the stimulus materials. Research using dual-processing models shows that complexity can lead to shallower processing of materials, and can increase the impact of biases in evaluating arguments (Chaiken, 1980; Chaiken & Maheswaran, 1994a). There is no standardised way of evaluating the “complexity” of expert testimony. However, it is likely that the differences in the difficulty of stimulus materials used in both the studies included in this thesis as well as those previously looking at forensic evidence will have influenced whether participants used more effortful or shallower processing, and the extent to which heuristics influenced decision making. It is possible that many of the discrepancies in findings are due to the studies having materials that encouraged different levels of cognitive processing. Although this should be investigated in future studies, it should also encourage the replication of previous research using different materials.

5.6.2 Written format vs. real actors or videos

A written format allowed for careful control of the stimulus materials. However, it reduced the ecological validity of the studies. In a study testing differences in trial presentation modality (transcript or video), there were no significant differences in guilt ratings (Pezdek, Avila-mora, & Sperry, 2010). However, participants did rate the testimony of the eyewitness expert featured in the trial as having more impact, being more useful, and more influential in the written transcript condition in comparison to those seeing the video. The researchers did point out this may have been because participants had time to re-read sections of the expert testimony in the transcript condition, but ultimate argued that the *content* of the testimony of an expert was more influential on guilt ratings than communication style. This suggests that having written testimony from the experts will have had minor influence on participants' judgements of the reliability and verdict.

The previous studies of expert evidence have shown that mock jurors will rate experts based on a wide range of extra-legal cues, not limited to gender, warmth, confidence, likeability, and even the amount of eye contact made with jurors (Brodsky, Griffin, & Cramer, 2010; Cramer, Harris, Fletcher, DeCoster, & Brodsky, 2011; Neal & Brodsky, 2008; Neal, Guadagno, Eno, & Brodsky, 2012). All of these cues have the potential to influence juror perception of experts within a courtroom, but are absent when the only representation of an expert is through a written transcript. While the focus being placed on the content of the materials is an advantage when looking specifically at the perception of evidence, it does mean that the evidence was evaluated without the normal cues jurors would be capable of using to judge expert competence.

5.6.3 Deliberations

The analysis was restricted to examining the responses from single jurors, which is unfortunately common in jury research: few studies looking at forensic evidence reliability include a jury deliberation section. Kalven and Zeisel's (1966) classic study of jury decision making found that the distribution of individual jurors' verdicts prior to deliberation was able to predict the post-deliberation, "in nine out of ten juries" (p. 488). As Ellsworth (1993) points out, this focussed researchers' attention on how pre-deliberation verdict choice was made, but very little is known about the study or the method of collecting data, as it is only referenced in a footnote.

Other studies have found differing rates of jury panels where the verdict of a minority at the beginning was able to dominate at the end of deliberation, with results ranging from 10% (using a combination of mock and actual jurors, Devine et al., 2001), to 5% (using actual juror trials, Hans, Hannaford-Agor, Mott, & Munsterman, 2003). The latter study also found there was an asymmetrical balance in the hung juries: if the minority at the beginning favoured acquittal, they were more likely to prevail. There is also an imbalance in who does the most talking during a deliberation. Ellsworth (1993) found in her study that four jurors out of the panel of twelve took up more than half of the deliberation time: the foreman spoke more than any other juror, men spoke more than women, and jurors with more "worldly success" (a combination of social status and education, see Hastie et al., 1983) talked more than other jurors. What would happen if jurors who took up more of the discussion time had better awareness of the unreliability of the forensic evidence?

A small number of people recognizing flaws could make a great difference in a deliberation. Although individual jurors may hold unreasonable biases or ideas about forensic evidence (particularly unreliable types), there is support for the idea that deliberation may act as an educational tool in itself. Deliberation can help correct errors, focus on relevant information, and attenuate biases (Bourgeois, Horowitz, ForsterLee, & Grahe, 1995; Studebaker & Penrod, 1997). In one study, deliberation led to the effects of clearly one-sided expert testimony disappearing (Brekke, Enko, Clavet, & Seelau, 1991). This indicates that deliberation may be able to address some of the inaccurate decision making brought on my heuristic processing. A small number of jurors who have a better understanding of scientific research (or just a better understanding of the testimony) may be able to help other jurors understand problematic elements, just as it has been shown to improve understanding of mitochondrial DNA (Dann, Hans, & Kaye, 2007), and statistics (Kaasa et al., 2007).

However, studies show that if biases are held by a majority then deliberation can also accentuate them (Hinsz, Tindale, & Vollrath, 1997; Kramer, Kerr, & Carroll, 1990). Ruva and Guenther (2015) argue that pre-trial publicity (PTP) can affect both individual jurors and juries. In a two part trial including quantitative and qualitative examination of deliberating juries, negative-PTP jurors were more likely to find defendant guilty and as less credible than those seeing no PTP. Those who saw the PTP also misattributed more facts given from the PTP to the trial, and were more confident in those attributions. An examination of the content of the deliberations found that PTP-exposed jurors also discussed those facts without crediting the source.

Future studies should consider utilizing the type of study conducted by Goodman-Delahunty and Wakabayashi (2012) to investigate the impact deliberation has on bias towards forensic evidence or whether it can help jurors evaluate evidence more accurately. Nolan (2003) has suggested that research into real juror deliberations may improve understanding of jury performance, as well as allowing appeal courts insight into whether a panel was influenced by anything that has led to a miscarriage of justice.

5.6.4 Representativeness of the samples

The final limitation to be addressed is the (un)representativeness of the samples. It has been almost two decades since Bornstein (1999) conducted a review of jury research, arguing that there were few difference between studies using student samples and those using more representative participant group or different trial presentation format.

Bornstein et al. (2017) reviewed 53 studies in a meta-analyses, testing for differences between student and non-student samples on guilty verdicts, culpability, and sentencing in criminal trials, and liability verdicts, continuous liability and damages in civil trials. There was no support for a difference in verdict, culpability ratings or damage awards. The two liability judgments showed significant differences, but were in opposite directions, and sentencing had a small non-significant difference but a very small effect size (studies published after 1994 had d = 0.01). However, the researchers found that students voted guilty more often than non-students for written trials ($d = 0.06$), but not for other media. In this thesis, there was a slight difference in verdict choice between those accessing through the undergraduate and public portals in Study Two (section 3.3.1), but subsequent analyses also found differences in age (which was correlated to FEEBSpp scores and ratings of methodological reliability). This further underscores the need to research what laypersons know about forensic science and how this affects verdicts. It may be that age and education act together to produce differences in expectations regarding forensic science.

The third study looked at the distribution of highest education level achieved before combining the samples, and found that the only differences were in age (which was to be expected) and ratings of anthropometric facial comparison. Devine and Caughlin (2014) found little support for an effect of education level on mock-juror decisions in their meta-analyses. However, given the small numbers of studies that have looked at forensic evidence and have included personal characteristics, this may not preclude an effect from existing, it may just indicate that it has only been studied in cases where it would make no difference.

In his early review of jury simulations, Bornstein (1999) supported following Diamond's (1997) recommendation that researchers using simulation methods (particularly those using student populations and written transcripts) should be considered "Stage One" research, to be followed up by more comprehensive studies that recruit representative laypersons, and use videotaped trials as stimulus materials. Although the limitations of the studies in this thesis do not detract from the overall findings, they should be considered Stage One studies, with future studies following the recommendations made above.

5.7 Contribution to jury decision making research

This thesis has made several notable contributions to jury decision making research concerning both forensic and unreliable evidence, and ways to educate jurors.

Krauss and Sales (2001) argued that the mechanisms underlining the way expert testimony would affect judgement, then verdict, had been under-researched and were not well understood. Most research into expert witnesses has looked at social framework testimony (Hurwitz et al., 1992; Vidmar & Schuller, 1989), which includes eyewitness research, or has asked jurors to consider research taken from single studies (Levett & Kovera, 2008, 2009; McAuliff et al., 2009). The studies in this thesis have added to the area by going beyond these types of expert evidence to examine juror responses to conflicting expert testimony based on comparison matching. Prior to the completion of this thesis, only Eastwood and Caldwell (2015) had looked at forensic evidence with the goal of studying the best way to educate jurors. This thesis has highlighted the need to continue studying juror perceptions of identification comparison evidence, including the frequency and type of mistaken ideas about methodology and the role of the examiner.

Two-and-a-half decades ago, Ellsworth challenged researchers to look for individual differences between jurors that would help to study the mechanisms behind decision making (1993). More recently, Cole (2015) has argued that the CSI Effect gained a large amount of interest in the media, and this lead to legal commentators and researchers looking into the effect. As discussed in Chapter One (section 1.7.5), there has been very little empirical evidence to support the operation the CSI effect. The use of the FEEBSpp subscales throughout studies two and three show that measuring the ideas behind the CSI Effect – that jurors would have exaggerated faith in inculpatory forensic evidence – is more informative as a measure of individual differences, as it may directly affect perception of the reliability of evidence. By continuing the work of Smith and Bull (2012, 2014), and finding that bias towards forensic evidence affects the perception of an unreliable science, this thesis has contributed to the search for mechanisms affecting juror decision making, and the investigation of extra-legal decision making factors.

Finally, Studies One and Two have shown support for the use of opposing expert testimony as a way of educating jurors about the limitations of unreliable forensic sciences, whereas Study Three has shown that cross-examination, even when scientifically informed, may not have the same effect. However, this thesis has also demonstrated the need to examine the particular arguments presented by the expert or cross-examining lawyer to locate which need a higher amount of knowledge of scientific methodology, and whether laypersons are able to learn about this within a court room setting.

5.8 Expert evidence: where are we now?

Beginning in 2012 in the US, the FBI reviewed over 3,000 criminal cases where hair comparison analysis had been used (President's Council of Advisors on Science and Technology, 2016). The results, which were released in 2015, half way through the completion of this thesis, indicated that where the testimony was used as incriminating evidence, it was invalid over 95% of the time. The PCAST report claims that as of March 2016, there was an intention to look at other forensic science comparison methods as part of its review. The PCAST report also made a number of recommendations, including performing ongoing studies of the scientific validity of current and newly developing identification techniques; having the FBI undergo a research program to improve forensic science; the direction of attorneys so that expert testimony met relevant standards; objective methods for DNA analysis, fingerprint comparison, and firearm analysis should be developed; and that Federal judges should take into account the foundational validity, "that testimony is the product of reliable principles and methods" (p. 145), and the validity as applied "that an expert has reliably applied the principles and methods to the facts of the case" (p. 145).

Edmond, Martire, and San Roque (2017), as part of a review of the use of expert reports concerning forensic evidence, reviewed the recent case of *JP v DPP* (2015, "*JP v DPP").* The case involved a minor who was charged with an aggravated break and enter, after a latent print examiner from NSW attributed a single fingerprint taken from the scene as coming from JP's thumb. The case is noteworthy because the expert examiner, and the certificate relied upon by the prosecution, were raised in an appeal (which was unsuccessful). The commentators point out that the while the appeal in *JP v DPP* was not the first time someone had challenged admissibility of fingerprint evidence, it did address recent commentary about forensic analyses and unreliable methods (Edmond et al., 2017). The certificate was challenged as it did not provide enough information about the method of analysis, or the procedures and standards that were applied, and as Edmond et al. (2017) point out “the certificate provides very little insight into what was done, what procedures were used and what standards applied. There are no references to limitation or uncertainties, no recognition of even the possibility of error, and no discussion of controversies” (p. 603). Justice Beech-Jones, the appellate judge, however, said that the oral evidence given by the expert made up for the inadequacies in the report, and accepted that the expert had not read the material the defence counsel referred to in regards to the potential for bias. Worryingly, the authors point out that “the magistrate treated defence concerns about methodological limitations, exaggerated expression and bias as hypothetical issues. Rather than require the Crown to demonstrate that its routine procedures protect examiners from scientifically notorious dangers, there is an expectation that the defence will identify actual mistakes and errors” (Edmond et al., 2017, p. 621). In light of the findings of this thesis, this is particularly concerning. Participants in the third study were not educated by the cross-examination, so if this had been in front of a jury trial the defence’s questioning may have had limited effect on jurors’ perception of the reliability of the evidence, particularly as fingerprints are seen as so trustworthy already. There is a need for the legal community to address issues at the admission stage, as well as to work on ways to educate jurors. However, the questions asked by the defence lawyer indicate that there may be growing awareness of the problems with methodology and reliability, and willingness to challenge forensic evidence, rather than attacking qualifications and experience.

Admissibility standards have had little impact on excluding unreliable evidence, and therefore has had little effect on shaping the quality of evidence reaching the court (Edmond, Cole, et al., 2014). While commentators, such as Edmond, Hamer and Cunliffe (2016), are not proposing that fingerprints, and other forms of forensic identification evidence, be kept out of courts, they argue instead that the “limitations and uncertainties” about the evidence be communicated to the triers of fact. They also point out that Australian courts are yet to work out a way to obtain independent advice on issues concerning contested forensic science.

Finally, more and more researchers and commentators are calling for additional effort to be made into standardising forensic examination procedures (Stoel, Dror, & Miller, 2014). In particular, identification analysts should be exposed to as little domain-irrelevant information as possible (Dror, 2012; Miller, 1984; Stoel et al., 2014), with several researchers now recommending the use of "sequential unmasking" as a way of limiting exposure to potentially biasing materials (Dror et al., 2015). Stoel, Dror, and Miller (2014) report that this has already been trialled in the Netherlands Forensic Institute in the firearm identification department, and has been rolled out to examiners looking at handwriting identification. A similar context management system has already been implemented in Australian in the Victorian Police Forensic Services Department, (Found & Ganas, 2012).

Together, these challenges and reports indicate a growing awareness of the risk of unvalidated and unreliable evidence, and the dangers of allowing that evidence into court. Instead, they currently place their faith in systems that do not ensure that jurors will hear all the relevant information about evidence, and do not help to engage with mainstream scientific knowledge (Edmond et al., 2016).

5.9 Recommendations for the legal community

In the PCAST report, the authors argued that evidence based on feature-comparison was particularly dangerous because: 1) "The vast majority of jurors have no independent ability to interpret the probative value of results based on the detection, comparison, and frequency of scientific evidence"; and 2) "The potential prejudicial impact is unusually high, because jurors are likely to overestimate the probative value of a "match" between samples." (p. 43). This has been clearly demonstrated throughout the studies in this thesis.

From the results of the first two studies, there is support for the use of opposing expert testimony in educating jurors about threats to reliability across the field of unreliable forensic sciences. Jurors were not overwhelmed by conflicting expert evidence. However, the design of the studies preclude making claims about the ability of opposing experts to sensitize jurors. The results also confirm that without assistance jurors make substantial errors in reasoning, and many have difficulty interpreting arguments and criticisms concerning forensic methodology.

Opposing experts may assist by informing jurors why tests were carried out and how, and may, by being more informative, lead to better processing of arguments.

There was limited support for the effectiveness of scientifically-informed cross-examination, even when using questions suggested by Edmond, Martire, et al. (2014) that directly asked the expert about the main criticisms facing identification sciences. Given that multiple surveys have shown that lawyers struggle with adequately presenting a defence against forensic expert testimony, and may not be able to provide as critiques that are as comprehensive as those written by an opposing expert (Bowers, 2014; Cashman & Henning, 2012; Freckelton et al., 2001), this suggests that opposing expert testimony is a more effective way of informing jurors about unreliable sciences. Qualification-focussed cross-examination was able to reduce guilty verdicts with one type of science. However, this is clearly not the ideal type of questioning as jurors would not be being informed about the methodological limitations. More research will be needed to investigate why the predicted effects did not occur in the cross-examination scenarios, particularly in studying the effects of complexity of testimony and responses to cross-examination. For these questioning techniques to work – they must be targeted towards revealing more about the tenants of validity so that jurors can evaluate the evidence effectively.

The effect of bias towards forensic evidence, as found in the second study, and the initial ratings of evidence in the third, are also of concern. Although in Australian jurisdictions, no voir dire of juries is conducted (Goodman-Delahunty et al., 2008), legal practitioners, particularly those preparing a defence against forensic identification evidence, should consider ways to address the problematic beliefs held by jurors within a trial. Participants held unrealistic expectations towards forensic evidence, and were disappointed by the amount of forensic evidence presented in the trial, with many believing more should have been done. Examination of criminal trials shows that the length of a jury trial has increased ten-fold in the past fifty years (with the average criminal trial now 2 weeks or more), and the content is becoming more complex (Dunford, 2004). A survey of 136 judges in Australia found that 54% let jurors ask questions, although only 43% instructed them on how to ask (Ogloff et al., 2006). Over-taxing jurors by presenting them with fast paced, complex evidence could lead to reliance on heuristics, but measures like note-taking can ease the strain and increase recall and comprehension (Hope, Eales, & Mirashi, 2014).

Finally, the research presented in this thesis emphasises that there are glaring holes in what we do know about layperson beliefs about forensic evidence. While there exists a vast amount of literature concerning the CSI Effect and lamenting the prejudicial effect of fictional depictions on juror beliefs (Cole, 2015; Maeder & Corbett, 2015), there is a troubling lack of research looking at what it is that jurors actually believe. Very few studies have asked basic questions about the perceived strength or reliability of evidence (Baskin & Sommers, 2010; Lieberman et al., 2008). At the moment, we do not know enough about how triers of fact will use the information about forensic evidence that they will receive in court (Koehler, 2017). Before we can refine educational measures designed to teach jurors about forensic methodology, resources should be committed to studying what laypersons know currently about forensic sciences and about scientific methodology.

5.10 Conclusion

As scientific innovations grow further away from the domain of general knowledge, jurors are more likely to need expert assistance to understand and evaluate evidence (Wheate, 2006). Forensic science has played an important part in driving exonerations. But erroneous or exaggerated expert forensic testimony has also been a significant contributor to wrongful convictions (Garrett & Neufeld, 2009; Gould, Carrano, Leo, & Young, 2013; Saks & Koehler, 2005).

Researchers have consistently expressed concern that jurors are not aware of the limitations of forensic identification evidence (McAuliff & Duckworth, 2010; McAuliff & Kovera, 2008; McAuliff et al., 2009; McQuiston-Surrett & Saks, 2009). However, comparatively less research has looked at the best ways to educate jurors. Building upon previous work (Austin & Kovera, 2015; Eastwood & Caldwell, 2015; Jones & Kovera, 2015; Levett & Kovera, 2008, 2009), the research reported in this thesis has shown that opposing expert testimony may be able to assist jurors in making more accurate decisions, but that more research is needed into cross-examination and why jurors respond differently to the two forms of presenting the same arguments. While the research has not directly compared opposing experts and cross-examination, study three has provided a foundation for this work to continue.

This thesis has highlighted the importance of juror perception of the reliability of forensic science. Researching what jurors already believe about forensic science when walking into a courtroom will be the next step in refining methods of education. In criminal trials where the strongest evidence linking a defendant to a crime is physical evidence, the perception of the reliability of that evidence is the biggest predictor of verdict choice. If jurors have exaggerated beliefs about the evidence they are more vulnerable to making inaccurate decisions.

Gross (2013) has suggested that while it is ultimately an unknowable, and untestable, figure, between 1 and 5% of trials for serious violent offences may have ended in a wrongful conviction. The National Registry of Exonerations in the US lists 2,127 people exonerated, with more than 18, 450 years lost behind bars (2017). The Criminal Cases Review Commission, which investigates miscarriages of justice in England and Wales, has had 627 cases heard by the courts, with 419 appeals allowed since 1997. Dioso-Villa (2015) estimates that between 1922 and 2015, there have been at least 71 wrongful convictions in Australia. Research into ways to help jurors evaluate forensic evidence must continue. It is crucial that the legal community knows that jurors are making accurate decisions: faith in convictions requires faith in juror decision making.

"It is a fact, that in all matters which require to be investigated through the evidence of expert witnesses, *the same remarkable discrepancies show themselves… Even in criminal cases, where the point to be decided is whether a particular poison was administrated, or whether a death was caused in a particular way, the evidence of the experts is generally more contradictory than would be supposed from the nature of the inquiry; and, in short, judges and lawyers are rapidly coming to the conclusion that skilled testimony*, *which ought to be the most decisive and convincing of them all, is of all the most suspicious and unsatisfactory"*

- Editorial, "Expert Witnesses", *Saturday Review*, 11th January 1862

References

Ackland, R. (2011). A return to village life: internet bursts the bubble protecting juries. *Sydney Morning Herald*. Retrieved from www.smh.com.au

Aksu, M. N., & Gobetti, J. P. (1996). The past and present legal weight of bite marks as evidence. *American Journal of Forensic Medicine and Pathology, 17*(2), 136-140. doi: Doi 10.1097/00000433-199606000-00010

Antonio, L., & Lipsitz, A. (Writers). (2000). Friends & Lovers, *CSI: Crime Scene Investigation*. New York, NY: CBS.

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, 16-24.

Ashbaugh, D. R. (1999). *Quantitative--qualitative friction ridge analysis: An introduction to basic and advanced ridgeology*. Boca Raton, Fla: CRC Press.

Ask, K., Granhag, P. A., & Rebelius, A. (2011). Investigators under influence: How social norms activate goal-directed processing of criminal evidence. *Applied Cognitive Psychology, 25*(4), 548-553. doi: 10.1002/acp.1724

Ask, K., Reinhard, M.-A., Marksteiner, T., & Granhag, P. A. (2010). Elasticity in evaluations of criminal evidence: Exploring the role of cognitive dissonance. *Legal and Criminological Psychology, 16*(2), 289-306. doi: 10.1348/135532510X510153

Association of Forensic Science, P. (2009). Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice, 49*(3), 161-164. doi: 10.1016/j.scijus.2009.07.004

Austin, J. L., & Kovera, M. B. (2015). Cross- Examination Educates Jurors About Missing Control Groups in Scientific Evidence. *Psychology, Public Policy, and Law, 21*(3), 252-264. doi: 10.1037/law0000049

Avon, S. L. (2004). Forensic odontology: the roles and responsibilities of the dentist. *Journal (Canadian Dental Association), 70*(7), 453.

Bacon, F. (1900). *Novum organum*. London : New York: Routledge, Dutton.

Balcetis, E., & Dunning, D. (2006). See What You Want to See: Motivational Influences on Visual Perception. *Journal of Personality and Social Psychology, 91*(4), 612-625. doi: 10.1037/0022-3514.91.4.612

Barkan, S. E., & Bryjak, G. J. (2011). *Fundamentals of Criminal Justice: A Sociological View*. Sudbury, MA: Jones & Bartlett Learning.

Baskin, D. R., & Sommers, I. B. (2010). Crime-Show-Viewing Habits and Public Attitudes Toward Forensic Evidence: The "CSI Effect" Revisited. *The Justice System Journal, 31*(1), 97-113. doi: 10.1080/0098261X.2010.10767956

Bentler, P. M., & Wu, E. J. C. (2002). *EQS for Windows user's guide*. Encino, CA: Multivariate Software, Inc. .

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20*(1), 115-129. doi: 10.1002/acp.1171

Berger, C. A., Champod, J., Curran, C., Dawid, J., & Kloosterman, A. (2011). Expressing evaluative opinions: a position statement. *Sci. Justice, 51*(1), 1-2.

Bernstein, D. E. (1995). The science of forensic psychiatry and psychology. *Psychiatry, Psychology and Law, 2*(1), 75-80. doi: 10.1080/13218719509524852

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting, 1*(3), 257-269. doi: 10.1002/for.3980010305

Biber, K. (2007). *Captive images: Race, crime, photography*: Routledge.

Boehm, V. R. (1968). Mr. Prejudice, Miss Sympathy, and the authoritarian personality: An application of psychological measuring techniques to the problem of jury bias. *Wis. L. Rev.*, 734.

