THE ROLE OF PRE-TRIAL ATTITUDES ABOUT FORENSIC SCIENCE EVIDENCE: DEVELOPING AND TESTING A FORENSIC EVIDENCE EVALUATION BIAS SCALE

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester by Lisa L. Smith

School of Psychology

University of Leicester

2011
The role of pre-trial attitudes about forensic science evidence: Developing and testing a forensic evidence evaluation bias scale.

Lisa L. Smith

The unique decision making task entrusted to lay juries in adversarial legal systems has attracted the attention of legal psychologists for decades, but more recently technological advances in forensic science have highlighted the importance of understanding how jurors perceive this often ambiguous and complicated type of evidence. This thesis begins by investigating the forensic awareness of lay participants, and the ability of mock jurors to discriminate between varying probative values of forensic evidence. The findings suggest that the perception of weak forensic evidence is affected by contextual information, and there was wide disagreement among participants about the probative value of weak evidence. In an effort to explain the variance in perceived evidence strength, a measure of pre-trial attitudes about forensic science was developed (the Forensic Evidence Evaluation Bias Scale – FEEBS) and administered to 446 participants ranging from students, to jury eligible members of the public, to actual jury venire persons. The results of exploratory and confirmatory factor analyses identified two distinct clusters of attitudes measured by the FEEBS, which correspond conceptually to the hypothesised juror beliefs described in the CSI Effect literature. These attitudes were found to have a significant indirect effect on verdict preference, for trial vignettes describing murder, robbery, and sexual assault scenarios containing weak (or absent) forensic DNA evidence. The implications of these findings for voir dire hearings are discussed, with reference to the cognitive models of juror decision making and the CSI Effect literature.
Acknowledgements

To my supervisors, Ray Bull and Robyn Holliday, thank you for all of your support and patience. I consider myself so lucky to have had the opportunity to work with such inspirational mentors throughout this process. I also would like to acknowledge the continued support I have received from John Bond – thank you John for being such an enthusiastic research colleague, and a never-ending source of wisdom and good advice.

To my parents, your love and encouragement have always been the source of my courage and determination and this achievement is the greatest example of this. Dad, I have you to thank for always nurturing my intellectual and adventurous nature, and without these skills I would never have made it this far. Mom, you have taught me to be creative and the importance of a strong work ethic, without which this project would never have been completed. I love you both!

To my Oma and Opa, you have always believed in me and I have always been able to count on your support and love every step of the way. I love you both, and I hope I’ve made you proud.

Last, but not least, thank you to all of my friends and family, both in Canada and the UK. You have always been there for me, and I know I can always count on you for support and advice about all things big and small – thanks.
CONTENTS

LIST OF TABLES AND FIGURES ........................................................................................................... 9

Tables .............................................................................................................................................. 9

Figures ........................................................................................................................................ 11

PUBLICATIONS & PRESENTATIONS RESULTING FROM THIS THESIS .......... 13

CHAPTER ONE: LITERATURE REVIEW ......................................................................................... 16

1.1 Jury Decision Making .................................................................................................................. 20

   1.1.2 Narrative-based models of juror decision making ............................................................... 22

   1.1.3 Probability theory models .................................................................................................... 28

   1.1.4 Decision making about physical evidence .......................................................................... 34

1.2 Measuring juror bias .................................................................................................................... 42

1.3 Attitudes toward forensic science and legal decision making .................................................. 54

1.4 Overcoming juror bias in the courtroom .................................................................................... 70

CHAPTER TWO: METHODOLOGICAL CONSIDERATIONS .................................................... 80

2.1 Mock juror paradigm ..................................................................................................................... 80

2.2 Online research ............................................................................................................................ 83

2.3 Dependent variables .................................................................................................................... 85

CHAPTER THREE: GENERAL FORENSIC AWARENESS AND EVIDENCE
EVALUATION ..................................................................................................................................... 88

3.1 Abstract .................................................................................................................................... 88

3.2 Introduction ................................................................................................................................. 89

3.3 Method ..................................................................................................................................... 91

   3.3.1 Participants .......................................................................................................................... 91

   3.3.2 Materials ............................................................................................................................. 92

   3.3.3 Procedure ........................................................................................................................... 98
### 3.4 Results ............................................................................................................................................. 99