Bogan, P. S., & Roberts, A. (2011). Voice *Identification : investigation, trial and scientific evidence*. Briston: Jordans.

Bond, J. W. (2007). Value of DNA Evidence in Detecting Crime. *Journal of Forensic Sciences, 52*(1), 128-136. doi: 10.1111/j.1556-4029.2006.00323.x

Bornstein, B. H. (1999). The Ecological Validity of Jury Simulations: Is the Jury Still Out? *Law and Human Behavior, 23*(1), 75-91. doi: 10.1023/A:1022326807441

Bornstein, B. H. (2004). The impact of different types of expert scientific testimony on mock jurors' liability verdicts. *Psychology, Crime & Law, 10*(4), 429-446. doi: 10.1080/10683160310001629292

Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and human behavior., 41*(1), 13. doi: 10.1037/lhb0000223

Bourgeois, M. J., Horowitz, I. A., ForsterLee, L., & Grahe, J. (1995). Nominal and interactive groups: Effects of preinstruction and deliberations on decisions and evidence recall in complex trials. *Journal of Applied Psychology, 80*(1), 58.

Bowers, C. M. (2010). *Forensic Dental Evidence An Investigator's Handbook* (2nd ed. ed.). Burlington: Elsevier Science.

Bowers, C. M. (2014). Cross-Examination: The Expert's Challenge and the Lawyers' Strategies (pp. 125-145).

Brekke, N. J., Enko, P. J., Clavet, G., & Seelau, E. (1991). Of juries and court-appointed experts: The impact of nonadversarial versus adversarial expert testimony. *Law and Human Behavior, 15*(5), 451.

Brewer, N., Harvey, S., & Semmler, C. (2004). Improving comprehension of jury instructions with audio-visual presentation. *Applied Cognitive Psychology, 18*(6), 765-776. doi: 10.1002/acp.1036

Brewer, P. R., & Ley, B. L. (2010). Media Use and Public Perceptions of DNA Evidence. *Science Communication, 32*(1), 93-117. doi: 10.1177/1075547009340343

Bright, D. A., & Goodman-Delahunty, J. (2006). Gruesome evidence and emotion: Anger, blame, and jury decision-making. *Law and Human Behavior, 30*(2), 183.

Brodsky, S. L., Griffin, M. P., & Cramer, R. J. (2010). The Witness Credibility Scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law, 28*(6), 892-907. doi: 10.1002/bsl.917

Bromby, M. C. (2003). At face value? *New Law Journal, Expert Witness Supplement vol 153*, 302-304.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes, 41*(3), 390-404.

Buck, J. A., London, K., & Wright, D. B. (2011). Expert Testimony Regarding Child Witnesses: Does It Sensitize Jurors to Forensic Interview Quality? *Law and Human Behavior, 35*(2), 152-164. doi: 10.1007/s10979-010-9228-2

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation, 32*, 275-318.

Budowle, B., Bottrell, M. C., Bunch, S. G., Fram, R., Harrison, D., Meagher, S., . . . Stacey, R. B. (2009). A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement*. *Journal of Forensic Sciences, 54*(4), 798-809. doi: 10.1111/j.1556-4029.2009.01081.x

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90*(2), 105.

Byrne, B. M. (2010). *Structural equation modeling with AMOS : basic concepts, applications, and programming* (2nd ed. ed.). New York: Psychology Press.

Cacioppo, J., & Petty, R. (1983). The Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research, 11*, 673.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131. doi: 10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition. *Psychological Bulletin, 119*(2), 197-253. doi: 10.1037/0033-2909.119.2.197

Caddy, B., & Cobb, P. (2004). Forensic Science. In P. C. White (Ed.), *Crime Scene to Court: The Essentials of Forensic Science. 2nd Edition* (2nd ed. ed., pp. 1-20). Cambridge: Royal Society of Chemistry.

Camerer, C. F., & Johnson, E. J. (Eds.). (1991) Toward a general theory of expertise: Prospects and limits. New York, NY, US: Cambridge University Press.

Camerer, C. F., & Kunreuther, H. (1989). Decision processes for low probability events: Policy implications (Vol. 8, pp. 565-592). New York.

Cao, X. (2015). The Influence of Fiction Versus Nonfiction on Political Attitudes. *Communication Research Reports, 32*(1), 83-92. doi: 10.1080/08824096.2014.989979

Cashman, K., & Henning, T. (2012). Lawyers and DNA: Issues in understanding and challenging the evidence. *Current Issues in Criminal Justice, 24*(1), 69-83.

Cashmore, J., & Trimboli, L. (2006). Child sexual assault trials: A survey of juror perceptions. *BOCSAR NSW Crime and Justice Bulletins*, 20.

Cavender, G., & Deutsch, S. K. (2007). CSI and moral authority: The police and science. *Crime, Media, Culture, 3*(1), 67-81. doi: 10.1177/1741659007074449

Cecil, J. S., Hans, V. P., & Wiggins, E. C. (1990). Citizen comprehension of difficult issues: Lessons from civil jury trials. *Am. UL Rev., 40*, 727.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752-766. doi: 10.1037/0022-3514.39.5.752

Chaiken, S., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252).

Chaiken, S., & Maheswaran, D. (1994a). Heuristic Processing Can Bias Systematic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment. *Journal of Personality and Social Psychology, 66*(3), 460-473. doi: 10.1037/0022-3514.66.3.460

Chaiken, S., & Maheswaran, D. (1994b). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460-460.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*: Guilford Press.

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the Diagnostic Utility of Facial Composites: Beliefs of Guilt Can Bias Perceived Similarity between

Composite and Suspect. *Journal of Experimental Psychology: Applied, 15*(1), 76-90. doi: 10.1037/a0014682

Chartrand, T. L., Dalton, A. N., & Cheng, C. M. (2008). The antecedents and consequences of nonconscious goal pursuit. *Handbook of motivation science*, 342-355.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. *Dual-process theories in social psychology, 15*, 73-96.

Chesterman, M. (1999). Criminal Trial Juries in Australia: From Penal Colonies to a Federal Democracy. *Law and Contemporary Problems, 62*(2), 69-102. doi: 10.2307/1192253

Cole, S. (2017). Changed Science Statutes: Can Courts Accommodate Accelerating Forensic Scientific and Technological Change?

Cole, S., A. (2002). *Suspect identities: a history of fingerprinting and criminal identification*. Cambridge, Mass.: Harvard University Press.

Cole, S., A. (2005). More than zero: accounting for error in latent fingerprint identification. *Journal of Criminal Law and Criminology, 95*(3), 985.

Cole, S., A., & Dioso-Villa, R. (2011). Should judges worry about the "CSI Effect"? *Court Review, 47*, 20-102.

Cole, S. A. (2008). Comment on 'Scientific validation of fingerprint evidence under Daubert'. *Law, Probability & Risk, 7*(2), 119-126. doi: 10.1093/lpr/mgm021

Cole, S. A. (2015). A surfeit of science: The "CSI effect" and the media appropriation of the public understanding of science. *Public understanding of science (Bristol, England), 24*(2), 130. doi: 10.1177/0963662513481294

Cole, S. A., & Dioso-Villa, R. (2006). CSI and its effects: Media, juries, and the burden of proof. *New Eng. L. Rev., 41*, 435.

Cole, S. A., & Dioso-Villa, R. (2009). Investigating the CSI Effect; Effect: Media and Litigation Crisis in Criminal Law. *Stanford Law Review, 61*, 1335-1535.

Committee on Evaluation of Sound Spectograms, A. o. B. a. S., & Sciences, N. R. C. (1979). *On the Theory and Practice of Voice Identificatio*.

Cooley, C., & Turvey, B. (2007). Observer effects and examiner bias: Psychological influences on the forensic examiner. *Crime reconstruction*, 51-83.

Cooley, C. M. (2004). Reforming the forensic science community to avert the ultimate injustice.(Capital Concerns: The Death Penalty in America). *Stanford Law & Policy Review, 15*(2), 381-446.

Cooper, J., Bennett, E. A., & Sukel, H. L. (1996). Complex Scientific Testimony: How Do Jurors Make Decisions? *Law and Human Behavior, 20*(4), 379-394. doi: 10.1007/BF01498976

Cooper, J., & Neuhaus, I. M. (2000a). The" Hired Gun" effect: Assessing the effect of pay, frequency of testifying, and credentials on the perception of expert testimony. *Law and Human Behavior, 24*(2), 149.

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

Cordner, S. (2015). Expert opinions and evidence: A perspective from forensic pathology. *Flinders Law Journal, 17*(2), 263-286.

Cosmides, L., & Tooby, J. (1994). Better than Rational: Evolutionary Psychology and the Invisible Hand. *The American Economic Review, 84*(2), 327-332.

Costabile, K. A., & Klein, S. B. (2005). Finishing strong: Recency effects in juror judgments. *Basic and Applied Social Psychology, 27*(1), 47-58.

Cramer, R. J., Harris, P. B., Fletcher, L. M., DeCoster, J., & Brodsky, S. L. (2011). A confidence-credibility model of expert witness persuasion: Mediating effects and implications for trial consultation. *Consulting Psychology Journal, 63*(2), 129-137. doi: 10.1037/a0024591

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences & the Law, 7*(2), 215-225. doi: 10.1002/bsl.2370070206

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1990). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology, 20*(13 14), 1197.

Cutler, B. L., Moran, G., & Narby, D. J. (1992). Jury selection in insanity defense cases. *Journal of Research in Personality, 26*(2), 165-182.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*(2), 185-191. doi: 10.1007/BF01062972

Dahir, V. B., Richardson, J. T., Ginsburg, G. P., Gatowski, S. I., Dobbin, S. A., & Merlino, M. L. (2005). JUDICIAL APPLICATION OF DAUBERT TO PSYCHOLOGICAL SYNDROME AND PROFILE EVIDENCE: A Research Note. *Psychology, Public Policy, and Law, 11*(1), 62-82. doi: 10.1037/1076-8971.11.1.62

Dalbert, C., Montada, L., & Schmitt, M. (1987). Glaube an eine gerechte Welt als Motiv: Validierungskorrelate zweier Skalen. *Psychologische Beitrage*.

Dann, B. M., Hans, V. P., & Kaye, D. H. (2007). Can jury trial innovations improve juror understanding of DNA evidence. *Judicature, 90*(4), 152-156.

Dartnall, S., & Goodman-Delahunty, J. (2006). Enhancing juror understanding of probabilistic DNA evidence.(Australia). *Australian Journal of Forensic Sciences, 38*(2), 85-96. doi: 10.1080/00450610609410635

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993), .

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23*(4), 482-505. doi: 10.1002/acp.1490

Davis, J. P., Valentine, T., & Davis, R. E. (2010). Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Science International, 200*(1), 165-176. doi: 10.1016/j.forsciint.2010.04.012

De Keijser, J., & Elffers, H. (2012). Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. *Psychology, Crime & Law, 18*(2), 191-207. doi: 10.1080/10683161003736744

De La Fuente, L., De La Fuente, E. I., & García, J. (2003). Effects of pretrial juror bias, strength of evidence and deliberation process on juror decisions: new validity evidence of the Juror Bias Scale scores. *Psychology, Crime & Law, 9*(2), 197-209. doi: 10.1080/1068316031000116283

Desmarais, S., & Read, J. D. (2011a). After 30 Years, What Do We Know about What Jurors Know? A Meta-Analytic Review of Lay Knowledge Regarding Eyewitness

Factors. *Law & Human Behavior (Springer Science & Business Media B.V.), 35*(3), 200-210. doi: 10.1007/s10979-010-9232-6

Desmarais, S. L., & Read, J. D. (2011b). After 30 Years, What Do We Know about What Jurors Know? A Meta-Analytic Review of Lay Knowledge Regarding Eyewitness Factors. *Law and Human Behavior, 35*(3), 200-210. doi: 10.1007/s10979-010-9232-6

Devenport, J., & Cutler, B. L. (2004). Impact of Defense-Only and Opposing Eyewitness Experts on Juror Judgments. *Law and Human Behavior, 28*(5), 569-576. doi: 10.1023/B:LAHU.0000046434.39181.07

Devenport, J. L., Stinson, V., Cutler, B. L., & Kravitz, D. A. (2002). How Effective Are the Cross-Examination and Expert Testimony Safeguards? Jurors' Perceptions of the Suggestiveness and Fairness of Biased Lineup Procedures. *Journal of Applied Psychology, 87*(6), 1042-1054. doi: 10.1037//0021-9010.87.6.1042

Devine, D. J. (2012). *Jury Decision Making: The State of the Science*. New York: New York University Press.

Devine, D. J., & Caughlin, D. E. (2014). Do They Matter? A Meta-Analytic Investigation of Individual Characteristics and Guilt Judgments. *Psychology, Public Policy, and Law, 20*(2), 109-134. doi: 10.1037/law0000006

Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law, 7*(3), 622-727. doi: 10.1037/1076-8971.7.3.622

Devlin, P. (1979). *The judge*: Oxford University Press Oxford.

Devlin, P. D. B. (1956). *Trial by jury*. London: Stevens.

DeWitt, J. S., Richardson, J. T., & Warner, L. G. (1997). Novel Scientific Evidence and Controversial Cases: A Social Psychological Examination. *Law & Psychology Review, 21*, 1-271.

Diamond, S. (1997). Illuminations and Shadows from Jury Simulations. *Law and Human Behavior, 21*(5), 561-571. doi: 10.1023/A:1024831908377

Diamond, S. S., Casper, J. D., Heiert, C. L., & Marshall, A.-M. (1996). Juror Reactions to Attorneys at Trial. *The Journal of Criminal Law and Criminology (1973-), 87*(1), 17-47. doi: 10.2307/1143971

DiFonzo, J. H. (2005). The crimes of crime labs. *Hofstra Law Review, 34*(1), 1-11.

DiFonzo, J. H., & Stern, R. C. (2007). Devil in a white coat: the temptation of forensic evidence in the age of CSI. *New England Law Review, 41*(3), 503-532.

Dioso-Villa, R. (2015). A repository of wrongful convictions in Australia: First steps toward estimating prevalence and causal contributing factors. *Flinders Law Journal, 17*(2), 163-202.

Dobbin, S. A., Gatowski, S. I., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2002). Applying Daubert: How well do judges understand science and scientific method. *Judicature, 85*(5), 244-247.

Douglas, K. S., Lyon, D. R., & Ogloff, J. R. (1997). The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: probative or prejudicial? *Law and Human Behavior, 21*(5), 485.

Dror, I. (2012). Combating Bias: The Next Step in Fighting Cognitive and Psychological Contamination. *Journal of Forensic Sciences, 57*(1), 276.

Dror, I., E., & Cole, S., A. (2010). The vision in "blind" justice: expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review, 17*(2), 161.

Dror, I., Thompson, W., Meissner, C., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). *Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making* (Vol. 60).

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice, 51*(4), 204-208. doi: 10.1016/j.scijus.2011.08.004

Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability & Risk, 9*(1), 47-67.

Dror, I. E., Péron, A. E., Hind, S. l., & Charlton, D. (2005). When emotions get the better of us: the effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology, 19*(6), 799-809. doi: 10.1002/acp.1130

Dunford, J. (2004). Looking forward: The direction of criminal law. *Bar News: The Journal of the NSW Bar Association*(Summer 2004-2005), 46.

Dwyer, J., Scheck, B., & Neufeld, P. (2003). *Actual innocence : when justice goes wrong and how to make it right* ([New ed.] ed.). New York: New American Library.

Eastwood, J., & Caldwell, J. (2015). Educating jurors about forensic evidence: using an expert witness and judicial instructions to mitigate the impact of invalid forensic science testimony. *Journal of Forensic Sciences, 60*(6), 1523-1528.

Edmond, E., & Roberts, A. (2011). Procedural Fairness, the Criminal Trial and Forensic Science and Medicine. *Sydney Law Review, 33*, 359-849.

Edmond, G. (2010). Impartiality, efficiency or reliability? A critical response to expert evidence law and procedure in Australia. *Australian Journal of Forensic Sciences, 42*(2), 83-99. doi: 10.1080/00450610903258128

Edmond, G. (2011a). The building blocks of forensic science and law: Recent work on DNA profiling (and photo comparison) (Book Review) (Vol. 41, pp. 127-152).

Edmond, G. (2011b). Review essay: The building blocks of forensic science and law: Recent work on DNA profiling (and photo comparison). *Social Studies of Science, 41*(1), 127-152.

Edmond, G. (2012). Is reliability sufficient? The Law Commission and expert evidence in international and interdisciplinary perspective (Part 1). *The International Journal of Evidence & Proof, 16*(1), 30-65. doi: 10.1350/ijep.2012.16.1.391

Edmond, G. (2013a). (Ad)ministering justice: Expert evidence and the professional responsibilities of prosecutors. *University of New South Wales Law Journal, The, 36*(3), 921-953.

Edmond, G. (2013b). Just truth? Carefully applying history, philosophy and sociology of science to the forensic use of CCTV images. *Studies in History and Philosophy of Biological & Biomedical Sciences, 44*(1), 80. doi: 10.1016/j.shpsc.2012.09.004

Edmond, G. (2014). The admissibility of forensic science and medicine evidence under the Uniform Evidence Law.(Australia, New South Wales). *Criminal Law Journal, 38*(3), 136-158.

Edmond, G. (2015a). Forensic science evidence and the conditions for rational (jury) evaluation. *Melb. UL Rev., 39*, 77.

Edmond, G. (2015b). What Lawyers Should Know about the Forensic Sciences. *Adel. L. Rev., 36*, 33.

Edmond, G., Biber, K., Kemp, R. I., & Porter, G. (2009). Law's looking glass: expert identification evidence derived from photographic and video images.(Australia). *Current Issues in Criminal Justice, 20*(3), 337-377.

Edmond, G., Cole, S., A., Cunliffe, E., & Roberts, A. J. (2014). Admissibility Compared: The Reception of Incriminating Expert Evidence (i.e., Forensic Science) in Four Adversarial Jurisdictions. *University of Denver Criminal Law Review, 3*.

Edmond, G., Hamer, D., & Cunliffe, E. (2016). A little ignorance is a dangerous thing: engaging with exogenous knowledge not adduced by the parties. *Griffith Law Review, 25*(3), 383-413. doi: 10.1080/10383441.2016.1238029

Edmond, G., Martire, K., Kemp, R., Hamer, D., Hibbert, B., Ligertwood, A., . . . White, D. (2014). How to cross-examine forensic scientists: a guide for lawyers. *Australian Bar Review, 39*(2), 174-197.

Edmond, G., Martire, K., & San Roque, M. (2017). Expert reports and the forensic sciences. *UNSWLJ, 40*, 590.

Edmond, G., & Mercer, D. (1999). Juggling science: From polemic to pastiche. *Social Epistemology, 13*(2), 215-233. doi: 10.1080/026917299298727

Edmond, G., & San Roque, M. (2014). Honeysett v. The Queen: forensic science, 'specialised knowledge' and the uniform evidence law. (Australia). *Sydney Law Review, 36*(2), 323.

Edwards, H. T. (2009). Solving the problems that plague the forensic science community. *Jurimetrics, 50*, 5.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: people's assessments of bias in themselves and others. *Personality & Social Psychology Bulletin, 31*(5), 680. doi: 10.1177/0146167204271570

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology, 32*(1), 53-88.

Ellsworth, P. (1993). Some Steps Between Attitudes and Verdicts. In R. Hastie (Ed.), *Inside the juror : the psychology of juror decision making*. Cambridge [England] ; New York, NY, USA: Cambridge England ; New York, NY, USA : Cambridge University Press.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive–experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 462-483). New York: Guilford.

Evidence Act 1995 (Cth) (Austl.).

Evidence Act (2006). http://www.legislation.govt.nz/act/public/2006/0069/latest/DLM393600.html.

Faigman, D. L., & Baglioni, A. J. (1988). Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence. *Law and Human Behavior, 12*(1), 1-17. doi: 10.1007/BF01064271

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data (How Many Falsify Research?). *PLoS ONE, 4*(5), e5738. doi: 10.1371/journal.pone.0005738

Feigenson, N. (2010). Visual evidence. *Psychonomic Bulletin & Review, 17*(2), 149-154. doi: 10.3758/pbr.17.2.149

Feigenson, N., & Spiesel, C. (2009). *Law on display: the digital transformation of legal persuasion and judgment*. New York: New York : New York University Press.

Findlay, M. (2009). *Australian criminal justice* (4th ed. ed.). South Melbourne, Vic.: Oxford University Press.

Findley, K. A. (2008). Innocents at risk: adversary imbalance, forensic science, and the search for truth.(Guilt vs. Guiltiness: Are the Right Rules for Trying Factual Innocence Inevitably the Wrong Rules for Trying Culpability?). *Seton Hall Law Review, 38*(3), 893-973.

Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review, 2006*(2), 291-397.

Fishfader, V. L., Howells, G. N., Katz, R. C., & Teresi, P. S. (1996). Evidential and extralegal factors in juror decisions: Presentation mode, retention, and level of emotionality. *Law and Human Behavior, 20*(5), 565.

Fong, G. T., & et al. (1986). The Effects of Statistical Training on Thinking about Everyday Problems. *Cognitive Psychology, 18*(3), 253-292. doi: 10.1016/0010-0285(86)90001-0

Forsterlee, L., Kent, L., & Horowitz, I. A. (2005). The cognitive effects of jury aids on decision-making in complex civil litigation. *Applied Cognitive Psychology, 19*(7), 867-884. doi: 10.1002/acp.1124

Found, B., & Ganas, J. (2012). The management of domain irrelevant context information in forensic handwriting examination casework. *Science & Justice*. doi: 10.1016/j.scijus.2012.10.004

Freckelton, I. R. (1999). *Australian judicial perspectives on expert evidence : an empirical study / Ian Freckelton, Prasuna Reddy, Hugh Selby*. Carlton, Vic.: Carlton, Vic. : Australian Institute of Judicial Administration Inc.

Freckelton, I. R., Reddy, P., & Selby, H. (2001). *Australian Magistrates' Perspectives on Expert Evidence: A Comparative Study*: Australian Institute of Judicial Administration Incorporated.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923), .

Garrett, B. L., & Mitchell, G. (2016). Forensics and Fallibility: Comparing the Views of Lawyers and Judges. *West Virginia Law Review, 119*, 2016-2063.

Garrett, B. L., & Neufeld, P. J. (2009). Invalid forensic science testimony and wrongful convictions. *Virginia Law Review, 95*, 1-2129.

Garvey, S. P., Hannaford-Agor, P., Hans, V. P., Mott, N. L., Munsterman, G. T., & Wells, M. T. (2004). Juror first votes in criminal trials. *Journal of Empirical Legal Studies, 1*(2), 371-398.

Gatowski, S., Dobbin, S., Richardson, J., Ginsburg, G., Merlino, M., & Dahir, V. (2001). Asking the Gatekeepers: A National Survey of Judges on Judging Expert

Evidence in a Post-Daubert World. *Law and Human Behavior, 25*(5), 433-458. doi: 10.1023/A:1012899030937

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. *The Oxford handbook of social cognition*, 282-312.

Ghoshray, S. (2007). Untangling the CSI effect in criminal jurisprudence: circumstantial evidence, reasonable doubt, and jury manipulation.(Crime Scene Investigation)(The CSI Effect: The True Effect of Crime Scene Television on the Justice System). *New England Law Review, 41*(3), 533-562.

Giannelli, P. C. (2007). Wrongful convictions and forensic science: the need to regulate crime labs. *North Carolina Law Review, 86*(1), 163-235.

Giannelli, P. C. (2008). Forensic science: under the microscope. *Ohio NUL Rev., 34*, 315.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality : the adaptive toolbox* (1st MIT Press paperback ed. ed.). Cambridge, Mass.: MIT Press.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of personality and social psychology., 59*(4), 601. doi: 10.1037/0022-3514.59.4.601

Gilovich, T. (1991). *how we know what isn't so: the fallibility of human reason in everyday life*. New York, N. Y.: Free Press.

Golding, J. M., Stewart, T. L., Yozwiak, J. A., Djadali, Y., & Sanchez, R. P. (2000). The impact of DNA evidence in a child sexual assault trial. *Child maltreatment, 5*(4), 373-383.

Goodman-Delahunty, J., Brewer, N., Clough, J., Horan, J., Ogloff, J. R., Tait, D., & Pratley, J. (2008). *Practices, policies and procedures that influence juror satisfaction in Australia : report to the Criminology Research Council*. Canberra: Canberra : Australian Institute of Criminology.

Goodman-Delahunty, J., & Hewson, L. (2010). *Improving Jury Understanding and Use of Expert DNA Evidence*: Australian Institute of Criminology.

Goodman-Delahunty, J., & Tait, D. (2005). Lay Participation in Legal Decision Making in Australia and New Zealand: Jury Trials and Administrative Tribunals. In M. F.

Kaplan & A. M. Martin (Eds.), *Understanding World Juries Through Psychological Research*: Routledge.

Goodman-Delahunty, J., & Wakabayashi, K. (2012). Adversarial forensic science experts: an empirical study of jury deliberation.(Australia)(Forensic Science and Justice: From Crime Scene to Court and Beyond). *Current Issues in Criminal Justice, 24*(1), 85-103.

Gould, J. B., Carrano, J., Leo, R. A., & Young, J. K. (2013). Predicting erroneous convictions: A social science approach to miscarriages of justice.

Green, D. W., Green, D. W., & McCloy, R. Reaching a verdict. *Thinking & Reasoning, 9*(4), 307-333.

Green, M. C., & Brock, T. C. (2000). The Role of Transportation in the Persuasiveness of Public Narratives. *Journal of Personality and Social Psychology, 79*(5), 701-721. doi: 10.1037/0022-3514.79.5.701

Greenberg, J., & Wursten, A. (1988). The psychologist and the psychiatrist as expert witnesses: Perceived credibility and influence. *Professional Psychology: Research and Practice, 19*(4), 373.

Greene, E., Downey, C., & Goodman-Delahunty, J. (1999). Juror decisions about damages in employment discrimination cases. *Behavioral Sciences & the Law, 17*(1), 107-121. doi: 10.1002/(SICI)1099-0798(199901/03)17:1<107::AID-BSL330>3.0.CO2-X

Gross, S. R. (2013). How Many False Convictions are There? How Many Exonerations are There?

Haber, L., & Haber, R. N. (2008). Scientific validation of fingerprint evidence under Daubert. *Law, Probability and Risk, 7*(2), 87-109. doi: 10.1093/lpr/mgm020

Hallgrimsdottir, H. K., Phillips, R., & Benoit, C. (2006). Fallen women and rescued girls: Social stigma and media narratives of the sex industry in Victoria, BC, from 1980 to 2005. *Canadian Review of Sociology/Revue canadienne de sociologie, 43*(3), 265-280.

Halpern, D. F., Blackman, S., & Salzman, B. (1989). Using statistical risk information to assess oral contraceptive safety. *Applied Cognitive Psychology, 3*(3), 251-260. doi: 10.1002/acp.2350030305

Hamill, R., Wilson, T. D., & Nisbett, R. E. (1980). Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology, 39*(4), 578-589. doi: 10.1037/0022-3514.39.4.578

Hand, L. (1901). Historical and practical considerations regarding expert testimony. *Harvard Law Review*, 40-58.

Hans, V. P., Hannaford-Agor, P. L., Mott, N. L., & Munsterman, G. T. (2003). The hung jury: The American Jury's insights and contemporary understanding. *CRIMINAL LAW BULLETIN-BOSTON-, 39*(1), 33-50.

Hans, V. P., Kaye, D. H., Michael Dann, B., Farley, E. J., & Albertson, S. (2011). Science in the Jury Box: Jurors' Comprehension of Mitochondrial DNA Evidence. *Law and Human Behavior, 35*(1), 60-71. doi: 10.1007/s10979-010-9222-8

Hardy, J. H. (2007). Odontology. In T. Thompson & S. M. Black (Eds.), *Forensic human identification : an introduction* (pp. 177-198). Boca Raton: CRC Press.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury*: The Lawbook Exchange, Ltd.

Hayes, R., & Levett, L. (2013). Community Members' Perceptions of the CSI Effect. *The Journal of the Southern Criminal Justice Association, 38*(2), 216-235. doi: 10.1007/s12103-012-9166-2

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15*(4), 445-464.

Heuer, L., & Penrod, S. (1994). Trial Complexity: A Field Investigation of Its Meaning and Its Effects. *Law and Human Behavior, 18*(1), 29-51. doi: 10.1007/BF01499142

Hewson, L., & Goodman-Delahunty, J. (2008). Using multimedia to support jury understanding of DNA profiling evidence. *Australian Journal of Forensic Sciences, 40*(1), 55-64. doi: 10.1080/00450610802050782

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43.

Hofer, B. K. (2002). Personal epistemology as a psychological and educational construct: an introduction. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal* *epistemology : the psychology of beliefs about knowledge and knowing* (pp. 3-14). Mahwah: Lawrence Erlbaum Associates.

Hofer, B. K., & Pintrich, P. R. (1997). The Development of Epistemological Theories: Beliefs About Knowledge and Knowing and Their Relation to Learning. *Review of Educational Research, 67*(1), 88-140. doi: 10.3102/00346543067001088

Hoffner, C., & Buchanan, M. (2005). Young Adults' Wishful Identification With Television Characters: The Role of Perceived Similarity and Character Attributes. *Media Psychology, 7*(4), 325-351. doi: 10.1207/S1532785XMEP0704_2

Hope, L., Eales, N., & Mirashi, A. (2014). Assisting jurors: Promoting recall of trial information through the use of a trial- ordered notebook. *Legal and Criminological Psychology, 19*(2), 316-331. doi: 10.1111/lcrp.12003

Horowitz, I. A., Bordens, K. S., Victor, E., Bourgeois, M. J., & Forsterlee, L. (2001). The effects of complexity on jurors' verdicts and construction of evidence. *The Journal of applied psychology, 86*(4), 641.

Horowitz, I. A., & ForsterLee, L. (2001). The effects of note-taking and trial transcript access on mock jury decisions in a complex civil trial. *Law and Human Behavior, 25*(4), 373-391. doi: 10.1023/A:1010655602400

Horowitz, I. A., ForsterLee, L., & Brolly, I. (1996). Effects of Trial Complexity on Decision Making. *Journal of Applied Psychology, 81*(6), 757-768. doi: 10.1037/0021-9010.81.6.757

Houck, M. M. (2006). CSI: Reality. *Scientific American, 295*(1), 84. doi: 10.1038/scientificamerican0706-84

Houck, M. M. (2015). *Professional issues in forensic science*: Academic Press.

House of Commons Science and Technology Committee. (2013). *Forensic Science, Second Report of Session 2013 – 14 of House of*

Common. London: The Stationery Office Limited.

Hovland, C., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly, 15*(4), 635-650. doi: 10.1086/266350

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion; psychological studies of opinion change.

Howes, L. M. (2015). The communication of forensic science in the criminal justice system: A review of theory and proposed directions for research. *Science & Justice, 55*(2), 145-154.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Hughes, T., & Magers, M. (2007). The perceived impact of crime scene investigation shows on the administration of justice. *Journal of Criminal Justice and Popular Culture, 14*(3), 259-276.

Hurwitz, S. D., Miron, M. S., & Johnson, B. T. (1992). Source Credibility and the Language of Expert Testimony 1. *Journal of Applied Social Psychology, 22*(24), 1909-1939. doi: 10.1111/j.1559-1816.1992.tb01530.x

Inman, K., & Rudin, N. (2001). *Principles and practice of criminalistics : the profession of forensic science*. Boca Raton, Fla: CRC Press.

Innocence Project. (2016). Unvalidated or Improper Forensic Science. Retrieved 14/12/16, from http://www.innocenceproject.org/causes/unvalidated-or-improper-forensic-science/

Innocence Project. (2017a). Calvin Lee Scott. Retrieved 9/11/2017, 2017, from https://www.innocenceproject.org/cases/calvin-lee-scott/

Innocence Project. (2017b). Eyewitness Misidentification. Retrieved 09/10/2017, from https://www.innocenceproject.org/causes/eyewitness-misidentification/

Ivkovic, S. K., & Hans, V. P. (2003). Jurors' Evaluations of Expert Testimony: Judging the Messenger and the Message. *Law & Social Inquiry, 28*(2), 441-482.

Janusz, P. (2000). Expert Evidence: The Road from Daubert to Joiner and Kumho Tire. *Court Review, 37*, 32-54.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323.

Jones, A. M., & Kovera, M. B. (2015). A Demonstrative Helps Opposing Expert Testimony Sensitize Jurors to the Validity of Scientific Evidence. *Journal of Forensic Psychology Practice, 15*(5), 401-422. doi: 10.1080/15228932.2015.1090225

Kaasa, S. O., Peterson, T., Morris, E. K., & Thompson, W. C. (2007). Statistical Inference and Forensic Evidence: Evaluating a Bullet Lead Match. *Law and Human Behavior, 31*(5), 433-447. doi: 10.1007/s10979-006-9074-4

Kahan, D. M., Hoffman, D. A., & Braman, D. (2009). Whose eyes are you going to believe? Scott v. Harris and the perils of cognitive illiberalism.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty : heuristics and biases*. Cambridge: Cambridge University Press.

Kalven, H., & Zeisel, H. (1966). *The American jury*. Boston: Little, Brown.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition, 2*(1), 42-52. doi: 10.1016/j.jarmac.2013.01.001

Kassin, S. M., & Garfield, D. A. (1991). Blood and guts: General and trial-specific effects of videotaped crime scenes on mock jurors. *Journal of Applied Social Psychology, 21*(18), 1459-1472.

Kassin, S. M., & Wrightsman, L. S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality, 17*(4), 423-442.

Kassin, S. M., & Wrightsman, L. S. (1988). *The American jury on trial : psychological perspectives*. New York: Hemisphere Pub. Corp.

Kaye, D. H., & Koehler, J. J. (1991). Can Jurors Understand Probabilistic Evidence? *Journal of the Royal Statistical Society. Series A (Statistics in Society), 154*(1), 75-81. doi: 10.2307/2982696

Keller, C. (2011). Modern courts 'too complex' for juries. *The Advertiser*. Retrieved from http://www.adelaidenow.com.au

Kelly, J. F. (1998). *Tainting evidence : inside the scandals at the FBI crime lab / John F. Kelly, Phillip K. Wearne*. New York: New York : Free Press.

Kennedy, D. (2011). Forensic Dentistry and Microbial Analysis of Bite Marks. *Australian Police Journal*, 6-15.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Resarch, in press*.

Kim, Y. S., Barak, G., & Shelton, D. E. (2009). Examining the “CSI-effect” in the cases of circumstantial evidence and eyewitness testimony: Multivariate and path analyses. *Journal of Criminal Justice, 37*(5), 452-460. doi: 10.1016/j.jcrimjus.2009.07.005

Kirby, M. D. (1998). The jury of the future.(Delivering Justice in a Democracy, part 3)(Australia). *Australian Bar Review, 17*(2), 113-126.

Koehler, J. J. (1997). One in Millions, Billions, and Trillions: Lessons from People v. Collins (1968) for People v. Simpson (1995). *Journal of Legal Education, 47*(2), 214-223.

Koehler, J. J. (2017). Intuitive Error Rate Estimates for the Forensic Sciences. *Jurimetrics, 57*, 153-168.

Koehler, J. J. (2018). How Trial Judges Should Think About Forensic Science Evidence.

Koehler, J. J., Schweitzer, N., Saks, M. J., & McQuiston, D. E. (2016). Science, technology, or examiner experience: What influences jurors’ judgments about forensic science testimony? *Psychology, Public Policy, and Law, 22*(4), 401.

Koppl, R., & Cowan, E. J. (2010). A Battle of Forensic Experts is not a Race to the Bottom. *Review of Political Economy, 22*(2), 235-262. doi: 10.1080/09538251003665644

Kovera, M. B., Levy, R. J., Borgida, E., & Penrod, S. D. (1994). Expert testimony in child sexual abuse cases - Effects of expert evidence type and cross- examination. *Law and Human Behavior, 18*(6), 653-674. doi: 10.1007/BF01499330

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? *Journal of Applied Psychology, 85*(4), 574-586. doi: 10.1037/0021-9010.85.4.574

Kovera, M. B., McAuliff, B. D., & Hebert, K. S. (1999). Reasoning About Scientific Evidence: Effects of Juror Gender and Evidence Quality on Juror Decisions in a Hostile Work Environment Case. *Journal of Applied Psychology, 84*(3), 362-375. doi: 10.1037/0021-9010.84.3.362

Kovera, M. B., Russano, M. B., & McAuliff, B. D. (2002). ASSESSMENT OF THE COMMONSENSE PSYCHOLOGY UNDERLYING DAUBERT Legal Decision

Makers' Abilities to Evaluate Expert Evidence in Hostile Work Environment Cases. *Psychology, Public Policy, and Law, 8*(2), 180-200. doi: 10.1037/1076-8971.8.2.180

Kramer, G. P., Kerr, N. L., & Carroll, J. S. (1990). Pretrial publicity, judicial remedies, and jury bias. *Law and Human Behavior, 14*(5), 409.

Krauss, D. A., Lieberman, J. D., & Olson, J. (2004). The effects of rational and experiential information processing of expert testimony in death penalty cases. *Behavioral Sciences & the Law, 22*(6), 801-822. doi: 10.1002/bsl.621

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology Public Policy And Law, 7*(2), 267-310. doi: 10.1037//1076-8971.7.2.267

Kravitz, D. A., Cutler, B. L., & Brock, P. (1993). Reliability and validity of the original and revised Legal Attitudes Questionnaire. *Law and Human Behavior, 17*(6), 661.

Kruse, C. (2010). Producing Absolute Truth: CSI Science as Wishful Thinking. *American Anthropologist, 112*(1), 79-91. doi: 10.1111/j.1548-1433.2009.01198.x

Kuhn, D. (1991). *The skills of argument*. Cambridge ; New York: Cambridge ; New York : Cambridge University Press.

Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1. doi: 10.1111/1467-9280.00302

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309-328. doi: 10.1016/S0885-2014(00)00030-7

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: the psychology of beliefs about knowledge and knowing* (pp. 121-144). Mahwah, NJ: Lawrence Erlbaum Associates.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How Well Do Jurors Reason? Competence Dimensions of Individual Variation in a Juror Reasoning Task. *Psychological Science, 5*(5), 289-296. doi: 10.1111/j.1467-9280

Kumkale, G. T., Albarracín, D., & Seignourel, P. J. (2010). The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the

Design of Persuasive Communication Campaigns. *Journal of Applied Social Psychology, 40*(6), 1325-1356. doi: 10.1111/j.1559-1816.2010.00620.x

Leadbeater, B., & Kuhn, D. (1989). Interpreting discrepant narratives: Hermeneutics and adult cognition. In J. D. Sinnott (Ed.), *Everyday problem solving: Theory and applications* (pp. 175-190). New York: Praeger.

Lecci, L., & Myers, B. (2002). Examining the Construct Validity of the Original and Revised JBS: A Cross-Validation of Sample and Method. *Law and Human Behavior, 26*(4), 455-463. doi: 10.1023/A:1016335422706

Lecci, L., & Myers, B. (2008). Individual Differences in Attitudes Relevant to Juror Decision Making: Development and Validation of the Pretrial Juror Attitude Questionnaire (PJAQ)1. *Journal of Applied Social Psychology, 38*(8), 2010-2038. doi: 10.1111/j.1559-1816.2008.00378.x

Lecci, L. B., & Myers, B. (2009). Predicting guilt judgments and verdict change using a measure of pretrial bias in a videotaped mock trial with deliberating jurors. *Psychology, Crime & Law, 15*(7), 619-634. doi: 10.1080/10683160802477757

Lee, E. (2016). How Long Does it Take to Shoot 1 Trillion Photos? Retrieved from http://blog.infotrends.com

Lee, L., Frederick, S., & Ariely, D. (2006). Try It, You'll Like It. *Psychological Science, 17*(12), 1054-1058. doi: 10.1111/j.1467-9280.2006.01829.x

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43*(6), 431.

Leippe, M. R. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1*, 909-959. doi: 10.1037/1076-8971.1.4.909

Leippe, M. R., & Eisenstadt, D. (2014). Eyewitness confidence and the confidence-accuracy relationship in memory for people. *The handbook of eyewitness psychology, 2*, 377-425.

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Seib, H. M. (2004). Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts. *Journal of Applied Psychology, 89*(3), 524-541. doi: 10.1037/0021-9010.89.3.524

Lempert, R. (1993). Civil juries and complex cases: Taking stock after twelve years. In R. E. Litan (Ed.), *Verdict: Assessing the Civil Jury System* (pp. 181-234). Washington, D.C: Brookings Institution.

Lempert, R. O. (2001). Citizen Participation in Judicial Decision Making: Juries, Lay Judges and Japan. *Saint Louis-Warsaw Transatlantic Law Journal, 2001-2002*, 1-14.

Levett, L., & Kovera, M. B. (2008). The Effectiveness of Opposing Expert Witnesses for Educating Jurors about Unreliable Expert Evidence. *Law Hum Behav, 32*(4), 363-374. doi: 10.1007/s10979-007-9113-9

Levett, L., & Kovera, M. B. (2009). Psychological mediators of the effects of opposing expert testimony on juror decisions. *Psychology, Public Policy, and Law, 15*(2), 124-148. doi: 10.1037/a0016309

Lieberman, J., & Sales, B. (1999). The effectiveness of jury instructions. *Handbook of Jury Research, 18*.

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychology, Public Policy, and Law, 14*(1), 27-62. doi: 10.1037/1076-8971.14.1.27

Ligertwood, A. (2015). What Lawyers Should and Can Do Now That They Know about the Forensic Sciences: A Response to Edmond's What Lawyers Should Know about the Forensic Sciences. *Adel. L. Rev., 36*, 153.

Lincoln, R., & Lindner, D. (2004). Judging juries. *The National Legal Eagle, 10*(1), 4.

Lockhart, J. J., & Satya-Murti, S. (2017). Diagnosing crime and diagnosing disease: bias reduction strategies in the forensic and clinical sciences. *Journal of Forensic Sciences, 62*(6), 1534-1541.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods, 1*(2), 130.

Maeder, E. M., & Corbett, R. (2015). Beyond frequency: perceived realism and the CSI effect. *Canadian Journal of Criminology and Criminal Justice, 57*(1), 83.

Magnussen, S., Melinder, A., Stridbeck, U., & Raja, A. Q. (2010). Beliefs about factors affecting the reliability of eyewitness testimony: A comparison of judges, jurors and the general public. *Applied Cognitive Psychology, 24*(1), 122-133.

Makens, J. (1965). Effect of Brand Preference upon Consumers' Perceived Taste of Turkey Meat. *Journal of Applied Psychology, 49*, 261.

Mancini, D. E. (2011). The CSI effect reconsidered: is it moderated by need for cognition? *North American Journal of Psychology, 13*(1), 155.

Mardia, K. V., Coombes, A., Kirkbride, J., Linney, A., & Bowie, J. L. (1996). On statistical problems with face identification from photographs. *Journal of Applied Statistics, 23*(6), 655-676. doi: 10.1080/02664769624008

Maricopa County Attorney's Office. (2005). CSI: Maricopa County: The CSI Effect and It's Real-Life Impact on Justice.

Martin, D., L. (2002). Lessons about justice from the "laboratory" of wrongful convictions: Tunnel vision, the construction of guilt and informer evidence. *UMKC Law Review, 70*, 847-1001.

Martire, K. A., & Bali, A. (2016). Judicial attitudes toward expert evidence. In U. E. O. Services (Ed.).

Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International, 240*, 61-68. doi: 10.1016/j.forsciint.2014.04.005

Matthews, R., Hancock, L., & Briggs, D. (2004). *Jurors' perceptions, understanding, confidence and satisfaction in the jury system: A study in six courts*: Research Development and Statistics Directorate, Home Office.

McAuliff, B. D., & Duckworth, T. D. (2010). I Spy with My Little Eye: Jurors' Detection of Internal Validity Threats in Expert Evidence. *Law and Human Behavior, 34*(6), 489-500. doi: 10.1007/s10979-010-9219-3

McAuliff, B. D., & Kovera, M. B. (2008). Juror Need for Cognition and Sensitivity to Methodological Flaws in Expert Evidence 1. *Journal of Applied Social Psychology, 38*(2), 385-408. doi: 10.1111/j.1559-1816.2007.00310.x

McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can Jurors Recognize Missing Control Groups, Confounds, and Experimenter Bias in Psychological Science? *Law and Human Behavior, 33*(3), 247-257. doi: 10.1007/s10979-008-9133-0

McClellan, P. (2012). The future role of the judge: umpire, manager, mediator or service provider?(New South Wales). *Law Society Journal, 50*(3), 69.

McQuiston-Surrett, D., & Saks, M. (2009). The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear. *Law and Human Behavior, 33*(5), 436-453. doi: 10.1007/s10979-008-9169-1

McQuiston-Surrett, D., & Saks, M. J. (2008). Communicating opinion evidence in the forensic identification sciences: accuracy and impact. (Symposium). *Hastings Law Journal, 59*(5), 1159-1189.

Mellis, C., Peat, J. K., Williams, K., & Xuan, W. (2001). *Health Science Research : A Handbook of Quantitative Methods*. Crows Nest, N.S.W.: Allen & Unwin.

Michael, A. (2007). *Reading 'CSI' Television Under the Microscope*. London: London : I.B.Tauris.

Mill, D., Gray, T., & Mandel, D. R. (1994). Influence of Research Methods and Statistics Courses on Everyday Reasoning, Critical Abilities, and Belief in Unsubstantiated Phenomena. *Canadian Journal of Behavioural Science/Revue canadienne des Sciences du comportement, 26*(2), 246-258. doi: 10.1037/0008-400X.26.2.246

Miller, L. S. (1984). Bias among forensic document examiners: a need for procedural change. *Journal of Police Science and Administration, 12*(4), 407-411.

Mills, C. J., & Bohannon, W. E. (1980). Character structure and jury behavior: Conceptual and applied implications. *Journal of Personality and Social Psychology, 38*(4), 662.

Mitchell, H. E., & Byrne, D. (1973). The defendant's dilemma: Effects of jurors' attitudes and authoritarianism on judicial decisions. *Journal of Personality and Social Psychology, 25*(1), 123.

Mnookin, J. L. (2003). Fingerprints: Not a Gold Standard. *Issues in Science and Technology, 20*(1), 47-54.

Mnookin, J. L. (2008). The Validity of Latent Fingerprint Identification: Confessions of a Fingerprinting Moderate. *Law, Probability and Risk, 7*, 127-142.

Modin, K. (2006). *The impact of photographic crime scene evidence on juror verdicts, jurors' feelings of confidence, beliefs about impartiality, and emotional reactions.* Alliant International University, Center for Forensic Studies, Fresno.

Monahan, J., & Steadman, H. J. (1994). *Violence and mental disorder : developments in risk assessment*. Chicago: University of Chicago Press.

Monahan, J., Walker, L., & Mitchell, G. (2009). The limits of social framework evidence. *Law, Probability & Risk, 8*(4), 307-321.

Moran, G., & Comfort, J. C. (1982). Scientific juror selection: Sex as a moderator of demographic and personality predictors of impaneled felony juror behavior. *Journal of Personality and Social Psychology, 43*(5), 1052.