- **3.4.1 Forensic awareness** ........................................................................................................... 99
- **3.4.2 Relationship between educational level and overall forensic awareness score** .......... 103
- **3.4.3 The relationship between forensic awareness and source of forensic knowledge** ............ 104
- **3.4.4 Forensic awareness scores by evidence type** ....................................................................... 106
- **3.4.5 Mobility and relevance as factors contributing to probative value** ............................ 109
- **3.4.6 Relationship between evidence strength and confidence** .............................................. 112
- **3.4.7 Source of forensic knowledge and evidence evaluation** ............................................... 115

### 3.5 Discussion .................................................................................................................................. 117

- **3.5.1 Overall forensic awareness** ............................................................................................ 117
- **3.5.2 Evidence evaluation** ......................................................................................................... 120
- **3.5.3 Confidence and evidence evaluation** ................................................................................ 121
- **3.5.4 Limitations and future research** ....................................................................................... 122

### CHAPTER FOUR: THE IMPACT OF CASE CONTEXT ON EVIDENCE EVALUATION ................. 124

- **4.1 Abstract** ................................................................................................................................. 124
- **4.2 Introduction** ........................................................................................................................... 124
- **4.3 Method** .................................................................................................................................. 126
  - **4.3.1 Participants** .................................................................................................................... 126
  - **4.3.2 Materials** ....................................................................................................................... 126
  - **4.3.3 Procedure** ..................................................................................................................... 127
- **4.4 Results** .................................................................................................................................. 129
  - **4.4.1 Source of forensic knowledge** ....................................................................................... 129
  - **4.4.2 Tests of evidence manipulation across conditions** ...................................................... 130
  - **4.4.3 Evaluation of the DNA evidence** .................................................................................. 131
  - **4.4.4 Probability of guilt estimates** ....................................................................................... 134
  - **4.4.5 Verdict decisions** .......................................................................................................... 137
CHAPTER FIVE: DEVELOPING AN EVIDENCE EVALUATION BIAS SCALE ... 142

5.1 Abstract ............................................................................................................. 142
5.2 Introduction ....................................................................................................... 142
5.3 Scale development .......................................................................................... 145
5.4 Reliability and item selection for initial scale .................................................. 147
  5.4.1 Pilot study .................................................................................................... 147
  5.4.2 Participants .................................................................................................. 148
  5.4.3 Materials ..................................................................................................... 148
5.5 Results ............................................................................................................... 149
  5.5.1 Item characteristics .................................................................................... 149
  5.5.2 Principal components analysis .................................................................. 152
5.6 Discussion .......................................................................................................... 156

CHAPTER SIX: VALIDATION OF THE FEEBS WITH A MURDER TRIAL SCENARIO ... 157

6.1 Abstract ............................................................................................................. 157
6.2 Introduction ....................................................................................................... 157
6.3 Method ............................................................................................................... 158
  6.3.1 Participants .................................................................................................. 158
  6.3.2 Materials ..................................................................................................... 158
  6.3.3 Procedure .................................................................................................... 160
6.4 Results ............................................................................................................... 160
  6.4.1 Principal components analysis of the FEEBS ............................................ 160
  6.4.2 Relationship with other juror attitude measures ....................................... 164
  6.4.3 Evaluation of trial evidence ....................................................................... 165
  6.4.4 Probability of guilt estimates ..................................................................... 166
  6.4.5 Verdict decisions ....................................................................................... 167
6.4.6 Relationship between juror attitudes and decision making..........................168

6.5 Discussion ........................................................................................................173

CHAPTER SEVEN: TESTING THE PREDICTIVE VALIDITY OF THE FEEBS WITH ROBBERY AND SEXUAL ASSAULT SCENARIOS.......................... 180

7.1 Abstract ........................................................................................................180

7.2 Introduction ...................................................................................................180

7.3 Method ..........................................................................................................182

7.3.1 Participants ...............................................................................................182

7.3.2 Materials ..................................................................................................182

7.3.3 Procedure ................................................................................................183

7.4 Results ..........................................................................................................184

7.4.1 Confirmatory factor analysis of the FEEBS ....................................................184

7.4.2 Comparison of juror attitude measures .......................................................188

7.4.3 Relationship between source of forensic knowledge and measures of pre-trial attitudes/bias.................................................................................190