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., . . . Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27*(3), 265-279. doi: 10.1016/j.ijlp.2004.03.004

Moriarty, J. C., & Saks, M. J. (2005). Forensic science: grand goals, tragic flaws, and judicial gatekeeping. *Judges Journal, 44*(4), 16.

Morrison, G. S. (2009). Forensic voice comparison and the paradigm shift (Vol. 49, pp. 298-308).

Morrison, G. S. (2014). Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison. *Science & Justice, 54*(3), 245-256. doi: 10.1016/j.scijus.2013.07.004

Morrison, G. S., Sahito, F. H., Jardine, G., Djokic, D., Clavet, S., Berghs, S., & Goemans Dorny, C. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. *Forensic Science International, 263*, 92-100. doi: http://dx.doi.org/10.1016/j.forsciint.2016.03.044

Moskowitz, G. B., & Grant, H. (2009). *The Psychology of Goals*. New York: New York : Guilford Publications.

Myers, B., & Lecci, L. (1998). Revising the Factor Structure of the Juror Bias Scale: A Method for the Empirical Validation of Theoretical Constructs. *Law and Human Behavior, 22*(2), 239-256. doi: 10.1023/A:1025798204956

Nance, D., & Morris, S. (2005). Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small RandomMatch Probability. *The Journal of Legal Studies, 34*(2), 395-444. doi: 10.1086/428020

Narby, D. J., Cutler, B. L., & Moran, G. (1993). A meta-analysis of the association between authoritarianism and jurors' perceptions of defendant culpability: American Psychological Association.

National Academy of Sciences. (2009). *Strengthening forensic science in the United States: a path forward*. Washington, DC: National Academies Press.

National Science Board. (2000). *Science & Engineering Indicators – 2000* (Vol. 1). Arlington, VA: National Science Board.

Neal, T. M., & Brodsky, S. L. (2008). Expert witness credibility as a function of eye contact behavior and gender. *Criminal Justice and Behavior, 35*(12), 1515-1526.

Neal, T. M. S., Guadagno, R. E., Eno, C. A., & Brodsky, S. L. (2012). Warmth and competence on the witness stand: implications for the credibility of male and female expert witness. *Journal of the American Academy of Psychiatry and the Law, 40*(4), 488-497.

Neisser, U. (1976). *Cognition and reality : principles and implications of cognitive psychology*. San Francisco: W.H. Freeman.

Nemeth, R. J. (2002). *The impact of gruesome evidence on mock juror decision making: The role of evidence characteristics and emotional response.*

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175-220. doi: 10.1037/1089-2680.2.2.175

Niedermeier, K. E., Kerr, N. L., & Messé, L. A. (1999). Jurors' Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect. *Journal of Personality and Social Psychology, 76*(4), 533-542. doi: 10.1037/0022-3514.76.4.533

Nolan, F. (2001). *Speaker identification evidence: its forms, limitations, and roles.* Paper presented at the Proceedings of the Conference ‘‘Law and Language: Prospect and Retrospect’’, Levi (Finnish Lapland).

Nolan, M. (2003). The role and future of the jury in the Australian legal system : More creativity, less criticism : an 'evidence based' approach to jury reform. *Legaldate, 15*(3), 5-7.

Nolan, T. W. (2007). Depiction of the " CSI effect" in popular culture: portrait in domination and effective affectation.(The CSI Effect: The True Effect of Crime Scene Television on the Justice System). *New England Law Review, 41*(3), 575-589.

Norell, K., Läthén, K. B., Bergström, P., Rice, A., Natu, V., & O' Toole, A. (2015). The Effect of Image Quality and Forensic Expertise in Facial Image Comparisons. *Journal of Forensic Sciences, 60*(2), 331-340. doi: 10.1111/1556-4029.12660

Norris, C., McCahill, M., & Wood, D. (2002). The growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space. *Surveillance & Society, 2*(2/3).

O'Brien, K., Goodman-Delahunty, J., Clough, J., & Pratley, J. (2008). *Factors affecting juror satisfaction and confidence in New South Wales, Victoria and South Australia*. Canberra: Canberra: Australian Institute of Criminology.

O'Brien, B. (2009). PRIME SUSPECT: AN EXAMINATION OF FACTORS THAT AGGRAVATE AND COUNTERACT CONFIRMATION BIAS IN CRIMINAL INVESTIGATIONS. *Psychology, Public Policy, and Law, 15*(4), 315-334. doi: 10.1037/a0017881

Odgers, S. (2015). What lawyers should do about forensic science evidence.(response to article by Gary Edmond in this issue, p. 33). *Adelaide Law Review, 36*(1), 147-151.

Ogloff, J. R. P., Clough, J., Goodman-Delahunty, J., & Young, W. (2006). *The jury project : stage 1 : a survey of Australian and New Zealand judges*. Melbourne: Melbourne: Australian Institute of Judicial Administration.

Oliver, E., & Griffitt, W. (1976). Emotional arousal and 'objective'judgment. *Bulletin of the Psychonomic Society, 8*(5), 399-400.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied, 4*(2), 101.

Otto, R. K. (1992). Prediction of dangerous behavior: A review and analysis of "second-generation" research. *Forensic Reports, 5*(1), 103-133.

Page, M., Taylor, J., & Blenkin, M. (2011). Uniqueness in the forensic identification sciences—Fact or fiction? *Forensic Science International, 206*(1), 12-18. doi: 10.1016/j.forsciint.2010.08.004

Park, J., & Feigenson, N. (2013). Effects of a visual technology on mock juror decision making. *Applied Cognitive Psychology, 27*(2), 235-246.

Pennington, N., & Hastie, R. (1986). Evidence Evaluation in Complex Decision Making. *Journal of Personality and Social Psychology, 51*(2), 242-258.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189.

Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror : the psychology of juror decision making* (pp. 192-225). Cambridge: Cambridge University Press.

Penrod, S. D., & Cutler, B. (1989). EYEWITNESS EXPERT TESTIMONY AND JURY DECISION-MAKING. *Law Contemp. Probl., 52*(4), 43-83. doi: 10.2307/1191907

Peterson, J. L., Lin, G., Ho, M., Chen, Y., & Gaensslen, R. (2003). The feasibility of external blind DNA proficiency testing. II. experience with actual blind tests. *Journal of Forensic Science, 48*(1), 1-9.

Peterson, J. L., Mihajlovic, S., & Gilliland, M. (1984). *Forensic evidence and the police: The effects of scientific evidence on criminal investigations*: US Department of Justice, National Institute of Justice Washington, DC.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *Journal of Consumer Research, 10*(2), 135-146.

Pezdek, K., Avila-mora, E., & Sperry, K. (2010). Does Trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Applied Cognitive Psychology, 24*(5), 673-690. doi: 10.1002/acp.1578

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Lay people's and police officers' attitudes towards the usefulness of perpetrator voice identification. *Applied Cognitive Psychology, 21*(1), 103. doi: 10.1002/acp.1281

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.

Podlas, K. (2002). Should we blame Judge Judy? The messages TV courtrooms send viewers. *Judicature, 86*(1), 38-43.

Podlas, K. (2006a). The "CSI effect" and other forensic fictions. *Loyola of Los Angeles Entertainment Law Review, 27*(2), 87-125.

Podlas, K. (2006b). "The CSI Effect": Exposing the Media Myth. *Fordham Intellectual Property, Media & Entertainment Law Journal, 16*, 429-1349.

Porter, G. (2009). CCTV images as evidence. *Australian Journal of Forensic Sciences, 41*(1), 11-25. doi: 10.1080/00450610802537960

Porter, G. (2013). Images as evidence. *Precedent*(119), 38-42.

Porter, G., & Kennedy, M. (2012). Photographic truth and evidence. *Australian Journal of Forensic Sciences, 44*(2), 183-192. doi: 10.1080/00450618.2011.634835

Potter, W. J. (1986). Perceived Reality and the Cultivation Hypothesis. *Journal of Broadcasting & Electronic Media, 30*(2), 159-174.

Prentice, D., Gerrig, R. J., Chaiken, S., & Trope, Y. (1999). Exploring the boundary between fiction and reality. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 529–546). New York, NY: Guilford Press.

President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: ensuring scientific validity of feature-comparison methods*. Washington DC: Retrieved from https://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review, 111*(3), 781-799. doi: 10.1037/0033-295X.111.3.781

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: perceptions of bias in self versus others. *Personality & Social Psychology Bulletin, 28*(3), 369.

Pyrek, K. (2007). *Forensic Science Under Siege: The Challenges of Forensic Laboratories and the Medico-Legal Investigation System*: Elsevier Science.

Quas, J. A., Bottoms, B. L., Haegerich, T. M., & Nysse-Carris, K. L. (2002). Effects of Victim, Defendant, and Juror Gender on Decisions in Child Sexual Assault Cases1. *Journal of Applied Social Psychology, 32*(10), 1993-2021. doi: 10.1111/j.1559-1816.2002.tb02061.x

R v Alrekabi [2007] NSWDC 110, .

R v Bornyk 2013 BCSC 1927.

R v Keogh (No 2) [2014] SASCFC 136 (19 December 2014), .

R v Murdoch [2005] NTSC 78, .

R v Tang [2006] NSWCCA 167, .

R v Ward (1993) 96 Cr.App.R. 1 at 52), .

Ratneshwar, S., & Chaiken, S. (1991). Comprehension's Role in Persuasion: The Case of Its Moderating Effect on the Persuasive Impact of Source Cues. *Journal of Consumer Research, 18*(1), 52-62. doi: 10.1086/209240

Reference Under s 433a of the Criminal Code by the Attorney-General for the Northern Territory of Australia of Convictions of Alice Lynne Chamberlain and Michael Leigh Chamberlain [1988] NTSC 64 (15 September 1988), .

Risinger, D. M. (2000). Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?

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. *California Law Review, 90*, 1-56.

Robbers, M. L. (2008). Blinded by science: The social construction of reality in forensic television shows and its effect on criminal jury trials. *Criminal Justice Policy Review, 19*(1), 84-102.

Roberts, P. (2015). Paradigms of forensic science and legal process: a critical diagnosis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1674). doi: 10.1098/rstb.2014.0256

Robinson v. Jacksonville Shipyards, Inc, No. No. 86-927-Civ-J-12, 760 1486 (Dist. Court, MD Florida 1991).

Roese, N. J., & Sherman, J. W. (2007). Expectancy. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology : handbook of basic principles* (Vol. 1, pp. 91-115). New York: Guilford Press.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

Rozelle, S. D. (2007). Daubert, Schmaubert: Criminal defendants and the short end of the science stick. *Tulsa L. Rev., 43*, 597.

Ruva, C. L., & Guenther, C. C. (2015). From the Shadows Into the Light: How Pretrial Publicity and Deliberation Affect Mock Jurors' Decisions, Impressions, and Memory. *Law and Human Behavior, 39*(3), 294-310. doi: 10.1037/lhb0000117

Saks, M. J., & Faigman, D. L. (2008). Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It. *Annu. Rev. Law Soc. Sci., 4*, 149-171. doi: 10.1146/annurev.lawsocsci.4.110707.172303

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309*, 892-895. doi: 10.1126/science.1111565

Saks, M. J., & Koehler, J. J. (2008). The individualization fallacy in forensic science evidence. *Vanderbilt Law Review, 61*(1), 199-219.

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. doi: 10.1016/S1355-0306(03)71747-X

Saks, M. J., & Spellman, B. A. (2016). *The psychological foundations of evidence law*: NYU Press.

Salerno, J. M., & McCauley, M. R. (2009). Mock jurors' judgments about opposing scientific experts: Do cross-examination, deliberation, and need for cognition matter? *American Journal of Forensic Psychiatry, 27*, 37-60.

Schklar, J., & Diamond, S. S. (1999). Juror Reactions to DNA Evidence: Errors and Expectancies. *Law and Human Behavior, 23*(2), 159-184. doi: 10.1023/A:1022368801333

Schofield, J. R. (1997). A Misapplication of Daubert: Compton v. Subaru of America Opens the Gate for Unreliable and Irrelevant Expert Testimony. *BYU L. Rev.*, 489.

Schweitzer, N. J., & Saks, M. J. (2007). The CSI Effect: popular fiction about forensic science affects the public's expectations about real forensic science. *Jurimetrics Journal of Law, Science and Technology, 47*(3), 357-367.

Schweitzer, N. J., & Saks, M. J. (2009). The gatekeeper effect: The impact of judges' admissibility decisions on the persuasiveness of expert testimony. *Psychology, Public Policy, and Law, 15*(1), 1.

Scurich, N., Krauss, D. A., Reiser, L., Garcia, R. J., & Deer, L. (2015). Venire Jurors' Perceptions of Adversarial Allegiance. *Psychology, Public Policy, and Law, 21*(2), 161-168. doi: 10.1037/law0000042

Semmler, C., & Brewer, N. (2002). Using a Flow-chart to Improve Comprehension of Jury Instructions. *Psychiatry, Psychology and Law, 9*(2), 262-270. doi: 10.1375/pplt.2002.9.2.262

Sheahan, C. L., Pica, E., & Pozzulo, J. D. (2017). Abuse Is Abuse: The Influence of Type of Abuse, Victim Age, and Defendant Age on Juror Decision Making. *Journal of Interpersonal Violence*, 0886260517731316. doi: 10.1177/0886260517731316

Shelton, D. E., Kim, Y. S., & Barak, G. (2006). A study of juror expectations and demands concerning scientific evidence: does the "CSI effect" exist? *Vanderbilt Journal of Entertainment and Technology Law, 9*(2), 331-368.

Shelton, D. E., Kim, Y. S., & Barak, G. (2009). An Indirect-Effects Model of Mediated Adjudication: The CSI Myth, the Tech Effect, and Metropolitan Jurors' Expectations for Scientific Evidence. *Vanderbilt Journal of Entertainment & Technology Law, 12*, 1-947.

SHJB v Minister for Immigration and Multicultural and Indigenous Affairs [2003] FCA 502, .

Shuman, D. W., Whitaker, E., & Champagne, A. (1994). An empirical examination of the use of expert witnesses in the courts—Part II: A three city study. *Jurimetrics*, 193-208.

Simon, D., Snow, C. J., & Read, S. J. (2004). The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction. *Journal of Personality and Social Psychology, 86*(6), 814-837. doi: 10.1037/0022-3514.86.6.814

Skolnick, P., & Shaw, J. I. (2001). A comparison of eyewitness and physical evidence on mock-juror decision making. *Criminal Justice and Behavior, 28*(5), 614-630.

Sleath, E., & Woodhams, J. (2011). *Perceptions of rape victim blaming*. Manuscript submitted for publication.

Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. (2016). The Perfect Match: Do Criminal Stereotypes Bias Forensic Evidence Analysis? *Law and Human Behavior*. doi: 10.1037/lhb0000190

Smith, A. E., & Haney, C. (2011). Getting to the point: attempting to improve juror comprehension of capital penalty phase instructions. *Law and Human Behavior, 35*(5), 339.

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' Use of Probabilistic Evidence. *Law and Human Behavior, 20*(1), 49-82.

Smith, L., & Bull, R. (2012). Identifying and measuring juror pre-trial bias for forensic evidence: development and validation of the Forensic Evidence Evaluation Bias Scale. *Psychology, Crime & Law, 18*(9), 797-815. doi: 10.1080/1068316x.2011.561800

Smith, L., & Bull, R. (2014). Validation of the factor structure and predictive validity of the Forensic Evidence Evaluation Bias Scale for robbery and sexual assault trial scenarios. *Psychology, Crime and Law, 20*(5), 450-466. doi: 10.1080/1068316X.2013.793340

Smith, L., Bull, R., & Holliday, R. (2011). Understanding Juror Perceptions of Forensic Evidence: Investigating the Impact of Case Context on Perceptions of Forensic Evidence Strength. *Journal of Forensic Sciences*(3), 409-414.

Smith, L. L. (2011). *The role of pre-trial attitudes about forensic science evidence: Developing and testing a forensic evidence evaluation bias scale.* University of Leicester.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47*(2), 149-155. doi: 10.1037/h0047060

Smith, S. M., Patry, M. W., & Stinson, V. (2007). But what is the CSI Effect? How crime dramas influence people's beliefs about forensic evidence. *The Canadian Journal of Police & Security Services, 5*(3), 187-195.

Stangor, C., & McMillan, D. (1992). Memory for Expectancy-Congruent and Expectancy-Incongruent Information: A Review of the Social and Social Developmental Literatures. *Psychological Bulletin, 111*(1), 42-61. doi: 10.1037/0033-2909.111.1.42

Stevens, D. J. (2008). Forensic Science, Wrongful Convictions, and American Prosecutor Discretion. *Howard Journal of Criminal Justice, 47*(1), 31-51. doi: 10.1111/j.1468-2311.2008.00495.x

Stinson, V., Patry, M. W., & Smith, S. M. (2007). The CSI effect: Reflections from police and forensic investigators. *The Canadian Journal of Police and Security Services, 5*(3), 1-9.

Stoel, R. D., Dror, I. E., & Miller, L. S. (2014). Bias among forensic document examiners: Still a need for procedural changes. *Australian Journal of Forensic Sciences, 46*(1), 91-97. doi: 10.1080/00450618.2013.797026

Stone, M. (1995). *Cross-examination in criminal trials / Marcus Stone* (2nd ed. ed.). London: London : Butterworths.

Stoney, D. A. (1991). What made us ever think we could individualize using statistics? (Vol. 31, pp. 197-199).

Studebaker, C. A., & Penrod, S. D. (1997). Pretrial publicity: The media, the law, and common sense. *Psychology, Public Policy, and Law, 3*(2-3), 428.

Swann Jr, W. B., & Giuliano, T. (1987). Confirmatory search strategies in social interaction: How, when, why, and with what consequences. *Journal of Social and Clinical Psychology, 5*(4), 511-524.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22*(8), 995-997. doi: 10.1177/0956797611414729

The National Registry of Exonerations. (2017). Summary View. Retrieved 23/11/17, from https://www.law.umich.edu/special/exoneration/Pages/browse.aspx

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: experts can do a lot with a little. *PLoS ONE, 9*(12), e114759. doi: 10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2013). Expertise in Fingerprint Identification. *Journal of Forensic Sciences, 58*(6), 1519-1530. doi: 10.1111/1556-4029.12203

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human Matching Performance of Genuine Crime Scene Latent Fingerprints. *Law and Human Behavior, 38*(1), 84-93. doi: 10.1037/lhb0000051

Thompson, W. (1995). Subjective interpretation, laboratory error and the value of forensic DNA evidence: Three case studies. *Genetica, 96*(1), 153-168. doi: 10.1007/BF01441161

Thompson, W. C. (2008). Beyond bad apples: analyzing the role of forensic science in wrongful convictions. *Southwestern University Law Review, 37*(4), 1027-1050.

Thompson, W. C., & Cole, S., A. (2007). Psychological Aspects of Forensic Identification Evidence. In M. Costanzo, D. Krauss & K. Pezdek (Eds.), *Expert psychological testimony for the court*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Thompson, W. C., Ford, S., Doom, T., Raymer, M., & Krane, D. (2003). Evaluating forensic DNA evidence: Elements of a competent defense review. *The Champion, 27*, 16-25.

Thompson, W. C., Taroni, F., & Aitken, C. G. G. (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences, 48*(1), 47-54. doi: 10.1520/jfs2001171

Thorley, C., Baxter, R. E., & Lorek, J. (2016). The impact of note taking style and note availability at retrieval on mock jurors' recall and recognition of trial information. *Memory, 24*(4), 560-574. doi: https://doi.org/10.1080/09658211.2015.1031250

Thornton, J., & Peterson, J. (2007). The General Assumptions and Rationale of Forensic Identification. In D. Faigman, D. Kaye, M. Saks & J. Sanders (Eds.), *Modern Scientific Evidence: The Law and Science of Expert Testimony, Vol. 1,* . St. Paul, MN: West Publishing Group.

Tom R, T. (2006). Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction. *Yale Law Journal, 115*, 1050-2643. doi: https://doi.org/10.2307/20455645

Trope, Y. (1986). Identification and Inferential Processes in Dispositional Attribution. *Psychological Review, 93*(3), 239-257. doi: 10.1037/0033-295X.93.3.239

Turvey, B. E. (2013). *Forensic Fraud Evaluating Law Enforcement and Forensic Science Cultures in the Context of Examiner Misconduct*. Burlington: Elsevier Science.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Tyler, T. R. (2006). Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction. *The Yale Law Journal, 115*(5), 1050-1085. doi: https://doi.org/10.2307/20455645

Ulery, B., Hicklin, R., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings Of The National Academy Of Sciences Of The United States Of Ame, 108*(19), 7733-7738. doi: 10.1073/pnas.1018707108

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2012). Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners (Repeatability of Latent Fingerprint Decisions). *PLoS ONE, 7*(3), e32800. doi: 10.1371/journal.pone.0032800

Valentine, T. (2015). *Forensic Facial Identification Theory and Practice of Identification from Eyewitnesses, Composites and CCTV*: Wiley.

van Wallendael, L. R., Surace, A., Parsons, D. H., & Brown, M. (1994). 'Earwitness' voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology, 8*(7), 661-677. doi: 10.1002/acp.2350080705

Vanderkolk, J. (2009). *Forensic Comparative Science: Qualitative Quantitative Source Determination of Unique Impressions, Images, and Objects*: Elsevier Science.

Vidmar, N. (2005). Expert evidence, the adversary system, and the jury. *American Journal of Public Health, 95*(1), S137-S143. doi: 10.2105/AJPH.2004.044677

Vidmar, N., & Diamond, S. S. (2000). Juries and expert evidence. *Brook. L. Rev., 66*, 1121.

Vidmar, N. J., & Schuller, R. A. (1989). Juries and expert evidence: Social framework testimony. *Law and Contemporary Problems, 52*(4), 133-176. doi: https://doi.org/10.2307/1191909

Walker, L., & Monahan, J. (1987). Social frameworks: A new use of social science in law. *Virginia Law Review*, 559-598. doi: https://doi.org/10.2307/1072923

Wechsler, H. J., Kehn, A., Wise, R. A., & Cramer, R. J. (2015). Attorney beliefs concerning scientific evidence and expert witness credibility. *International Journal of Law and Psychiatry*. doi: 10.1016/j.ijlp.2015.03.008

Weinstock, M. (1999). *Epistemological understanding and argumentive competence as foundations of juror reasoning skill (Order No. 9916928).* (Doctoral dissertation), Available from ProQuest Dissertations & Theses Global. (304500849).

Weinstock, M. (2011). Knowledge-telling and knowledge-transforming arguments in mock jurors' verdict justifications. *Thinking & Reasoning, 17*(3), 282-314. doi: 10.1080/13546783.2011.575191

Weinstock, M., & Cronin, M. A. (2003). The everyday production of knowledge: individual differences in epistemological understanding and juror-reasoning skill. *Applied Cognitive Psychology, 17*, 161-181. doi: https://doi.org/10.1002/acp.860

Weinstock, M., & Flaton, R. (2004). Evidence coverage and argument skills: Cognitive factors in a juror's verdict choice. *Journal of Behavioral Decision Making, 17*(3), 191-212. doi: 10.1002/bdm.470

Wells, G. L. (1992). Naked Statistical Evidence of Liability: Is Subjective Probability Enough? *Journal of Personality and Social Psychology, 62*(5), 739-752. doi: 10.1037/0022-3514.62.5.739

Whalen, D. H., & Blanchard, F. A. (1982). Effects of photographic evidence on mock juror judgement. *Journal of Applied Social Psychology, 12*(1), 30-41. doi: https://doi.org/10.1111/j.1559-1816.1982.tb00846.x

Wheate, R. (2006). Australian juries and scientific evidence. *Australian Journal of Forensic Sciences, 38*(2), 75-84. doi: https://doi.org/10.1080/00450610609410634

White, D., Phillips, P. J., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). *Perceptual expertise in forensic facial image comparison.* Paper presented at the Proc. R. Soc. B.

Wiener, R. L., Richmond, T. L., Seib, H. M., Rauch, S. M., & Hackney, A. A. (2002). The psychology of telling murder stories: do we think in scripts, exemplars, or prototypes? *Behavioral Sciences & the Law, 20*, 119-139. doi: https://doi.org/10.1002/bsl.476

Wilson, D., & Sutton, A. (2004). Watched Over or Over-watched? Open Street CCTV in Australia. *Australian & New Zealand Journal of Criminology, 37*(2), 211-230. doi: 10.1375/acri.37.2.211

Wilson, T. D., Lisle, D. J., Kraft, D., & Wetzel, C. G. (1989). Preferences as Expectation-Driven Inferences: Effects of Affective Expectations on Affective Experience. *Journal of Personality and Social Psychology, 56*(4), 519-530. doi: 10.1037/0022-3514.56.4.519

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement, 73*(6), 913-934. doi: 10.1177/0013164413495237

Wood v R [2012] NSWCCA 21, 24 February 2012, § 4 December 2008.

Zabell, S. L. (2005). Fingerprint evidence.(Science for Judges III: Maintaining the Integrity of Scientific Research and Forensic Evidence in Criminal Proceedings). *Journal of Law and Policy, 13*(1), 143-179.

6 APPENDICES

6.1 Appendix A: Study 1 - Materials Presented to Participants

6.1.1 Information Sheet

Juror Reasoning about Evidence

In this study you will be asked to answer a few questions about yourself and will then read through a short transcript of a criminal court case. When you have reached a verdict, you will then be asked to write some short answers to questions about how you reached that verdict.

This study is expected to take no longer than 50 minutes, and has been approved by the School of Psychology ethics sub-committee.

The questionnaires and all contents of this study are anonymous: your answers will be strictly confidential. **Your participation in this study is strictly voluntary and you may withdraw from the study at any time**, without negative consequences. If you should wish to withdraw from the study before completion your answers to all questions answered will be destroyed and not used in analysis, however after completion there will be no way to trace your identity back to your answers, and thus, no way to remove your responses.

This section was only included in information sheets given to non-undergraduate student participants

At the completion of this experiment you will invited to provide your details to win a $150 JB Hi-fi voucher. The winner will be selected at random during mid-October. At that time, only the winner will be contacted. Contact details provided for this purpose will be kept separate from your study responses.

If you have any concerns about this study then please, in the first place, speak to Charlotte Scobie (email: charlotte.scobie@adelaide.edu.au) or contact the supervisor of this project, Dr. Carolyn Semmler, (email: carolyn.semmler@adelaide.edu.au; ph: 8313 4628).

Duration: 60 minutes

Where: Room 218, Hughes Building

Consent:

* I consent to take part in the research project entitled "Juror Decision Making".

* I acknowledge that I have read and understood the information provided in this information sheet.

* I have been informed that, while information gained during the study may be published, I will not be identified and my personal results will not be divulged.

* I understand that I am free to withdraw from the study at any time.

* I am aware no identifying information other than a contact email or phone number will be requested, and will be kept separately from experimental data for the purpose of entering the draw to win the JB Hifi voucher. I understand that other than the chance to win this voucher, there may be no personal benefit to myself.

Signed: ______________________________ Date: _______________

6.1.2 Consent Form

Human Research Ethics Committee (HREC)

1. I have read the attached Information Sheet and agree to take part in the following research project:

Title:	The Impact of Individual Differences in Juror Reasoning about Anthropometric Facial Comparison with Opposing Expert Testimony
Ethics Approval Number:	**13/57**

2. I have had the project, so far as it affects me, fully explained to my satisfaction by the research worker. My consent is given freely.

3. Although I understand the purpose of the research project it has also been explained that involvement may not be of any benefit to me. (*If part of payed participant pool this alternative text was shown:* Although I understand the purpose of the research project it has also been explained, that other than receiving a single payment of $15, involvement may not be of any benefit to me.)

4. I have been informed that, while information gained during the study may be published, I will not be identified and my personal results will not be divulged.

5. I understand that I am free to withdraw from the project at any time and that this will not affect my study at the University now or in the future.

6. I am aware that I should keep a copy of this Consent Form, when completed, and the attached Information Sheet.

Participant to complete:

Name: ....Signature: -------------------------------

Date:...

Researcher/Witness to complete:

I have described the nature of the research to ____

(print name of participant)

and in my opinion she/he understood the explanation.

Signature: ___________________ Position: _________________________ Date: __________

6.1.3 Participant Survey

Just a few details...

Age: ______________

Gender: ______________

Are you an Australian citizen? Yes ☐ No ☐

Is English your first language? Yes ☐ No ☐

Have you previously served as a juror? Yes ☐ No ☐

Are you currently studying, or did you study, any of the following subjects at a senior high school/tertiary level? (select all that apply)

- ☐ Biology
- ☐ Physics
- ☐ Chemistry
- ☐ Mathematics
- ☐ Legal Studies
- ☐

I didn't study any of these.

6.1.4 Transcript: "*Crown v Holt"*

Judge: Mark Holt, you are charged that on the 12th of November 2011, at Ballarat in the state of Victoria, while in the company of Pam Westman and David Turner, and armed with an offensive and dangerous weapon, you robbed Bernard Hirsch, owner of Quick-Mart, of cash and cigarettes, contrary to section 86 of the *Crimes Act*.

Mark Holt, how do you plead?

Holt: Not guilty, your Honour.

Judge: Members of the jury, Mark Holt has been charged with one count of robbery while armed with an offensive weapon. The co-accused pleaded guilty in the first day of this joint trial. The issue here is whether Mark Holt was the third offender. The prosecution will begin the trial by outlining their case against Mr Holt and summarising the evidence of their witnesses. The defence will also outline their case in an opening statement.

The standard of proof in a criminal trial is 'beyond reasonable doubt'. Therefore, it is the Crown who carries the burden of proof and the responsibility for proving the charge against the accused. If, by the end of the evidence, you decide that the Crown has not discharged this burden, you must find the accused not guilty. If you are convinced that the Crown has discharged this burden, you must find the accused guilty as charged. We will now hear an opening statement from the Crown.

Prosecutor: The Crown alleges that on the 12th of November 2011, at 4:25 am, the accused along with Pam Westman and David Turner, entered the premises of Quick-Mart while armed with a machete, and robbed owner Bernard Hirsch of $640 in cash, 157 packets of cigarettes of different brands, and a quantity of chewing gum of a particular brand. The robbery was filmed by means of a surveillance device installed on the premises. Pam Westman and David Turner were arrested 3 hours after the offence. Police searched the car in which they were travelling and located a plastic bag containing 44 packets of cigarettes and eight packets of chewing gum of the type and brand taken in the course of the robbery. One of the co-accused had $300 in bank notes in their pockets. The clothing of the co-accused matched that worn by two of the participants in the robbery as displayed in the surveillance video. The video shows a third person wearing a blue t-shirt who was holding a machete. The defendant first came to the attention of the police eight months after the robbery and was linked to the offence through fingerprints found on some of the stolen packs of cigarettes found in the car driven by the co-accused on the day of the robbery. We allege that Mark Holt was the third person who participated in the robbery, and robbed Bernard Hirsch while armed with an offensive weapon.

Judge: You will now hear an opening statement from the defence.

Defence: As you have just heard, the Crown has charged Mark Holt with one count of robbery while armed with an offensive weapon. It is true that Mark Holt was friends with Pam Westman and David Turner and has arranged meetings with them on several occasions to purchase goods from them. On the night in question Mark Holt met Pam Westman and David Turner in order to purchase marijuana. The meeting took place in the early hours of the morning, in the company of a third man who was unknown to him, in a vehicle which matches the description of the car the co-accused were driving when they were arrested. This meeting took place in a laneway in central Melbourne. Mr. Holt got into the car to buy some marijuana and six packets of Benson and Hedges full flavour. In the course of this transaction he handled various packets while looking for his preferred brand.

Although you will not hear from some of the witness directly, here is a summary of some of the evidence in this case.

Witness (owner of store): Bernard Hirsch, owner of Quick-Mart, testified that the offender who was wearing blue and who produced the machete, asked for Benson and Hedges full flavour. This is the brand the defendant says was his favourite brand. Mr. Hirsch was invited by the police to view a photographic line-up containing a photo of the accused. He was not able to make a positive identification but said there were several photos who “might be” the third offender.

Witness (forensic scientist): Rebecca Dixon, officer at the Victorian Fingerprint Unit. The accused voluntarily submitted a full set of fingerprints as part of an investigation into an unrelated event. Those fingerprints were added to the National Automated Fingerprint Identification System (NAFIS). The fingerprints taken from crime scenes throughout Australia are added to this system. Fingerprints found on several cigarette packets found in the possession of Pam Westman and David Turner matched those on the set submitted by Mark Holt.

Witness (friend of accused): Caitlin Winters, friend of the accused, gave evidence that early in November she had lent the accused a machete and a leather sheath. The machete was returned on November 27th. The accused had previously borrowed the machete and other camping gear when going on holidays.

Witness (Mark Holt, accused): Mark Holt testified that he had met Pam Westman and David Turner to purchase marijuana. He had made similar purchases from them before but was not closely acquainted with them. In early November he had called them asking if they had any marijuana to sell. They had called back a few days later in the early-morning and said they had come across a selection of cigarettes they wanted to sell at very cheap prices. Mr. Holt arranged to meet the co-accused in a side street in Melbourne about an hour later. When he arrived they were sitting in a car. A third man he did not know was standing in the shadows nearby smoking and watching the car. He looked through a garbage bag full of cigarettes to locate his preferred brand. He then left and went to work.

The following section was shown to participants in the control condition only

Judge: Although the surveillance evidence was examined by an expert in forensic anatomy, the photo-comparison results were inconclusive, and will not be admitted in this trial.

The following section was shown to participants in the two expert conditions

Prosecution: I now call Dr. Ashley Hamilton. Dr. Hamilton is an expert in the field of forensic anatomy and has a specialty in analysing faces from surveillance images.

Judge: Please raise your right hand. Do you solemnly swear or affirm to tell the truth, the whole truth and nothing but the truth?

Prosecution Expert: I do.

Prosecution: Please state your full name and occupation.

Prosecution Expert: Dr. Ashley Hamilton. I am an independent forensic science expert and researcher.

Judge: Please summarise your educational qualifications and background for the jury.

Prosecution Expert: I studied biological science at the University of Melbourne, completed Honours, and then a PhD, in which I researched anthropometry and facial reconstruction.

Prosecution: Please tell the jury how you analysed the surveillance footage and what you found.

Prosecution Expert: I was supplied with the videotape of the offence and stills extracted from it, together with forensic photographs of the accused and his two co-offenders. I played the video at different speeds, froze the video and magnified the images on my computer. I used a magnifying glass to further enhance the images. Facial mapping employs three techniques: photo-anthropometry, morphological analysis and photograph superimposition.

I first used photo-anthropometry which is a technique that attempts to metrically compare the proportional relationships of one photo to another rather than determine absolute visual similarities, as is done in morphological comparisons. It involves the analysis of anthropometric landmarks, dimensions and angles to quantify facial characteristics and proportions from a photograph.

I compared measurements and dimensions of faces by magnifying photographs of the offender to the same size as the suspect who was seen in the surveillance footage. I used photograph superimposition: a process of overlaying two comparably enlarged images through the use of computer technology to provide a visual display of the definitive similarities or differences between the offender and suspect. This was in order to see whether the features aligned or one could be overlayed over the other - this is photograph superimposition. I identified distinctive individual characteristic and habits, which I call “unique identifiers”. These are features distinctive to an individual. There were definitive resemblances with respect to the facial outline, lateral projections of the cheekbones, form and projection of the chin, jaw line, width and shape of mouth, thickness of upper and lower lips, placement of nose (i.e. the right nostril), and placement of eyes.

Morphological Analysis of the face revealed definitive resemblance of the most noticeable features. Morphological analysis is the feature by feature approach to evaluating faces, heads and bodies. It involves the comparison of two images – one from the crime scene and one of the suspect. It involves subdividing the face, head and body into components to obtain a thorough qualitative analysis and to determine visual similarities or differences.

These were a pentagonal face shape, pronounced projection of cheekbones, wide and thick upper and lower lips, wide square chin with dimple and several details of additional morphological features such as slight ear projection and black coloured hair with visible sideburns. More significantly were the definitive resemblance of the wide and thick upper and lower lips and wide square chin with dimple which constitute as ‘unique identifiers’ or most distinctive features. These enhance the support for positive identification.

Prosecutor: In your opinion, what did the photo-comparison ultimately reveal? Can you positively identify the accused, Mark Holt, as the man wearing the blue shirt in the surveillance footage?

Prosecution Expert: Given the number and quality of matches between facial features on the surveillance footage and pictures I was given of the accused, I formed the opinion that they show the same person.

Judge: Are there any questions in cross examination from the defence?

Defence: Dr. Hamilton, is there any set number of comparable feature matches that you require before you make a conclusion that you're identifying the same person in two different photos?

Prosecution Expert: Not in that sense because it's not a match of quantity, it's a match of quality. Whether you have a lot of points but yet none of them uniquely or distinctly characterizes that person it would weigh less than if you only have a small number of points but those are strong points, if they were unique identifiers.

Defence: Doesn't that mean there is a high amount of subjectivity in your analysis?

Prosecution Expert: My analysis is based on my experience and a strict set of protocols which I have developed.

Defence: CCTV footage is rarely of high quality. How do you overcome difficulties that arise from angle, lighting, and poor resolution?

Prosecution Expert: This is all done on a computer, and the police provide me with images of the suspect that are as close as possible to the angle in the surveillance footage.

Defence: So, you were told prior to making the analysis that Mark Holt was already a suspect. Isn't it possible that you were already inclined to "match" his face with that on the security camera? Did you have any other photo-matching experts verify your work or give a second opinion?

Prosecution Expert: I was the only one who examined the work, yes. But I was not biased.

Defence: Doctor, would you say that photo-matching is a highly accurate procedure?

Prosecution Expert: There is always some risk of error, of a false positive when there is no true match. But my opinion is based on experience in the lab and the field. And some of my previous work has been supported by DNA evidence.

Defence: No further questions.

Judge: Thank you, Dr. Hamilton.

The following section was shown to participants in the prosecution and opposing expert condition

Defence: I now call Dr. Louise Day. Dr. Day is an expert in the field of forensic psychology and has a specialty in facial perception.

Judge: Please raise your right hand. Do you solemnly swear or affirm to tell the truth, the whole truth and nothing but the truth?

Defence Expert: I do.

Defence: Please state your full name and occupation.

Defence Expert: Dr. Louise Day. I am an expert in the field of forensic psychology.

Judge: Please summarize your educational qualifications and background for the jury.

Defence Expert: I studied psychology at the University of Melbourne, completed Honours, and then a PhD, in which I researched perception and memory sensitivity for facial features.

Defence: You have some objections as to the evidence submitted by the expert for the prosecution?

Defence Expert: Yes, I do. In my view, Dr. Hamilton has testified using methods from an emerging discipline which is without statistical justification or sufficient analysis. It is not possible to say whether the results received were either reliable or valid.
Facial mapping is not a reliable identification technique. Experts in this area differ in the procedures they recommended and employ. I have conducted a survey of a number of practitioners of facial mapping. There have been no formal studies carried out to evaluate the validity of the techniques they use, or to calculate the reliability of results. We do not know how often errors occur, or even the likelihood that a proposed match really is a correct identification. We know that photographs can differ vastly when lighting and angles are changed. We know that very little can be done to correct for poor resolution, which is often the result that occurs when surveillance footage like the one used as evidence against the accused is slowed down and screen-captured to produce still images. We don't know if the different methods experts use to adjust for these problems actually work or how much they are able to compensate for the distortion.

There have also been many cases of experts being influenced by the contextual situation of the images and the manner in which images were presented to them, particularly where there was a subjective component to the assessment to be made. This was such a case. Here, the letter of instruction sent to Dr. Hamilton contained the opinion of the referring police officer that there had already been an identification of the suspect on the basis of fingerprint comparison. Experts tend to believe that they are immune from such influences and pressures. But research shows this is not so, and for that reason there may well have been a contamination of Dr. Hamilton's examination opinions.

Photo-comparison is different from that of fingerprint evidence where there is an empirical statistic of evidence that there are no two fingerprints the same. Similarly, with DNA evidence there is established evidence to enable an expert to say that there is a likelihood of a result occurring in a particular community. Here there is no such statistical justification for the technique nor its application. We do not know how many people have the same shape of chin, or thickness of lips. And we therefore cannot estimate how likely it would be that two people would have the same combination of features. Without that information it is not possible to reliably say that only one person would have that combination of features. There is no evidence here of identification. Dr. Hamilton at best could say that there is a resemblance.

Defence: So in your opinion, we cannot confidently say that the person shown in the surveillance footage was the accused, Mr. Holt.

Defence Expert: No, we cannot say that.

Judge: Are there any questions in cross examination from the prosecution?

Prosecution: Dr. Day, isn't it true that Dr. Hamilton's work has been verified by DNA evidence in the past? Doesn't this lend support to the techniques used?

Defence Expert: Dr. Hamilton's work may have been found to be correct previously, but a small amount of success cannot change the bias and subjectivity in this instance, and it cannot change that there is no empiric statistical evidence to say that the features compared are only going to have these particular measurements for one individual. Without knowing an error rate, we cannot know the success rate.

Prosecution: Fields like fingerprinting had to start somewhere. Photo-comparison is a groundbreaking new field and hasn't had time to gather that evidence. Scientists accepted fingerprinting without measuring everyone, without even measuring large samples. And yet it is still accepted in court today. Should this evidence be excluded in this trial just because it is new? Should we ignore the surveillance footage just because a different expert would use a different computer program to find the same result?

Defence Expert: Well, you can't say that a different expert would find the same result because no one has tested the techniques or how the application differs between experts. I am not saying that the footage should be excluded as evidence, but I am arguing against admitting misleading and potentially faulty examination of it to the jury.

Prosecution: No further questions.

Judge: Members of the jury, you have now heard all of the relevant facts and evidence in the case. It is up to you whether to accept the evidence that has been presented. If, after careful consideration, you do not accept the evidence then you are free to reject it.

When you retire to consider your verdict, you will have heard or received in court, or otherwise under my supervision, all the information that you need to make your decision. It would be unfair to both the Crown and the accused to use any material obtained outside the courtroom because the parties would not be aware of it and, therefore, would be unable to test it or make submissions to you about it.

In the Australian legal system, a defendant is presumed innocent and that the burden is on the prosecution to convince you beyond a reasonable doubt that Mark Holt participated in the burglary of Bernard Hirsch while armed with an offensive weapon. The onus is on the Crown to prove that the accused is guilty of this robbery and that burden never leaves it. It never becomes the responsibility of the accused to prove that he is not guilty or, in this case, to prove that someone else committed the robbery. Before you convict you must be satisfied *beyond a reasonable doubt*. If there is a reasonable doubt as to the accused's guilt, then the accused is entitled to the benefit of that doubt. That does not mean that you must be satisfied beyond any doubt whatever, if, indeed you can ever be satisfied of anything to that extent. The accused is not entitled to the benefit of any whimsical, fanciful, or far-fetched doubt which an agile mind might conjure up. Being satisfied beyond reasonable doubt means this—if you regard it as a reasonable possibility that someone else participated in the robbery of Bernard Hirsch, then the accused must be acquitted. If you consider this possibility so insubstantial and so remote that no reasonable person would take it into account for a moment, then you are satisfied beyond reasonable doubt and the accused must be found guilty. You must now decide whether or not the defendant, Mark Holt, is guilty as charged.

6.1.5 Argument skill questionnaire

The next section is aimed at getting you to answer questions about the trial and the possible verdicts by discussing the evidence mentioned in the trial.

It is not enough to simply say that "you agree with the prosecution/defence" or that they "made good points". You need to justify your position by discussing details from the various testimonies and information from the judge. You do not need to discuss every detail from the trial, only what you feel you need to argue your point.

You must consider **only the evidence and information presented to you** in the trial transcript. It is the right of the accused to only be judged on the information that is presented in a trial and which they have been able to respond to.

1. Why did you choose this verdict? (Please write as much or as little as you want, the lines provided are just for your convenience.)

2. What other factors went into your decision to choose that verdict?

3. Was there anything in the trial that suggested this was not the proper verdict?

4. Why didn't you choose the other verdict?

5. Other jurors might have chosen a different verdict. How might someone who chose the alternate verdict explain their choice?

6.1.6 Epistemology Questionnaire

You will now be presented with a series of statements. Your job is to read each pair, and then decide whether both statements could be valid, or whether one statement is more valid than the other.

If you think that only one opinion could be correct (**regardless of whether you agree with either statement**) then pick the first option.

If you think that both opinions could be acceptable, justifiable opinions backed up by evidence, then pick the second option. If you select this choice, then there is a further question to answer.

For example:

Peter says that doing questionnaires is the best thing in the world; he could do them all day!
Danni says that doing questionnaires is really boring and a waste of time.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be better or more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

If you have any questions please ask!

1. **Robin thinks the first piece of music they listen to is better.**
   Chris thinks the second piece of music they listen to is better.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

2. Robin thinks people should take responsibility for themselves.

Chris thinks people should work together to take care of each other.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

3. Robin has one view of why criminals keep going back to crime.

Chris has a different view of why criminals keep going back to crime.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

4. Robin believes one book's explanation of what atoms are made up of.

Chris believes another book's explanation of what atoms are made up of.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more “right” than the other?

☐ - One could be more right.

☐ - Neither is more “right” that the other.

5. Robin thinks the first painting they look at is better.

Chris thinks the second painting they look at is better.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more “right” than the other?

☐ - One could be more right.

☐ - Neither is more “right” that the other.

6. Robin thinks lying is wrong.

Chris thinks lying is permissible in certain situations.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more “right” than the other?

☐ - One could be more right.

☐ - Neither is more “right” that the other.

7. Robin thinks one book’s explanation of why the Crimean wars began is right.

Chris thinks another book’s explanation of why the Crimean wars began is right.

A) Can only one of their views be right, or could both be right?

- ☐ - Only one can be right (GO TO NEXT QUESTION)
- ☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more “right” than the other?

- ☐ - One could be more right.
- ☐ - Neither is more “right” that the other.

8. Robin believes one book’s explanation of how the brain works.

Chris believes another book’s explanation of how the brain works.

A) Can only one of their views be right, or could both be right?

- ☐ - Only one can be right (GO TO NEXT QUESTION)
- ☐ - Both can be right (GO TO B)

B) If both could be right, could one view more “right” than the other?

- ☐ - One could be more right.
- ☐ - Neither is more “right” that the other.

9. Robin thinks the first book they both read is better.

Chris thinks the second book they both read is better.

A) Can only one of their views be right, or could both be right?

- ☐ - Only one can be right (GO TO NEXT QUESTION)
- ☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more “right” than the other?

- ☐ - One could be more right.
- ☐ - Neither is more “right” that the other.

10. Robin thinks the government should limit the number of children families are allowed to have to keep the population from getting too big.

Chris thinks families should have as many children as they choose.

A) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

B) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

-

11. Robin agrees with one book's explanation of how children learn language.

Chris agrees with another book's explanation of how children learn language.

C) Can only one of their views be right, or could both be valid?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

D) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

☐ - Neither is more "right" that the other.

12. Robin believes one mathematician's proof of the math formula is right.

Chris believes another mathematician's proof of the math formula is right.

C) Can only one of their views be right, or could both be right?