7.4.4 Evaluation of trial evidence ......................................................................193

7.4.5 Probability of guilt estimates ...................................................................195

7.4.6 Verdict decisions ........................................................................................197

7.4.7 Juror bias measures and decision making in ‘DNA present’ conditions ....198

7.4.8 Juror bias measures and decision making in the ‘DNA absent’ conditions.....205

7.5 Discussion ....................................................................................................210

CHAPTER EIGHT: INVESTIGATING THE FEEBS AND NEGATIVE EVIDENCE TESTIMONY WITH A SAMPLE OF POTENTIAL JURORS 217

8.1 Abstract ........................................................................................................217

8.2 Introduction ...................................................................................................218

8.3 Method ..........................................................................................................219

8.3.1 Participants ...............................................................................................219

8.3.2 Materials ..................................................................................................220
8.3.3 Procedure

8.4 Results
8.4.1 Comparison of pre-trial questions
8.4.2 Scores on the FEEBS subscales
8.4.2 Evaluation of evidence
8.4.3 Probability of guilt estimates and verdict preferences

8.5 Discussion

CHAPTER NINE: OVERALL DISCUSSION
9.1 Considering the findings in the context of previous literature and theory
9.2 Theoretical and practical implications of the FEEBS
9.2.1 Theoretical basis for the FEEBS
9.2.2 Practical implications of the FEEBS
9.3 Comparisons of different samples across studies
9.4 Strengths and weaknesses of the research
9.4.1 Strengths of the research design
9.4.2 Weaknesses of the research design
9.5 Future directions
9.6 Conclusions

REFERENCES

APPENDIX A
APPENDIX B
APPENDIX C
APPENDIX D
APPENDIX E
APPENDIX F
LIST OF TABLES AND FIGURES

TABLES

Table 1. Percentage of responses to forensic awareness questionnaire for each category of agreement and overall item mean score

Table 2. Forensic awareness score by educational level

Table 3. Forensic awareness scores by source of forensic knowledge

Table 4. ANOVA summary table for forensic awareness scores by forensic knowledge source

Table 5. Mean forensic awareness scores for each evidence type

Table 6. ANOVA summary table for forensic awareness scores by evidence type

Table 7. Descriptive statistics for strength ratings by evidence category

Table 8. ANOVA summary table for strength ratings by evidence category

Table 9. Confidence ratings by strength category of evidence

Table 10. ANOVA summary table for confidence ratings by evidence strength category

Table 11. Mean strength ratings for DNA evidence in all conditions

Table 12. Mean probability of guilt estimations for each DNA evidence strength condition

Table 13. Frequency of guilty and not-guilty verdicts by evidence condition

Table 14. Percentage of participants’ strength ratings for DNA evidence in the case context condition

Table 15. Final scale items with descriptives and item-total correlations

Table 16. Summary of principal components analysis for 10-item bias scale (N=219)
Table 17. Summary of principal components analysis for 10-item bias scale (N=159)
Table 18. Descriptive statistics for scale measures
Table 19. Descriptive statistics for ratings of evidence strength
Table 20. Descriptive statistics for probability of guilt estimates
Table 21. Predictors of final probability of guilt
Table 22. Predictors of DNA strength rating
Table 23. Summary of goodness-of-fit statistics for 2-factor model of the FEEBS
Table 24. Descriptive statistics for attitude scale measures
Table 25. Correlations between attitude scale measures
Table 26. Descriptive statistics for influence ratings of sources of forensic knowledge
Table 27. Correlations between ratings of importance for sources of forensic knowledge
Table 28. Predictors of scores on the FEEBpp abd FEEBpd subscales
Table 29. Mean strength ratings of evidence types by condition
Table 30. Mean probability of guilt estimates by condition
Table 31. Frequency of verdict decisions by condition
Table 32. Multiple regression analyses predicting final probability of guilt in the robbery and sexual assault with DNA evidence conditions
Table 33. Multiple regression analyses predicting perceived strength of DNA evidence in the robbery and sexual assault with DNA evidence conditions
Table 34. Multiple regression analyses to predict probability of guilt estimates in ‘no DNA evidence’ conditions
Table 35. Multiple regression analyses to predict perceived strength of arresting officer testimony in ‘no DNA evidence’ conditions