☐ - Only one can be right (GO TO NEXT QUESTION)

☐ - Both can be right (GO TO B)

D) If both could be right, could one view be more "right" than the other?

☐ - One could be more right.

- ☐ - Neither is more “right” that the other.

6.2 Appendix B: Study 2 - Materials Presented to Participants

6.2.1 Information Sheet

In this study you will be asked to answer a few questions about yourself and will then read through a short transcript of a non-violent criminal court case. When you have reached a verdict, you will then be asked to write some short answers to questions about how you reached that verdict.

This study is expected to take no longer than 50 minutes, and has been approved by the University of Adelaide School of Psychology ethics sub-committee.

The questionnaires and all contents of this study are anonymous: your answers will be strictly confidential. However, your IP address will be recorded to ensure that no one can take the survey twice.

Your participation in this study is strictly voluntary and you may withdraw from the study at any time without negative consequences. If you should wish to withdraw from the study before completion your answers will be destroyed and not used in analysis, however after completion there will be no way to trace your identity back to your answers, and thus, no way to remove your responses.

If you have any concerns about this study then please, in the first place, speak to Charlotte Scobie (charlotte.scobie@adelaide.edu.au) or contact the supervisor of this project, Dr. Carolyn Semmler, (carolyn.semmler@adelaide.edu).

For questions concerning the ethical conduct of the research, please contact, Dr. Paul Delfabbro (paul.delfabbro@adelaide.edu.au)

Would you like to go into the draw to win a &150 JB HIfi voucher?

○ Yes (1)

○ No (2)

At the completion of this experiment you will be invited to provide your details to win a $150 JB Hi-fi voucher. The winner will be selected at random once the participation quota has been met (hopefully by the end of January!). At that time, only the winner will be contacted. Contact details provided for this purpose will be kept separate from your study responses.

6.2.2 Consent Form

Selecting "yes" below indicates you have read and understand the following statements:

1. I have had the project, so far as it affects me, fully explained to my satisfaction. My consent is given freely.

2. I understand the purpose of the research project and understand that involvement may not be of any benefit to me.

3. I have been informed that, while information gained during the study may be published, I will not be identified and my personal results will not be divulged.

4. I understand that I am free to withdraw from the project at any time and that this will not affect my study at the University now or in the future.

This section was only shown if participants accessed the public link

The following point concerns only those participants who are planning on entering the draw to win one of the JB Hi-Fi vouchers:

5. I am aware no identifying information other than a contact email or phone number will be requested, and will be kept separately from experimental data for the purpose of

Do you agree to take part in this study?

- Yes (1)
- No (2)

6.2.3 Instructions

In the following survey you will be doing a mixture of tasks: responding to some questions about yourself and your opinions on forensic science, reading a transcript of a criminal trial, and then coming to a verdict. Finally, you will be asked to give your opinion on aspects of the trial and also to write some short answer questions about how you reached that verdict. It is up to you how much you want to write for these questions.

It is important that you answer ALL the questions, and that you read through the transcript carefully, and at a moderate speed. You will not be able to return to pages once you have progressed to a new page. It is also important that you complete the survey in one sitting - you will not be able to save your progress and return later, or to take breaks. If an error occurs (e.g. Internet connection failure) it is possible you will be locked out of the survey and unable to complete it. If this occurs, please contact the researcher (charlotte.scobie@adelaide.edu.au).

Please imagine that you have been selected to be a juror on this trial. A juror's role is an incredibly important one, as their reasoning, evaluation of evidence, and decision making may have massive consequences on the lives of multiple individuals. You should read the transcript carefully, weigh the evidence, and pay attention to the instructions given by the judge.

Will you find the accused guilty or not guilty? To show us that you have read these instructions, please type "jury study 2014" into the "Other" box before clicking to the next page. Have you read the instructions?

○ Yes (1)

○ No (2)

○ Other (3) ____________________

6.2.4 Participant Survey

Q2.1 Are you an Australian Citizen?

◯ Yes (1)

◯ No (2)

Q2.2 Age?

______ Move the bar until your age appears on the right hand side. (1)

Q2.3 Do you work in the courts or the justice system, (e.g. practicing Lawyers, Police Officers, government employees whose duties are connected with the investigation of offences, the administration of justice or the punishment of offenders)?

◯ Yes (1)

◯ No (2)

Q2.4 *Displayed if answer to 2.3 was "Yes"* Unfortunately you are not eligible to take part in this survey. If you have any questions please contact charlotte.scobie@adelaide.edu.au

Q2.5 Gender?

◯ Male (1)

◯ Female (2)

◯ Other (3) ____________________

Q2.6 Have you previously served as a juror?

- ○ Yes (1)
- ○ No (2)

Q2.7 In what type of trial did you serve?

- ○ Civil (1)
- ○ Criminal (2)
- ○ Civil and criminal (3)

Q2.8 What is you current level of completed study?

- ○ High School (1)
- ○ Tertiary (2)
- ○ Postgraduate (3)
- ○ Trade (4)
- ○ TAFE (5)
- ○ Other (6) ____________________

Q76 Are you currently studying an Undergraduate degree?

- ○ Yes (9)
- ○ No (10)

Q77 Please enter the course title of your degree.

6.2.5 Forensic Evidence Evaluation Bias Scale

Q2.9 Juror Opinion Scale	Strongly disagree (1)	(2)	(3)	(4)	Strongly Agree (5)
Every crime can be solved with forensic science. (1)	○	○	○	○	○
Every criminal leaves some physical evidence behind at every crime scene. (2)	○	○	○	○	○
If forensic evidence suggests a defendant is guilty, this should be enough to convict even if other evidence (e.g. eyewitness testimony, alibi) suggest otherwise. (3)	○	○	○	○	○
Forensic evidence always eventually identifies the guilty person. (4)	○	○	○	○	○
Forensic evidence always provides a conclusive answer. (5)	○	○	○	○	○
Science is the most reliable way to identify the perpetrators of crimes. (6)	○	○	○	○	○
If no forensic evidence is recovered from a crime scene, it means the investigators did not look hard enough. (7)	○	○	○	○	○
If there is no forensic evidence presented in a particular case, then the jury should not convict. (8)	○	○	○	○	○
Police should not charge someone with a serious crime unless forensic evidence is available to prove their guilt. (9)	○	○	○	○	○
If no forensic evidence is recovered from a crime scene, the defendant is probably innocent of the crime. (10)	○	○	○	○	○

6.2.6 Transcript (altered sections only)

As there were only minor changes between the transcript used in study 1 and 2, the changed segments have been included in this Appendix, and the particular lines have been highlighted. All other segments remained the same.

You are now about to read the trail transcript of R v Holt. The trial should take between eight and fifteen minutes to read. Imagine that you are a juror who has been selected to be part of the panel for this trial.

To keep the transcript brief, some of it is in summary form, and is not transcribed. The summarized sections do not feature cross-examination. When reading these sections, please consider only the information that is provided to you, and not extra information that you might wish to know (e.g. extra details about relationships, further forensic tests etc.).

Once you start you cannot take breaks, and you will not be able to return to a previous page. You should read the transcript at a comfortable pace - as close as possible to your normal reading speed.

At the bottom of each page you will be able to take notes about pieces of evidence that interest you. DO NOT copy and paste sections of the trial. The text entry boxes will allow you to enter only a short amount of text and should be used as reminders of points of interest. Your responses will be made available to you as you make a verdict and answer questions in the trial later.

Ready to begin?

Judge (Introduction section): Members of the jury, Mark Holt has been charged with one count of robbery while armed with an offensive weapon. *The co-accused are being tried in a separate trial, and are refusing to testify concerning the name of the third offender.* The issue here is whether Mark Holt was the third offender. The prosecution will begin the trial by outlining their case against Mr Holt and summarising the evidence of their witnesses. The defence will also outline their case in an opening statement.

Please use this space to record evidence, pieces of testimony, or other points of interest. These points will be made available to you when answering the questions. [This was displayed at the bottom of every page]

1. (1)	
2. (2)	
3. (3)	
4. (4)	
5. (5)	

Prosecutor(opening statement): The Crown alleges that on the 12th of November 2011, at 4:25 am, the accused along with *Elliot Westman* and David Turner, entered the premises of Quick-Mart while armed with a machete, and robbed owner Bernard Hirsch of $640 in cash, 157 packets of cigarettes of different brands, and a quantity of chewing gum of a particular brand. The robbery was filmed by means of a surveillance device installed on the premises. *The video shows that three offenders were involved in the robbery but is not of a sufficient quality to enable viewers to clearly identify the offenders. Elliot Westman* and David Turner were arrested 3 hours after the offence. Police searched the car in which they were travelling and located a plastic bag containing 44 packets of cigarettes and eight packets of chewing gum of the type and brand taken in the course of the robbery. One of the co-accused had $300 in bank notes in their pockets. The clothing of the co-accused matched that worn by two of the participants in the robbery as displayed in the surveillance video. The video shows a third person wearing a blue t-shirt who was holding a machete.

The defendant first came to the attention of the police eight months after the robbery and was linked to the offence through fingerprints found on some of the stolen packs of cigarettes found in the car driven by the co-accused on the day of the robbery. We allege that Mark Holt was the third person who participated in the robbery, and robbed Bernard Hirsch while armed with an offensive weapon

These are features distinctive to an individual. There were definitive resemblances with respect to the facial outline, lateral projections of the cheekbones, form and projection of the chin, jaw line, width and shape of mouth, thickness of upper and lower lips, placement of nose (i.e. the right nostril), and placement of eyes. Morphological Analysis of the face revealed definitive resemblance of the most noticeable features. Morphological analysis is the feature by feature approach to evaluating faces, heads and bodies. It involves the comparison of two images – one from the crime scene and one of the suspect. It involves subdividing the face, head and body into components to obtain a thorough qualitative analysis and to determine visual similarities or differences. These were a pentagonal face shape, angular pronounced projection of cheekbones, wide and thick upper and lower lips, visible lips seam, wide square chin with dimple and several details of additional morphological features such as slight ear projection and black coloured hair with visible sideburns. *There was also a mezzo-cranic head shape*. More significantly were the definitive resemblance of the wide and thick upper and lower lips and wide square chin with dimple which constitute as 'unique identifiers' or most distinctive features. These enhance the support for positive identification.

Prosecutor (Expert evidence description and direct examination): In your opinion, what did the photo-comparison ultimately reveal?

Prosecution Expert: Given the number and quality of matches between facial features on the surveillance footage and pictures I was given of the accused; I would say they show a high amount of similarities.

Prosecutor: Did you also perform the analysis on the images of the other offenders who were arrested just after the robbery?

Prosecution Expert: Yes, and I found strong support for a match between images from the video and images of the other accused.

Prosecutor: And the accuracy of these techniques was backed up by DNA evidence linking the other offenders, Turner and Westman, to the crime?

Prosecution Expert: Yes. The video shows two of the offenders wearing distinctive caps. I was able to analyse the footage and identify the wearer of one cap as Turner, and the other as Westman.

Prosecutor: And these caps were found in the car and then tested for DNA?

Prosecution Expert: Yes. And traces of DNA from the hats confirmed that I had identified the correct person.

Prosecutor: So therefore we have pretty powerful support for your techniques. Your honour, I have no further questions.

Judge: Members of the jury, you have now heard all of the relevant facts and evidence in the case. It is up to you whether to accept the evidence that has been presented. If, after careful consideration, you do not accept the evidence then you are free to reject it.

When you retire to consider your verdict, you will have heard or received in court, or otherwise under my supervision, all the information that you need to make your decision. It would be unfair to both the Crown and the accused to use any material obtained outside the courtroom because the parties would not be aware of it and, therefore, would be unable to test it or make submissions to you about it.

The expert here is giving an opinion; the opinion is in the sense an opinion about the ultimate issue in the case. But only you the jury can decide that ultimate question of whether the accused is guilty or not guilty. As a jury, you hold an important and crucial role in our legal system that carries with it the burdens and responsibilities of deciding whether the accused is guilty. You have heard all of the facts deemed relevant in this trial and it is up to you as individuals and as a group to decide what they mean. Your reasoning and decision making are pivotal in deciding this issue.

In the Australian legal system, a defendant is presumed innocent and the burden is on the prosecution to convince you beyond a reasonable doubt that Mark Holt participated in the burglary of Bernard Hirsch while armed with an offensive weapon. The onus is on the Crown to prove that the accused is guilty of this robbery and that burden never leaves it. It never becomes the responsibility of the accused to prove that he is not guilty or, in this case, to prove that someone else committed the robbery. Before you convict you must be satisfied beyond a reasonable doubt. If there is a reasonable doubt as to the accused's guilt, then the accused is entitled to the benefit of that doubt. That does not mean that you must be satisfied beyond any doubt whatever, if, indeed you can ever be satisfied of anything to that extent. The accused is not entitled to the benefit of any whimsical, fanciful, or far-fetched doubt which an agile mind might conjure up. Being satisfied beyond reasonable doubt means this—if you regard it as a reasonable possibility that someone else participated in the robbery of Bernard Hirsch, then the accused must be acquitted. If you consider this possibility so insubstantial and so remote that no reasonable person would take it into account for a moment, then you are satisfied beyond reasonable doubt and the accused must be found guilty.

You must now decide whether or not the defendant, Mark Holt, is guilty as charged.

6.2.7 Experiment Responses

Here are your notes from the trial (you may need to scroll down to see the questions):

If participants made notes, they were displayed in this section

1. How do you find the defendant?

Guilty (1)

Not guilty (2)

2. Please use the slider to indicate your answer (in %)

______ How probable is it that the accused committed the crime? (1)

______ How certain are you that your verdict is the correct one? (2)

3. How would you rate the strength of the two sides in this trial?

	Very Weak (1)	Weak (2)	Somewhat Weak (3)	Average (4)	Somewhat Strong (5)	Strong (6)	Very Strong (7)
How strong was the case for the prosecution? (1)	○	○	○	○	○	○	○

How strong was the case for the defence? (2)	○	○	○	○	○	○	○

4. These questions concern the evidence presented by the facial mapping expert.

	Strongly Disagree (1)	(2)	(3)	(4)	(5)	(6)	Strongly Agree (7)
The methods used by the prosecution expert to identify the suspect were based on good scientific principles. (1)	○	○	○	○	○	○	○
The evidence presented by the prosecution expert was not reliable. (2)	○	○	○	○	○	○	○
The techniques used by the expert for the prosecution were scientifically appropriate. (3)	○	○	○	○	○	○	○
The findings of Dr. Hamilton can be used to reliably identify Mark Holt as the third offender. (4)	○	○	○	○	○	○	○
In my opinion, the procedure used by the prosecution expert to identify the suspect could protect against bias. (5)	○	○	○	○	○	○	○
The methods used by the prosecution expert cannot be used to declare a reliable match between images from	○	○	○	○	○	○	○

surveillance footage and images of a suspect. (6)	

6.2.8 Short Answer Questions

In normal jury trials, jurors are asked to enter a deliberation phase where different opinions and ideas are shared. Each opinion is important as each individual will provide a unique perspective that will contribute to the group discussion. Discussing different pieces of evidence and their interpretation means that the best verdict can be reached by the group.

Please consider what you would say if asked to explain why you believe a guilty verdict should be given.

Answer each of the following questions as if you were in this deliberation phase and talking to jurors who had seen the same trial as you. Describe the evidence or testimony that helped you reach your decision, and why you thought it was important. Remember: not all evidence may support your verdict, but is still important to consider. It is the right of the accused to only be judged on the information that is presented in a trial and which they have been able to respond to. It is also not enough to simply say that “you agree with the prosecution/defence” or that they “made good points”. You need to justify your position by discussing details from the various testimonies and information from the judge. You do not need to discuss every detail from the trial, only what you feel you need to argue your point.
Feel free to write as much as you wish for each question - but you must answer all the questions. You may also find that you are repeating yourself, or are mentioning the same pieces of evidence when answering different questions. This is normal, but please describe each relevant piece of evidence that supports your argument! If you think that you have answered a question in a previous answer, that is ok - just give the question number. There is no right or wrong opinion or a correct answer to any question.

1. Why did you pick not guilty/guilty? Particular version of question displayed depended on verdict choice

2. What other factors went into your decision to choose not/guilty?

3.a. Was there anything in the trial that suggested not guilty /guilty was not the proper verdict?

b. Why didn't this information lead you to rule out this verdict?

4. Why didn't you choose not guilty/guilty? How would you convince other jurors that this verdict is not the best choice?

5. Imagine another juror on the panel believes the not guilty/guilty verdict is the best choice. How might they explain their choice?

Thank you for completing the Jury Decision Making Experiment!

You will now be redirected to another website where you can enter your contact details.

If you have found this survey interesting, please feel free to send the link to a friend! Each person who completes it goes into the draw to win a $150 JB Hifi voucher.

If you have any questions about the survey please contact charlotte.scobie@adelaide.edu.au

6.3 Appendix C: Study 2 - Scoring Decisions for Content Analysis

6.3.1 Non-facial comparison responses

Guilty verdict justification

Code	Explanation (+ any exclusion criteria)	*Example*
Cigarette prints	Fingerprints found on recovered stolen cigarette packets that were matched with the defendant's prints	*"His fingerprints found on the cigarette packets"* (ID ekqR)
Cigarette brand	Some of the cigarettes stolen were of the favoured brand of the defendant. **Exclusion criteria:** The reference was clearly not about the defence's argument that the defendant had been looking for his preferred brand by going through the bag of the co-accused after the robbery	*"Additionally, the fact that they had the same preference for cigarettes may be coincidence, but it may also indicate that Holt was involved"* (ID Rzfb)
Machete	The defendant borrowed a machete from a friend for a time period that included the date of the robbery	*"Borrowed a machete from a friend at the time of the crime"* (ID n7AB)
Eyewitness ID	The eyewitness was presented with a line-up that included the defendant, but did not single him out instead saying there were several men you "could have been" the one. **Exclusion criteria:** Most references relating to the eyewitness ID were arguing that the identification had little probative weight as the eyewitness was unable to identify the defendant. To qualify for this category, the argument had to emphasize that the defendant was one of the few who could have been the third man.	*"...him being identified by the victim..."* (ID gEFj)
Defendant knew co-accused	The defendant was previously acquainted with the co-accused.	*"They might explain their choice by discussing the accused knew the other two men involved in the crime…"* (ID hs2d).
Defendant's character	Focus was on the character, "typical" behaviour, or morals of the defendant. **Exclusion criteria:**	*"The defense suggested that the accused was actually seeking to buy marijuana and cigarettes in secret dealing (which is*

	-unrelated to the defence's account of defendant's whereabouts on the night during the opening statement -could contain mentions of buying of stolen cigarettes, knowing co-accused etc. however the emphasis of argument was on the character. Mentions of other evidence were not counted towards other categories	*in itself illegal), this makes it seem more plausible that it is in the character of the accused to engage in an illicit activity such as theft of a Quick-mart." (*ID 2Nkb) In this instance mentions of the cigarettes would not be counted towards the "cigarettes" category as they were not used as a standalone argument.
Defendant's lack of alibi	Defendant could not provide an additional witness to support his version of the night's events (either him going to work early, being at the meeting point to buy the cigarettes, or leaving home)	"*Where was the person from work to prove that his aliby was correct?"* (ID R8Xf)
Defendant's story	The story provided by the defendant is suspicious or not believed by the participant. **Exclusion criteria:** -could contain mentions of buying of stolen cigarettes, knowing co-accused etc. but the focus was on the implausibility of the version of events given by the defence NOT the particular piece of evidence	*"it does not seem convincing that he came just after the crime had been committed to buy the cigs in the middle of the night, would henot have ben asleep if he had work early and would have maybe bought the cigs the next day at a more convenient time"* (ID j9A1) *"..but I believe that mr holt could have easily asked fr his preferred choice rather than wasting his time looking through each different branding" (ID bEaf)*
Defendant is only suspect	Participant argues that as there are no other suspects mentioned in the trial, the defendant is more likely to be guilty.	*"… no other person has been identified as the robber*" (ID ekqR)
Strength was "beyond reasonable doubt"	The case presented by the prosecution is strong enough to support a guilty verdict; minimal doubt that the defendant is innocent.	*"I would suggest that the other jurors review the evidence proving beyond reasonable doubt that he is guilty"* (ID EKEr)

Prosecution had forensic science	Prosecution had forensic evidence supporting the guilt of the defendant, with the emphasis on forensic evidence being a special type of evidence with more probative weight. *Note:* most responses including this argument were ultimately removed from the final analysis due to having little detail, or shortened responses. **Exclusion criteria:** -Doesn't mention or refer to a particular science (could indicate fingerprints, AFC, or both).	*"[Other evidence suggests that the defendant might not have been the one in the video] …forensic evidence and expertise proved otherwise"* (ID Ftu5).
Defence only had circumstantial evidence	Defence had only circumstantial evidence to present, or that the defence had no forensic evidence to offer supporting the innocence of the defendant.	"*[Defence's case] is weak, very vague and no hard evidence*" (ID bn69)

Not-guilty verdict justification

Code	Explanation (+ any exclusion criteria)	*Example*
Cigarettes (explanation of prints)	Participant accepts explanation of how prints appeared on the cigarettes.	*"…the defence presented a reasonable explanation as to why his fingerprints were on the cigarette packets. I see no particular reason why it is more likely that they were there because he committed the crime than because of the explanation he gave"* (ID tZ5f)

Eyewitness couldn't ID	The eyewitness failed to pick defendant from a police line-up but said that it "could have been" several men.	*"The store owner not being able to positively identify the accused suggested that guilty was not the proper verdict"* (ID FEe9)
Co-accused refusing to ID	The co-accused are refusing to identify the defendant as the man who robbed the store with them.	
Previous customer of co-accused	The defendant knew the two men and had previously bought goods from them.	*"The accused state that he regularly buys cigarettes from the co-accused"* (ID cjnj)
Machete (explanation)	The defendant's justification for having the machete is reasonable.	*"The fact that the machete had been previously borrowed, along with camping gear, for holidays prove an acceptable reason for having borrowed it at that time..."* (ID i2rP)
Machete wasn't found or tested	The machete borrowed by the defendant at the time of the robbery was not matched with the one seen in the video.	*"...but without some other form of corroboration (I.e. video surveillance footage identifying that it was one and the same as the machete owned by Caitlin Winters) then it was inconclusive as evidence"* (ID i2rp)
Could be innocent	Participant discusses the danger of sending a potentially innocent man to gaol.	*"A false conviction will not only ruin the accuses career, family life, etc, but will likely make a petty criminal into a hardened criminal." (*ID 5uCn)
Set up	Suggests that the defendant may have been set up by someone to take the fall.	*"MH was set up and was actually doing what he claimed"* (ID tL9v)

Possible 3rd Man	The defendant mentions he saw a third man by the car smoking when buying the cigarettes.	*"He gives a a possible third man in the ally way while he was buying cigarettes, which makes me believe that it was a different person"* (ID 5vw9)
Story/explanation	Arguments that the defendant's story as a whole is plausible. **Exclusion criteria:** -emphasis is on the scenario the defendant gives for his actions on the evening of the robbery, and *does not* include arguments about the reason for borrowing the machete, or the explanation of how the fingerprints got on the cigarettes.	*"The alibi of the defendant seems plausible"* (ID e0Fn).
Doubts	The participant has doubts about the defendant's guilt and therefore has not given a guilty verdict.	*"...because the evidence isn't strong enough for this gentleman's guilt to be considered beyond reasonable doubt"* (ID T5uV).
No "hard evidence"	Participant argues that the prosecution lacks a conclusive piece of evidence that would tie the defendant to the robbery.	*"A lot of the evidence 'suggested' he was guilty, but not completely guilty"* (ID I5B0)

6.3.2 Anthropometric facial comparison responses

Guilty verdict justification

Code	Explanation (+ any exclusion criteria)	*Example*

AFC (unspecified)	Mentions the anthropometric facial comparison evidence but no additional specific comment about why it supports guilty verdict.	*"The scientific evidence of facial recognition…"* (ID n7ab)
Qualifications	The expert's academic history or experience in the field were argued to support her findings	"*Dr. Ashley Hamilton completed biological science at University of Melbourne and completed a PhD and had focused on anthropometry and facial reconstruction"* (ID 7UcC)
Method	The method described by the expert is argued to be robust and reliable.	"*...because her analysis was clear and the science behind her technique could not be argued*" (ID lkrP).
Confidence	The expert is confident in her findings, and the participant trusts her conclusions because of this.	*"She seems confident with the methods used…"* (ID 7UcC)
Match from co-accused	Participant argues that the comparison	"*...the expert correctly identified the other two co-accused...*" (ID R8Xf)
Combination with other evidence	The AFC evidence would not be enough to prove guilt by itself, but when combined with the other evidence presented by the prosecution is enough to be beyond reasonable doubt.	*" [the AFC evidence] combined with the seemingly unlikely story of the defendant made it seem very likely that he was guilty"* (ID XkxW)

Not-guilty verdict justification

Code	Explanation (+ any exclusion criteria)	*Example*
Bias	Argues that the expert could be biased, but does not give any additional detail. **Exclusion criteria:** -criticism was that the findings, method or expert were prejudiced towards the defendant for some reason that is not specified	"*The expert opinion showed too much potential for bias and misinterpretation for me to consider it convincing*" (ID 2Nkb)
Unreliable	The conclusions or the techniques of the expert were untrustworthy. **Exclusion criteria:** -criticism is that the findings, method or expert should not be trusted, but the reason is not specified.	*"Days account, telling the court of the unreliable findings of Hamilton"* (ID DIUF)
Footage quality	Poor quality CCTV footage used in the comparison meant that the results were not reliable. This also included responses from participants in the control condition who mentioned the lack of footage in their reasoning.	*"I personally work with CCTV cameras and know how low quality the resolution of them can be, which wold make facial recognition not as reliable"* (ID OXKR)/
No statistical tests	Participant argues that the results are unreliable because there have been no empirical tests (either by the expert or the wider field) to support the findings.	*"…the expert didn't provide any sort of empirical evidence to show that she could accurately identify the defendant to any reasonable degree…"* (ID XkxW)
No peer review	The techniques used by the expert had not been evaluated by other people in the	"*The evidence of the photo-mapping doctor sounds plausible, but it is based on*

	same/similar field. Several participants referred to the AFC as being "new". This was included within this category.	*a method developed and tested by only one person, there is limited support for it's validity and reliability. Before this method is used to convict a person, I would want to [see] it replicated by other scientists"* (ID sctb)
No 2nd Expert	Test was only done by one expert, not confirmed by a different scientist.	*"I think that a single facial mapping procedure is not sufficient alone, I believe that there should have been a second opinion..." (ID 5vw9)*
Confusing language	Participant argues that the expert's language either intentionally or unintentionally obfuscated the nature of the tests or findings.	" *The expert witness tended to use complicated sounding words that camouflaged his meaning and as a lay person not cognizant with this field it felt like he was trying to baffle me with cow manure (or some similar product)"* (ID madH)
Risk of error	Participant argues that since the *expert cannot say there was a definite match* between the CCTV footage and the comparison image of the defendant, there must be some doubt about it being the defendant; OR because an error rate is mentioned, the test must not be reliable in the first place. **Exclusion criteria:** -Does not include comments about the lack of a known error rate. The few arguments	" *there's always some risk of error, as she said"* (ID Ftu5).

	commenting about the error rate were counted in the "No statistical tests" category. -Does not include comments about the tests not being unreliable.

6.4 Appendix D: Study 3 – Materials Presented to Participants

6.4.1 Information Sheet

In this study you will be asked to answer a few questions about yourself, and will then read through a short transcript of a criminal trial. When you have reached a verdict you will then be asked to rate some aspects of the trial.

As a final part, we are asking you to help us prepare a pilot study and quickly read through a newspaper article concerning a separate murder trial. In the future we will be investigating ways the media of reports criminal trials, so please read it carefully. You will be helping our next research study a lot!

Please be advised that the study trial involves a murder where a woman was thrown over a balcony, and the additional article involves a murder where a woman was shot.

This study, and reading the article, is expected to take no longer than 60 minutes, and has been approved by the University of Adelaide School of Psychology ethics sub-committee. We have included a "Save and Continue" feature in case of internet connection failure and other unforeseen issues. HOWEVER, if the survey is not completed after 4 hours your response will be deleted and you will be locked out.

The questionnaires and all contents of this study are anonymous: your answers will be strictly confidential. However, your IP address will be recorded to ensure that no one can take the survey twice**.**

Your participation in this study is strictly voluntary and you may withdraw from the study at any time without negative consequences. If you should wish to withdraw from the study before completion your answers will be destroyed and not used in analysis, however after completion there will be no way to trace your identity back to your answers, and thus, no way to remove your responses.

It is possible to do this study on a smart phone or tablet however it may require some scrolling! You must start and finish the study

If you have any concerns about this study then please, in the first place, speak to Charlotte Scobie (email: charlotte.scobie@adelaide.edu.au) or contact the supervisor of this project, Dr. Carolyn Semmler, (email: carolyn.semmler@adelaide.edu).

For questions concerning the ethical conduct of the research, please contact Dr. Paul Delfabbro (paul.delfabbro@adelaide.edu.au)

6.4.2 Consent Form

Selecting "yes" below indicates you have read and understand the following statements:

1. I have had the project, so far as it affects me, fully explained to my satisfaction. My consent is given freely.

2. I understand the purpose of the research project and understand that involvement may not be of any benefit to me.

3. I have been informed that, while information gained during the study may be published, I will not be identified and my personal results will not be divulged.

4. I understand that I am free to withdraw from the project at any time and that this will not affect my study at the University now or in the future.

This section was only shown if participants accessed the public link

The following point concerns only those participants who are planning on entering the draw to win one of the JB Hi-Fi vouchers:

5. I am aware no identifying information other than a contact email or phone number will be requested, and will be kept separately from experimental

Do you agree to take part in this study?

○ Yes (1)

○ No (2)

6.4.3 Instructions

In the following survey you will be doing a mixture of tasks: responding to some questions about yourself and your opinions on forensic science, reading a transcript of a criminal trial, and then coming to a verdict. Finally, you will be asked to give your opinion on aspects of the trial. It is up to you how much you want to write for these questions.

It is important that you answer ALL the questions, and that you read through the transcript carefully, and at a moderate speed. You will not be able to return to pages once you have progressed to a new page. It is also important that you complete the survey in one sitting. If you are unable to do this and are locked out of the survey, please contact the researcher (charlotte.scobie@adelaide.edu.au).

Please imagine that you have been selected to be a juror on this trial. A juror's role is an incredibly important one, as their reasoning, evaluation of evidence, and decision making may have massive consequences on the lives of multiple individuals. You should read the transcript carefully, weigh the evidence, and pay attention to the instructions given by the judge.

If you wish to take notes using paper and pen, please do so.

Will you find the accused guilty or not guilty? To show us that you have read these instructions, please type "jury study 2016" with NO spaces into the "Other" box before clicking to the next page.

Have you read the instructions?

- ◯ Yes (1)
- ◯ No (2)
- ◯ Other (3) ____

6.4.4 Participant Survey

Q1 Are you an Australian Citizen?

- ◯ Yes (1)
- ◯ No (2)

Q2 Age?

Move the bar until your age appears on the right hand side. (1)	

Q3 Do you work in the courts or the justice system, (e.g. practicing Lawyers, Police Officers, government employees whose duties are connected with the investigation of offences, the administration of justice or the punishment of offenders)?

- ◯ Yes (1)
- ◯ No (2)

Q4 Gender?

- ◯ Male (1)
- ◯ Female (2)
- ◯ Other (3) ____

Q5 What is your current level of *completed* study?

- ○ Year 11 or below (1)
- ○ Year 12 (2)
- ○ Certificate I to IV (3)
- ○ TAFE diploma/Advanced diploma (4)
- ○ Bachelor degree (5)
- ○ Postgraduate studies (7)
- ○ Other (6) ____

Q6 What is the field of your highest qualification?

- Natural and physical sciences (1)
- Information technology (2)
- Engineering and related technologies (3)
- Architecture and building (4)
- Agriculture, environment and related studies (5)
- Health (6)
- Education (7)
- Management and commerce (8)
- Society and culture (9)
- Creative arts (10)
- Food, hospitality and personal services (11)
- Other (12) ____________________

Q7 We are interested in how people get information about science and technology. Where do you get most of your information about science and technology? Please select all that apply

- Newspaper (including articles read online) (4)
- Magazine (including articles read online) (5)

- TED Talks (16)
- Youtube videos (17)
- Internet (Other) (6)
- Books (fiction) (7)
- Books (non-fiction) (8)
- TV (non-fiction, e.g. documentary or news feature type programs) (9)
- TV (fiction, e.g. drama or crime programs, includes movies) (10)
- Radio (11)
- Government Agency (12)
- Family/friends/colleagues (13)
- Other (14) ____________

Q8 We would now like to ask your opinion about forensic science in the courtroom.

How reliable would you consider these types of evidence?

Please answer using the slider scale from 0 to 10, with 10 being VERY RELIABLE, and 0 being NOT AT ALL RELIABLE,

Hair comparison (1)	
Bloodstain pattern (2)	
Handwriting comparison (3)	
Forensic dentistry (4)	
DNA comparison (5)	
Voice identification (6)	
Eyewitness testimony (7)	

Facial comparison (from photos) (8)	
Fingerprint comparison (9)	
Firearm identification (10)	

6.4.5 Transcript

You are now about to read the trail transcript of *R v Heinlein*. The trial should take between fifteen and thirty minutes to read.

Imagine that you are a juror who has been selected to be part of the panel for this trial.

To keep the transcript brief, some of it is in summary form, and is not transcribed. The summarized sections do not feature the direct and cross-examination, but only summaries of all the important details. When reading these sections, please **consider only the information that is provided to you**, and not extra information that you might wish to know (e.g. extra details about relationships, further forensic tests etc.).

Once you start you cannot take breaks, and you will not be able to return to a previous page.

You should read the transcript at a comfortable pace - as close as possible to your normal reading speed.

Ready to begin?

Judge: John Heinlein, you are charged with murder according to section 18 of the Crimes Act, and stand accused that in the early hours of the 1st of April, 2014, in the state of New South Wales, you pushed Veronica Barnes off the balcony of her home, with an intent to kill her. How do you plead?

Heinlein: Not guilty, your Honour.

Judge: Members of the jury, John Heinlein has been charged with one count of murder. Murder, as defined here according to section 18(1)(a) of the *Crimes Act*, is made out where a voluntary act or omission of the accused causes the death of the deceased and the act is committed with an intent to inflict grievous bodily harm, or an intent to kill. Any person who commits murder shall be guilty of an offense and shall be imprisoned for life.

The Crown has to prove beyond reasonable doubt that at the time he did the deliberate act which caused the death of Victoria Barnes, Heinlein was present at the scene of the crime and had an intention to kill the deceased, or an intention to inflict grievous bodily harm upon her. Of course, you can infer or conclude what a person's state of mind is at any particular point from a consideration of the person's state of mind leading up to that particular time and sometimes afterwards.

The standard of proof in a criminal trial is 'beyond reasonable doubt'. Therefore, it is the Crown who carries the burden of proof and the responsibility for proving the charge against the accused. If, by the end of the evidence, you decide that the Crown has not discharged this burden, you must find the accused not guilty. If you are convinced that the Crown has discharged this burden, you must find the accused guilty as charged.

A criminal trial is a pain-staking inquiry into the facts. At the end of the day, the jury is the best informed. In coming to your decision, rely only on the evidence you have heard. Once you have heard and seen the witnesses, you are the best people to assess the case and to assess the witnesses and their evidence. Keep an open mind until you decide.

We will now hear an opening statement from the Crown.

Prosecutor: The Crown alleges that on the 1st of April, 2014, John Heinlein pushed Victoria Barnes over the balcony of her apartment building.
The facts were these: Veronica Barnes and the accused had never met before the 1st of April. Veronica Barnes was a hard working massage therapist who was saving up to buy her own business. She lived alone.

It was a Friday night and she had left her lights off when going out. Tempting to any potential thief. Five other apartments were similarly broken into and had items stolen from them. Unfortunately, Ms. Barnes came home from an evening out while someone was in her apartment. There was a struggle, and then she was thrown over a balcony. Her killer removed a Bang & Olufsen speaker system, some jewellery, and her Macbook. Numerous items were stolen from apartments in the same building whose owners had similarly left their lights on.

Mr. Heinlein attempted to sell Ms. Barnes's Macbook to a Cash Converters days later. He claims to have found it on the road. He has no alibi for the two hours prior to the time police arrived on the scene. He is known to have been drunk and aggressive on that evening, and had lost his job and needed money.

Unfortunately, he left behind several fingerprints on the belt of Ms. Barnes.

This evidence was presented to a forensic expert, who will reveal to the members of the jury the method they used to test this evidence, and how it conclusively shows that Mr. Heinlein was in the apartment that night. You will be left with no doubt that Mr. Heinlein was the one to throw Ms. Barnes over the balcony to her death.

Judge: You will now hear an opening statement from the defence.

Defence: As you have heard, the Crown has charged John Heinlein with one count of murder. The evidence, however, does not support this charge.

The Crown's case primarily rests on forensic evidence that suggests Mr. Heinlein was in Ms. Barnes's apartment on that night. This evidence is not reliable and cannot prove that it was Mr. Heinlein who caused Ms. Barnes to fall over the balcony. Through close cross examination of the prosecution's expert I will show to you, members of the jury, how you should be skeptical of the evidence. This is not CSI. This is a man's future threatened by science that is hardly worthy of the name.

Ms. Barnes's apartment is in a nice area but the maintenance has not kept up with the wear and tear. Ms. Barnes's deadlock had been broken for more than 9 months. There have been multiple break in attempts to apartments in the same area in the last 2 years. The police testimony will show you that the state of the apartment indicates Ms. Barnes was robbed of several expensive items: a speaker system, worth a substantial amount of money, Ms. Barnes' jewelry, and her laptop. Four other apartments were robbed on that night and most of their missing possessions have yet to be found.

Mr. Heinlein had been out drinking that night. He stayed with some friends at the Fitzroy Hotel until 11:50pm. At about 1:30 am as he was walking down a side street he saw a garbage bag in the middle of the road. As he was curious he walked closer and saw a laptop inside. He took the laptop and decided to pawn it for money. It was not a sensible thing to do. But he was drunk and needed cash.

Mr. Heinlein only came to the attention because he was, admittedly, foolish enough to try to take the laptop to a pawn brokers. There is no reliable evidence suggesting that it was the accused in the apartment that night. The forensic evidence does not rule out the possibility that it was someone other than Mr. Heinlein who pushed Ms. Barnes over the balcony, and therefore does not provide enough support to send an innocent man to gaol.

You will now read summarized statements from a series of witnesses. You will not be reading the complete direct and cross-examination transcripts as this would take too long. Please read the statements carefully.

Callum Knight (paramedic): *Mr. Knight and his partner were the first to respond to calls from a neighbor of Ms. Barnes, who said that someone had fallen from an upper story in the apartment building and needed medical attention. Upon arriving at the apartment complex at approximately 1:05am, they entered the rear courtyard and attended to Ms. Barnes however it was obvious she was already deceased. They notified the police and remained with the body until they arrived.*

Police Officer Pileggi (first responder): *Officer Pileggi and her partner were the nearest patrol car to the apartment when the alert went out, and arrived at the courtyard at 1:20am. Open entering the courtyard they saw a body lying on the pavement. Officer Pileggi went up to Ms. Barnes apartment, while her partner remained with the body. The door to the apartment had been pulled shut, but the main lock had clearly been pried off and was lying on the floor. On her way to Ms. Barnes's apartment Officer Pileggi noticed that several other rooms had their locks similarly pried off. The lounge room and kitchen of the apartment appeared to have been the site of some disturbance as a small table and chair had been knocked over, there were scattered items on the floor, and broken glass was found in several spots (including the kitchen and by a window). The television in the living room had been knocked onto the floor, and several cabinets in the same room had their doors opened. The doors to the balcony area were open.*

Police Officer Edmond (arresting officer): *Officer Edmond was on duty on the morning of the 7th of April. The Bondi Junction police station had received a list of recently acquired items from* *the Bondi Cash Converters and the serial number from Veronica Barnes's laptop triggered an alert. Officer Edmond contacted Cash Converters and obtained the driver's license of John Heinlein which the store is required to obtain from sellers of valuable items. The police obtained a warrant to search his house, where they found some of the video games and a broken record player that matched descriptions of items stolen from other apartments on that same night. Mr. Heinlein was taken to the station and was subsequently arrested.*

Sam Walker (manager of Bondi Cash Converters): *Walker was working when Heinlein came in with a Macbook and several video games to trade on the 6th of April. He remembers that Heinlein appeared slightly hung-over and was somewhat annoyed at having to provide his license when trading in the Macbook. Walker also said that Heinlein was very eager to get the cash. The Macbook was taken into storage and another employee entered the serial number into a file for the police to check the next day.*

Graeme Harrison (Medical Examiner): *Veronica Barnes was 28 years of age when she died on the 1st of April, 2014. Ms. Barnes was declared dead at the scene by paramedics who responded to a call from her apartment building in Alexandria.*

Primary impact on the concrete surface of the inner court yard of the apartment building resulted in severe deceleration injuries to the thorax, head and face, abdominal and pelvic viscera and pelvic girdle, and vertebral column. Determining the region of the body impacting the ground first has proven difficult due to the multiplicity of injuries, and the difficulty in *excluding injuries sustained if the body 'bounced' after primary impact. Death was most likely instantaneous caused by blunt force trauma to the skull and neck.*

A forensic toxicology test using blood and stomach contents found that Ms. Barnes had alcohol in her system with levels over twice the legal limit. In conclusion, examination of the deceased showed injuries consistent with a fall from a height of above 25 meters.

Jane Harris (Property Manager): *Mrs. Harris is the property manager of Ms. Barnes's apartment building. She told the court that on the 1st April. 4 other apartments were broken into. This is the third time in two years that the complex has been robbed. Mrs. Harris said that residents frequently had visitors enter via the back alley, and through the door to the fire escape stairwell, which was often propped open with a traffic cone to stop the door from locking. This was visible from a side street. She had put up "dozens" of signs warning residents not to do this.*

Josiah Kane (Former employer of John Heinlein): *Josiah told the court that he had terminated John Heinlein's employment at the end of February. John was a competent employer although he sometimes turned up to work hung-over. He had let him go because the company had insisted their branch cut 2 positions, but Mr. Kane had rarely had cause to be dissatisfied with Heinlein's performance. Heinlein had come back to the shop once to ask for his job back and to see if Mr. Kane knew of any other opportunities for refrigeration mechanics. Mr. Kane considered John to be a good man.*

John Heinlein (accused): *Mr. Heinlein testified that he did not know Veronica Barnes and had never been to her apartment. On the night of the 31st of March he had gone drinking with a few friends to a pub they liked, the Fitzroy Hotel. While there he had gotten into an argument about the line waiting for the billiards table, and one of the players had knocked over a tray of drinks. Mr. Heinlein doesn't remember the words that were spoken but remembers being very angry and shouting at the man who had knocked over the drinks. At one point he thought the* *argument might escalate but then a bouncer had come over and asked him to leave, and after arguing with the bouncer he had done so.*

After that he told the court that he had spent some time wandering around the streets to "cool his head", as he did not want to pay for a taxi but didn't want to take the trains back home while still angry. He was walking through the streets and throwing small rocks at things, "just for something to do with his hands", when he came across a garbage bag lying on the pavement. When he threw a rock at the bag he heard it hit something hard. He then had a look in the bag and saw a laptop, some video games, and a broken vinyl record player in it. The laptop still had the cord and looked like it was new and hardly scratched. He thought it must have been thrown away because it had broken down, but he had a friend who was good with fixing computers and sometimes bought scrap. He hadn't talked to anyone while out walking. When he eventually got home that night he had tried the laptop and saw that it was working. When he saw that it still had someone's work and photos on it he was suspicious. He did think about turning it into the police but "never got around to it".

A few days later he got "a bit" drunk at home after hearing back about a few jobs that had passed him up, and when he realized he had run out of money he decided to try pawning the laptop. He said that he knew it was stupid idea but sometimes he does "stupid things" when craving alcohol. He told the court that he had never met Veronica Barnes and even though he sometimes drank too much and should have turned in the laptop, he was not a violent person, and not a criminal.

There were only minor changes in the transcript between the fingerprint, anthropometric facial comparison, and voice Identification conditions. Therefore, the full transcript of the fingerprint conditions are presented here, followed by the changed segments in the

remaining conditions, with the particular lines have been highlighted. All other segments remained the same.

The following section was shown to participants in the fingerprint conditions only:

You will now read through the transcript regarding the forensic evidence as it was taken from the court.

Prosecution: I now call Dr. Scott Andrews. Dr. Andrews is an expert in the field of forensic science and has a specialty in analyzing fingerprints.

Judge: Please raise your right hand. Do you solemnly swear or affirm to tell the truth, the whole truth and nothing but the truth?

Expert: I do.

Prosecution: Please state your full name and occupation.

Expert: Dr. Scott Andrews. I am an independent forensic anthropologist and researcher.

Prosecution: Now, starting from your bachelor's degree, could you give us all a brief rundown of your education and training?

Expert: I went to the University of Technology, Sydney, where I studied a Bachelor of Science and completed Honours, that was in 2003. I then did a Master of Science there as well

Prosecution: What did you do your Master's thesis on?

Expert: Longevity of latent prints and ridge pattern analysis.

Prosecution: Please tell the jury how you analyzed the fingerprints found at the scene and what you found.

Prosecution Expert: Fingerprint identification works by comparing patterns made from friction ridges and furrows, which are the raised and recessed formations on your thumb and fingers. In 2D print comparison we are examining the print left by the ridges, and closely examine Galton details, or points of identity, which are the tiny points in the ridges or lines. Initially you can see if you have prints worth comparing by examining the larger patterns - whorls, loops and arches. And then you examine the minute - there we look for dots, islands… they all have funny names like that. You get trifurcations or hooks etc. The high level of variation between these patterns made by the ridges is what makes a fingerprint unique. The configuration of these patterns are also unchanging as we age.

In an examination we compare the ridge characteristics and structure between a questioned sample and a known exemplar. There is also a distinction made between patent prints - which are visible to the eye and can be made from substances like blood or paint, and latent prints which are initially invisible as they are made from the natural oils and sweat secreted by the body, and need to be made visible.

In this case I was supplied with the prints taken from the patent leather belt worn by the victim at the time of her death and prints found on the balcony railing, together with comparison fingerprints taken by the police. The prints from the belt and railing had been examined with an Alternate Light Source (ALS) and treated with cyanoacrylate before using a fingerprint powder. This was done by a Scene of Crime officer. Photos were taken of these prints, and it was these photos which were used for comparison with samples from the suspect.

Initially, prints need to be examined to see if they are actually usable. After an initial examination I determined I had one usable partial thumb print and 2 separate partial finger prints. The other prints were smudged and unable to be used as they did not have enough points of comparison. These three prints were mostly clear and contained the delta, the middle whorl, of the prints.

I then compared these prints to the corresponding locations on the prints supplied to me by the police. The general features, the loops the whorls, and the minute, in which I found matching islands, dots, bifurcation etc., all show a great match between the prints.

Prosecutor: In your opinion, what did the fingerprint analysis ultimately reveal?