10
Table 36. Significant Pearson correlations between responses to pre-trial questions and FEEBS subscales

Table 37. Summary of mean FEEBS scores between studies

Table 38. Mean evidence strength ratings by condition

Table 39. Multiple regression analysis predicting final probability of guilt

Table 40. Multiple regression analysis predicting perceived strength of arresting officer testimony in no DNA condition

Table 41. Multiple regression analysis predicting perceived strength of negative evidence expert testimony in no DNA plus negative evidence condition

FIGURES

Figure 1. Relative strengths of forensic evidence as a function of mobility and relevance

Figure 2. Source of forensic knowledge as indicated by participants (N=163)

Figure 3. Graphical representation of mean strength ratings by mobility and relevance of evidence

Figure 4. Strength ratings of evidence by category (mobility/relevance) and source of forensic knowledge

Figure 5. Confidence ratings by evidence category (mobility/relevance) and source of forensic knowledge

Figure 6. Participants’ reported source of forensic knowledge (N = 160)

Figure 7. Effect of case context on mean strength ratings for all evidence categories

Figure 8. Mean Probability of guilt estimates by DNA evidence condition
Figure 9. Partial mediation model for relationship between evidence bias (FEEBpp) and probability of guilt for murder trial scenario

Figure 10. Confirmatory factor analysis of the two-factor FEEB scale with standardised regression weights

Figure 11. Partial mediation model for relationship between FEEBpp and final probability of guilt in the robbery condition

Figure 12. Partial mediation model for relationship between FEEBpp and final probability of guilt in the sexual assault condition

Figure 13. Partial mediation model for relationship between FEEBpd and final probability of guilt in the robbery with no DNA condition

Figure 14. Distribution of verdict preferences by condition
The following peer reviewed articles, based on the studies conducted in this thesis, have been accepted and/or submitted for publication.


*This article contains the studies described in Chapters three and four of this thesis.


*This article contains the studies described in Chapters five and six of this thesis.


*This article contains the results presented in Chapter seven of this thesis.
The following conference presentations were based on the findings described within this thesis.


“Consider what you think justice requires, and decide accordingly. But never give your reasons; for your judgment will probably be right, but your reasons will certainly be wrong.” ~Lord Mansfield

The application of the physical sciences to criminal investigations and legal proceedings can be traced back to 1248, when a Chinese book produced medico-legal advice about distinguishing death by drowning from strangulation (Saferstein, 2007). However, the majority of early forensic science developments took place in the 1800s, which coincided with substantial discoveries in fields such as chemistry and toxicology (White, 2004). The development of a fingerprint classification system in 1901, by Edward Richard Henry, was integral in the development of identification techniques, and was closely followed in 1910 by Edmund Locard’s principle of trace evidence which revolutionised the way crime detection was approached by both the police and the scientific community (Caddy & Cobb, 2004).

These early contributions to forensic science paved the way for more recent discoveries such as DNA profiling in the 1980s by Sir Alec Jeffreys, which changed the forensic identification landscape dramatically (Williams & Johnson, 2008). Current forensic science techniques are capable of recovering and analysing a wide range of materials (e.g. biological material, glass, fibres, paint, gunshot residue) which have the potential to be used in court to establish a connection between a source and a
criminal act or crime scene. Techniques such as these are now commonplace in criminal courtrooms, and play an important role in jury and judge decisions in an ever increasing number of cases.

As well as increasing the range of types of material which can be recovered and analysed for forensic purposes, technological advances have also worked to decrease the amount of trace material required to conduct useful comparative analyses. Some material samples smaller than one nanogram can be detected and analysed using current forensic science technology (Caddy & Cobb, 2004). Although these extremely sensitive detection methods can be of great benefit to investigating crime, the interpretation of such evidence requires careful consideration of alternative explanations for the presence of tiny traces of material (Gallop & Stockdale, 2004; Rossmo, 2009).