Prosecution Expert: Given the number and quality of matches between points in the latent prints taken from the belt of the victim and samples I was given of the accused, John Heinlein, I would say they show a high amount of similarities, and that there are no significant differences between the samples. I conclude to a reasonable degree of scientific certainty that the two fingerprint samples originate from the same person.

The following section was shown to participants in the fingerprint + scientifically informed condition only:

Judge: Are there any questions in cross examination from the defence?

Defence: Yes, thank you, Your Honour. Dr. Andrews, would your method in this analysis differ to the method used by a different expert?

Expert: There might be some small differences but I couldn't say. I have developed a method that I trust and as it has been used in previous criminal trials that have secured convictions, I know that it works.

Defence: But if we had a different expert from your field in that chair right now, they would probably be using a different technique?

Expert: It is possible.

Defence: Would they also have made the same number of matches as you or made the same measurements?

Expert: I would hope so, if their technique was sound.

Defence: But you can't say for sure?

Expert: No. But while the technique might differ in some aspects, the fundamentals of what we do are always the same: isolating small samples for comparison, and checking them.

Defence: But essentially, you are saying there is no standard protocol in this type of analysis?

Expert: Not really, no.

Defence: You said in the trial that you had a "high amount of similarities", can you explain how you selected the terminology used to express your opinion?

Expert: I chose those words because I thought they would convey to the jury what I meant most clearly - that there were enough similarities between the samples that I was confident the questioned material could have come from the source.

Defence: Is your use of "high similarities" based on a scale or a specific quantitative amount?

Expert: Not in that sense because it's not a match of quantity, it's a match of quality overall.

Defence: Doesn't that mean there is a high amount of subjectivity in your analysis, especially as you have said the method might differ between experts?

Prosecution Expert:

My analysis is based on my experience and a strict set of protocols which I have developed over the course of many years in the field. The number I selected in this case was based on the quality of the material. I wouldn't have continued the analysis if I wasn't sure of the result.

Defence: Doctor, can you direct us to any validation studies that support your analysis?

Expert: Research is always being conducted.

Defence: But do you know of any large scale population studies that actually show the statistical likelihood of combinations of ridge characteristics? How can you conclude that there are "enough" similarities? There is no empirical evidence allowing us to calculate the combination of features in a population.

Expert: I don't know of any large scale studies, but I do trust in the research that has been conducted so far, and in the history of the field.

Defence: But do you think that there is sufficient statistical justification for these types of findings?

Expert: I think that there is justification through my experience, my experience being asked to testify in court, and through the analysts like me who are employed by the police to do this type of analysis, that is justification. Besides, I am not giving statistical evidence, I just said that there was a high number of similarities. When large scale studies are done I will be happy to present statistical evidence.

Defence: Do you believe that even a sincere analyst may be influenced by cognitive and contextual effects, and not know it?

Expert: Well I think that with the right training, the right methods, that should be avoided. It would be morally wrong to submit findings of a test when you even suspect you might have been biased.

Defence: Can you tell the jury what you knew about the accused and the circumstances of this case *before* you were asked to analyse the evidence?

Expert: I knew very little. I was given photos of the latent prints taken and the comparison prints by the police, and that was it.

Defence: So you knew this was a criminal case, since the police had contacted you.

Expert: Yes.

Defence: How many comparison prints did you get?

Expert: I received a full set.

Defence: But only from one person?

Expert: Yes.

Defence: So you knew that a crime had been committed and you were only given the prints from one person, is that correct?

Expert: Yes.

Defence: You are certain that the knowledge that the police were suspecting this one man in no way influenced you to confirm that their suspect was the individual who left the prints?

Expert: I do not believe so. I believe my experience and the methodology I have created protect against this. I have been doing this work for over a decade and the police have never had reason to question my work, or my ethic.

Defence: Did you consider having a second expert who was unaware of your conclusions check your work?

Expert: No. I was confident in my findings, and had no doubts. It is also not normal procedure to do so.

Defence: Have you ever had a second person from your field, with similar training or experience, check and confirm your work?

Expert: Well, of course in my training I have had people verify it. But it is not what happens in the field. Once you have shown that you are proficient, that you can find matches or non-matches, there is no need. Do you have a second lawyer check your arguments? Besides, you just get better with experience.

Defence: But how can you know that for sure, when you have never had another expert look at your work and make sure that you haven't been influenced by the polices' suspicions, or just through random error.

Expert: Like I said, I trust in my training and experience. This field wouldn't have lasted, there wouldn't be experts in this type of analysis, if the analyses couldn't be trusted.

Defence: No further questions.

The following section was shown to participants in the fingerprint + qualification based cross examination only:

Judge: Are there any questions in cross examination from the defence?

Defence: Yes, thank you, Your Honour. Dr. Andrews, how many trials you have testified in?

Expert: I have testified in 11 trials, but I have worked in more cases than that. Above 30 I should think.

Defence: What was the reason for the cases where you were not asked to appear?
Expert: Some didn't make it to trial, in a few I think the prosecution dropped the case, in other cases the suspect accepted a plea bargain.

Defence: But overall, you have appeared in only a third of the cases you were initially consulted on?

Expert: Yes.

Defence: Earlier you said that you attend workshops each year, about 30 to 40 hours or so, is that correct?

Expert: Yes.

Defence: And some of that time is spent on training or developing skills in evidence examination?

Expert: Yes, it is.

Defence: How much time?

Expert: Pardon?

Defence: If some of the time is spent on things other than evidence presentation then I want you to tell the jury how much time is actually spent by you training for the specific, note that I say the *specific*, type of examination you used to present this evidence. How much time is spent comparing fingerprints from different surfaces, or of differing quality?

Expert: Well it would be about… it would be about a few hours. We spend more time discussing developments, new areas where the methods are being used or new computer programs. Once you have a set of skills you only need to build upon them. To refresh them, so to speak.

Defence: So in a year you spend about… 3 or 4 hours total training, or "building upon" as you say, the type of analysis that you are presenting here to support the presence of the accused.

Expert: That is correct, but as I've said, I do this type of analysis frequently and I do not consider it strictly necessary to complete workshops for skills that are in constant use and which I am proficient in. I presume you don't need to go to workshops teaching you the basics of cross examination?

Defence: <laughing> No, I do not. But what you are saying is that in your forensic work, you are only infrequently called upon to examine forensic materials, and your training is also only a few hours per year?

Expert: I do analytic work outside of what the police ask me to do, and the techniques are very similar, almost identical. I don't need the training because I am familiar with what I am doing.

Defence: Have you ever had a second person from your field, with similar training or experience, check and confirm your work?

Expert: Well, of course in my training I have had people verify it.

Defence: But have you attended any sort of proficiency testing since you started giving this type of evidence that shows that you are still performing to a high degree of accuracy?

Expert: No, I have not attended any proficiency training. However, if the police were at all dissatisfied with my work, they would not be asking me to appear in court repeatedly.

Defence: In the end, we can never be 100% sure about the guilt of suspects in a criminal trial, can we?

Expert: Well, no.

Defence: A true test of proficiency would be carried out blind, with you and the examiner not knowing until the end which in a series of comparisons were supposed to be matches or non-matches, and those answers would be known conclusively. Wouldn't that be the best way to show your expertise?

Expert: That would be, yes. But that type of testing is not currently done, in Australia or, as far as I know, internationally. I have had the training, appeared in many criminal trials, and I have developed a method that I trust and as it has been used in previous criminal trials that have secured convictions, I know that it works.

Defence: Have you ever had your work, either your research work or descriptions of your techniques, analytic work etc., published?

Expert: I had two papers from work I did in my Masters, and I also have appeared at three conferences.

Defence: So the years those papers were published, that would have been about 2005 or 2006?

Expert: One in 2005, the other in 2007. The conferences have been… Well one was at the end of my Masters, early 2006? Yes, 2006. The other two conferences were in 2010 and 2012.

Defence: Were those conference papers peer reviewed?

Expert: The abstracts were.

Defence: But not the research or the information, or the methods you were presenting?

Expert: No. Just the abstracts. But the abstracts contain all the necessary information, descriptions of what I was doing etc. My work would not have been accepted into any conferences, especially one which was in France, The European Academy of Forensic Science Conference, without them passing a review committee. **Defence:** So you haven't had any experts with similar or greater experience than you, confirm that your work was of a high, international standard, except at a conference several years ago?

Expert: No, but like I said, the police wouldn't ask me to work for them if they weren't satisfied with what I did.

Defence:
Your forensic work, has it always been with the police, appearing for the prosecution?

Expert: Yes.

Defence: Have you ever been concerned that solely working for the prosecution, in this case the police, can put analysts in a certain frame of mind, such as “needing to solve the case to get justice”, or “wanting to find the criminal”?

Expert: No, I would not allow bias to effect my work.

Defence: But it has been shown to affect other experts in forensic domains. You are certain that the knowledge that the police were asking you to do this, that you knew this was a case where a man had threatened and possibly injured or killed a defenseless woman, did not play into your mindset?

Expert:
I do not believe so. I believe my experience and the methodology I have created protect against this. I have been doing this work for over a decade and the police have never had reason to question my work, or my ethic.

Defence: But how can you know that for sure, when you have never had another expert look at your work and make sure that you haven’t been influenced by the polices’ suspicions or ideas?

Expert: Like I said, I trust in my training, experience, and method.

Defence: No further questions.

The following section was shown to participants in the anthropometric facial comparison conditions only:

Prosecution (opening statement): Unfortunately, he left behind footage on a security camera attached to the building. This evidence was presented to a forensic expert, who will reveal to the members of the jury the method they used to test this evidence, and how it conclusively shows that Mr. Heinlein was in the apartment that night. You will be left with no doubt that Mr Heinlein was the one to throw Ms. Barnes over the balcony to her death.

…

Prosecution: I now call Dr. Scott Andrews. Dr. Andrews is an expert in the field of forensic science and has a specialty in analysing photographs.

Judge: Please raise your right hand. Do you solemnly swear or affirm to tell the truth, the whole truth and nothing but the truth?

Prosecution Expert: I do.

Prosecution: Please state your full name and occupation.

Expert: Dr. Scott Andrews. I am an independent forensic anthropologist and researcher.

Prosecution: Now, starting from your bachelor's degree, could you give us all a brief run down of your education and training?

Expert: I went to the University of Technology, Sydney, where I studied a Bachelor of Science and completed Honours, that was in 2003. I then did a Master of Science there as well.

Prosecution: What did you do your Master's thesis on?

Expert: Morphological analysis and shadow patterns in photographs.

Prosecution: Please tell the jury how you analysed the security camera footage found at the scene and what you found.

Prosecution Expert: I was supplied with the videotape of the offence and stills extracted from it, together with forensic photographs of the accused. I played the video at different speeds, froze the video and magnified the images on my computer. I used a magnifying glass to further enhance the images.

Facial mapping employs three techniques: photo-anthropometry, morphological analysis and photograph superimposition.

I first used photo-anthropometry which is a technique that attempts to metrically compare the proportional relationships of one photo to another rather than determine absolute visual similarities, as is done in morphological comparisons. It involves the analysis of anthropometric landmarks, dimensions and angles to quantify facial characteristics and proportions from a photograph.

I compared measurements and dimensions of faces by magnifying photographs of the offender who was seen in the surveillance footage to the same size as the suspect. I used photograph superimposition: a process of overlaying two comparably enlarged images through the use of computer technology to provide a visual display of the definitive similarities or differences between the offender and suspect. This was in order to see whether the features aligned or one could be overlayed over the other - this is photograph superimposition. I identified distinctive individual characteristic and habits, which I call “unique identifiers”. These are features distinctive to an individual. There were definitive resemblances with respect to the facial outline, lateral projections of the cheekbones, form and projection of the chin, jaw line, width and shape of mouth, thickness of upper and lower lips, placement of nose (i.e. the right nostril), and placement of eyes.

Morphological Analysis of the face revealed definitive resemblance of the most noticeable features. Morphological analysis is the feature by feature approach to evaluating faces, heads and bodies. It involves the comparison of two images – one from the crime scene and one of the suspect. It involves subdividing the face, head and body into components to obtain a thorough qualitative analysis and to determine visual similarities or differences.

These were a pentagonal face shape, angular pronounced projection of cheekbones, wide and thick upper and lower lips, visible lips seam, wide square chin with dimple and several details of additional morphological features such as slight ear projection and black coloured hair with visible sideburns. There was also a mezzo-cranic head shape. More significantly were the definitive resemblance of the wide and thick upper and lower lips and wide square chin with dimple which constitute as 'unique identifiers' or most distinctive features. These enhance the support for positive identification.

Prosecutor: In your opinion, what did the photo-comparison ultimately reveal?

Prosecution Expert: Given the number and quality of matches between facial features on the surveillance footage and pictures I was given of the accused, I would say they show a high amount of similarities, and there are no significant differences between the samples.

The following section was shown to participants in the voice identification conditions only:

Police Officer Pileggi (first responder):

At 1:07am on the 1st of April a call came through to emergency services. The voice on the *phone was panicked and said she had come home to find a man in her apartment. She identified herself as Veronica Barnes, gave her address, and said she had surprised him in her bedroom and had fled to the bathroom to hide. On the recording a man's voice is audible, and can be heard shouting through the door, with a mixture of obscenities and death threats, before the sound of the door being kicked in and then the woman's voice screaming. The phone then cut out.*

Prosecution: I now call Dr. Scott Andrews. Dr. Andrews is a forensic expert in the field of anthropology and phonetics and has a specialty in analysing voices from recordings.

Judge: Please raise your right hand. Do you solemnly swear or affirm to tell the truth, the whole truth and nothing but the truth?

Prosecution Expert: I do.

Prosecution: Please state your full name and occupation.

Expert: Dr. Scott Andrews. I am an independent forensic anthropologist and researcher.

Prosecution: Now, starting from your bachelor's degree, could you give us all a brief run down of your education and training?

Expert: I went to the University of Technology, Sydney, where I studied a Bachelor of Science and completed Honours, that was in 2003. I then did a Master of Science there as well.

Prosecution: What did you do your Master's thesis on?

Expert: Speech patterns, and regional dialectical changes.

Prosecution: Please tell the jury how you analysed the surveillance footage and what you found.

Prosecution Expert: I was supplied with the recording of the emergency call made to the police, together with recordings made of the accused. The recordings were taken from the interviews between the chief investigating officer and the accused, with his permission, along with scripted recordings of the accused saying the lines from the emergency call. I was therefore able to edit the recordings to enhance the audio I needed.

The type of voice identification analysis I do employs two main techniques: auditory analysis and the acoustic-phonetic approach.

Auditory analysis involves the analysis of sounds by transcribing recordings using phonetic symbols and diacritics. Phoneticians who do this are able to listen to recordings and notice unusual or distinctive features of voice samples. I first examined the recordings for evidence of distinct vocal fold activity or voice quality, which is examining the way a voice "sounds". In these recordings I was able to notice that both the known and questioned samples display a degree of breathlessness during long bouts of shouting, and that the voice has characteristics you can think of as "creakiness" which can happen when airflow through the glottis is very slow. I then examined the recordings for features distinctive to a certain dialect such as the way certain phonemes are pronounced. I was able to notice that there was /l/ vocalization and the /æ/ pronunciation was not the short version of the vowel common to New South Wales; both of these variation suggest the speaker could be from South Australia.

When comparable phonetics are identified in a sample, the acoustic-phonetic approach can be used. This involves making quantitative measurements of the acoustic properties of speech so that the waveforms and spectrograms can be visually analyzed. Waveforms show a plot of amplitude, and spectrograms show frequency. This means we can examine pacing of words and the frequency content of the voice. It is important that a recording made by the police be as close as possible to the original voice recording. This is especially important when examining spectrograms. I was able to isolate the known sample from the emergency call and the questioned sample of the suspect being recorded saying the same words. I picked 12 segments - some single words and some combinations of two or three, with no overlaps of words, and compared them. The fundamental frequencies and frequency ranges, which can show the unique characteristics of an individual's voice, are nearly identical in all comparisons. The vocal expression, pronunciation of the words and voice range are also a great match.

Prosecutor: In your opinion, what did the voice-identification ultimately reveal?

Prosecution Expert: Given the number and quality of matches between voice features in the emergency call audio and samples I was given of the accused, John Heinlein, I would say they show a high amount of similarities, and that there are no significant differences between the samples. I conclude to a reasonable degree of scientific certainty that the two voice samples originate from the same person.

Prosecutor: All 12 segments were matches?

Prosecution Expert: Yes. All 12.

Prosecution: And I gather that that is incredibly rare?

Prosecution Expert: Yes, it is.

Prosecutor: Thank you, doctor. Members of the jury we therefore have support for the presence of the accused in the apartment building.

Judge: Members of the jury, you have now heard all of the relevant facts and evidence in the case. It is up to you whether to accept the evidence that has been presented. If, after careful consideration, you do not accept the evidence then you are free to reject it.

When you retire to consider your verdict, you will have heard or received in court, or otherwise under my supervision, all the information that you need to make your decision. It would be unfair to both the Crown and the accused to use any material obtained outside the courtroom because the parties would not be aware of it and, therefore, would be unable to test it or make submissions to you about it.

The expert here is giving an opinion; the opinion is in the sense an opinion about the ultimate issue in the case. But only you the jury can decide that ultimate question of whether the accused is guilty or not guilty. As a jury, you hold an important and crucial role in our legal system that carries with it the burdens and responsibilities of deciding whether the accused is guilty. You have heard all of the facts deemed relevant in this trial and it is up to you as individuals and as a group to decide what they mean. Your reasoning and decision making are pivotal in deciding this issue.

In the Australian legal system, a defendant is presumed innocent and the burden is on the prosecution to convince you beyond a reasonable doubt that John Heinlein committed the murder of Veronica Barnes. The onus is on the Crown to prove that the accused is guilty of this robbery and that burden never leaves it. It never becomes the responsibility of the accused to prove that he is not guilty or, in this case, to prove that someone else committed the robbery. Before you convict you must be satisfied beyond a reasonable doubt. If there is a reasonable doubt as to the accused's guilt, then the accused is entitled to the benefit of that doubt. That does not mean that you must be satisfied beyond any doubt whatever, if, indeed you can ever be satisfied of anything to that extent. The accused is not entitled to the benefit of any whimsical, fanciful, or far-fetched doubt which an agile mind might conjure up. Being satisfied beyond reasonable doubt means this—if you regard it as a reasonable possibility that someone else committed the murder of Veronica Barnes, then the accused must be acquitted. If you consider this possibility so insubstantial and so remote that no reasonable person would take it into account for a moment, then you are satisfied beyond reasonable doubt and the accused must be found guilty.

You must now decide whether or not the defendant, John Heinlein, is guilty as charged.

6.4.6 Experiment Responess

How do you find the defendant?

- ◯ Guilty (1)
- ◯ Not guilty (2)

Please use the slider to indicate your answer (in %)

How probable is it that the accused committed the crime? (1)	
How certain are you that your verdict is the correct one? (2)	

6.4.7 Sensitivity Check

You have now completed the main part of the study! As a second activity, we are asking you to read through a brief article and answer some questions about the trial featured in it. This will only take a few minutes.

Please scroll across to read the article.

» **CRIME**

Batro Case Rests on Evidence From Expert Dentist

"Next time you bite into an apple, take a look at the marks your teeth make"

NEXT time you bite into an apple, take a look at the marks your teeth make. That pattern could be as unique to you as your fingerprints, which is bad news for alleged murderer Michael Batro.

Batro is currently on trial in the Hobart Supreme Court. He stands accused ofmurdering his one-time lover, Stephanie Varèse, in late 2014. Batro and Varèse were engaged for 4 months earlier in the year, but Varèse called off the engagement in June after concerns that Batro was becoming "possessive" and "manipulative", the court heard last week.

Initially the case for the prosecution seemed shaky as the police were forced to admit that there were no fingerprints found on Varèse's body, and while there were several unidentified prints found on the outside of the sedan where she was found, none could be matched with Batro. The prosecution has suggested that the accused, an avid crime show watcher, may have worn gloves while strangling her and then later while dumping her body. A claim that the defence has scoffed at as "more suited to an episode of *Criminal Minds*".

But a heavy blow was dealt to the defence on Thursday when Dr. Chris Thomas presented evidence concerning a bite mark found on Varèse's thigh.

Police enlisted the services of Dr. Thomas in June. Thomas is a local forensic orthodontist (dentist) who has worked with the police for almost two decades, helping them to identify remains as well as to analyse bite marks. A graduate of The University of Technology in Sydney and The University of Adelaide in South Australia, where he completed his Graduate Diploma in Forensic Odontology in 1997, the expert told the jury that he has testified in over 20 criminal cases involving bite marks throughout Australia, 15 of which have gone on to secure convictions. He works as a normal dentist three days a week, conducting forensic work when it is needed.

Humans normally have 32 teeth, which have various sizes, functions and shapes. Forensic dentists are able to develop a dental profile from a person's mouth and teeth. Racial characteristics, orthodontic work, removals, caps, diet, and the normal wear and tear of every day eating can affect the shape of dental features, which can lead to a highly distinctive bite mark pattern.

Skin can also differ in its ability to register details, such as bruising, and we all know that there can be a loss of elasticity over time that will differ from person to person.

In the Varèse case, the coroner who examined the body noted what looked like a bite mark on the victim's thigh and photographed the marks. A wax impression was then made of Batro's teeth and, along with the photos, was sent to Thomas for examination. As Thomas did not need to check the results with another examiner, saying he was "confident and very sure" of his analysis, the turn around from the arrest of Batro to his appearance in court has been much shorter than can happen in cases waiting for DNA analysis.

Dr. Thomas demonstrates the unique bite impressions of human teeth.

Thomas told the court the bite mark on the thigh clearly showed four of the upper front teeth and five of the lower teeth, which he said was enough for a positive identification. Under cross-examination, defence lawyer Bec Brandt forced Thomas to admit that there was no standard number for comparison agreed upon by experts. The expert was not deterred, telling Brandt that the type of analysis he did uses a holistic approach as well as looking at specific points.

"From examination of the marks on the victim's thigh, and the biting edges of the accused's teeth, I am certain there is a match. The chances of it not being him are one in a million," Thomas told reporters yesterday. "I have no doubt that Batro will be in gaol by the end of the month."

However, there was a different story in the courtroom when the forensic expert was forced to redact a previous statement where he'd said that bite marks were unique to an individual, and this could lead to a "match" between samples. The objection from the defence was based on a lack of statistical evidence that could support these claims.

Despite the long history of the field, it was only recently that dental records were computer generated and kept with enough detail and precision to make them useful for comparison between dentists. This has made it hard for researchers to conduct large scale studies comparing the frequency of features. Nevertheless, the field has a rich history in the criminal justice system.

The first reported incidences of bite marks were during the Salem Witch trials, although the first time it was used in the American judicial system was in 1954 in Texas, when a burglar left a bite mark in a block of cheese. Interest in the field has increased since the 60s. Forensic odontology was even used in the trial against serial killer Ted Bundy and in attempts to identify Hitler's remains.

The Batro trial continues next week, and will be decided by jury.

How reliable would you consider the evidence mentioned in the text?

Please answer using the slider scale from 0 to 10, with 10 being VERY RELIABLE, and 0 being NOT AT ALL RELIABLE,

Forensic dentistry (4)	

Phew! All done.

If you have found this survey interesting, please feel free to send the link to a friend. Each person who completes it can go into the draw to win either a $100, $30, or $20 JB Hifi voucher. Winners will be drawn when the survey participant quota ends..

If you have any questions about the survey please contact charlotte.scobie@adelaide.edu.au

Q84 Do you want to go into the draw to win either a $100, $30, or $20 JB Hifi voucher, or to be informed of the study results? Your contact information will be kept separate from your study results.

- ○ Yes (4)
- ○ No (5)

Thank you for completing the Jury Decison Making Experiment