To illustrate how advances in detection and analytic techniques have changed the way evidence is interpreted it is useful to consider a simple example of two crimes scenes, both with biological evidence recovered and analysed. In the first crime, imagine that the offender was injured during the commission of the offence, and as a result left a visible blood stain at the scene. Analysis of DNA from the blood stain confirms that it does not belong to the victim of the crime and is therefore assumed to belong to the perpetrator. When a suspect is identified, their DNA can be compared to that of the blood stain, and if a match is obtained this can be used as evidence suggestive of the guilt of that suspect. Consider a second, similar crime scene where no visible traces of biological material are apparent, however DNA evidence is recovered from extremely small samples referred to as ‘touch DNA’ or low copy
number DNA. This type of sample refers to DNA extracted from skin cells which have remained on an object after being touched by an individual (Williams & Johnson, 2008). Like the example with the visible blood stain, this DNA evidence can also be used as a comparison sample with a suspect’s DNA profile. The fundamental difference between these examples is in the probative value of each DNA sample.

In the first example the source of the DNA profile (a visible blood stain at the scene) carries a high level of probative value, as the inference about the guilt of the person who left the sample at the scene could be relatively straightforward. It is reasonable to assume that the blood stain was most likely left by a person involved in the offence, as an innocent person is unlikely to have left such a stain at the crime scene. This line of reasoning is not necessarily valid for the evidence in the second example. The DNA profile obtained from the ‘touch DNA’ evidence does not necessarily have a direct inferential link to the guilt of the person who matches the trace. This is due to the fact that with such tiny samples there are a number of potentially plausible, and perhaps innocent, explanations for the presence of the material. Issues such as contamination and secondary transference need to be considered when evaluating the probative strength of these types of evidence, and this makes reasoning about the guilt of a suspect more complicated (Gallop & Stockdale, 2004; Rossmo, 2009; R v Hoey, 2007). Champod, Lennard, Margot, and Stoilovic (2004) also highlight the risk of legally irrelevant biological traces being profiled as a direct result of increasingly sensitive material recovery techniques. These types of issues are particularly relevant in cases where the suspect has legitimate
access to the crime scene, and therefore traces of their presence are not necessarily legally relevant (Evett, 1992).

Of course, forensic evidence, such as that described above, is usually only one type of evidence that may be presented at a trial in support (or contradiction) of a suspect’s guilt. Typically the courtroom hears testimony describing different sources of evidence which may include eye-witness testimony, alibi evidence, circumstantial evidence, as well as information from any number of expert witnesses. However, there have been instances when juries appear to have decided on a verdict based almost entirely on physical evidence (or the lack thereof). One such example often cited in the literature is the case of Barry George, who was charged with the murder of Jill Dando in 2000. The Crown’s case against George focused almost exclusively on relatively ambiguous forensic evidence, namely a single particle of gunshot residue recovered from inside George’s jacket pocket and a single fibre from the crime scene which was found to be consistent with a pair of trousers owned by George (R. v George, Court of Appeal, 2007). The jury returned a majority 10-1 verdict of guilty, and the absence of any eyewitnesses, murder weapon, or clear motive leads to the conclusion that the jury found the forensic evidence to be extremely compelling. This conviction was overturned in 2008, after further consideration of the gunshot residue evidence led to the conclusion that it was unreliable and of neutral probative value (Jones, 2008). Since the Court of Appeal were not able to know how much weight the jury had given to this evidence, the conviction was considered unsafe.

The case of Barry George is not an isolated incident, and although this case featured widely in the media due to the high profile of the victim, there are many
wrongful convictions of a similar nature which go unnoticed by the public. The
Innocence Project, which works to exonerate wrongfully convicted individuals, reports
that in more than fifty percent of their DNA exonerations the cause for the wrongful
conviction was “unvalidated or improper forensic science” (Innocence Project, 2010).
Testimony of this type is thought to be overly persuasive to jurors who lack the
background knowledge to realise that such evidence may not be valid, or indicative of
guilt.

In order to understand the cognitive mechanisms underlying these
controversial jury decisions, it is first necessary to consider some of the theories which
have dominated the juror decision making literature. The following sections discuss
the most relevant cognitive theories as well as the literature investigating the
influence of juror bias and individual differences on courtroom decision making. A
more recent development in the literature is also discussed which has been referred to
as the CSI Effect, which hypothesises that the popularity of forensic science fiction in
the media is to blame for changes in juror decision making in the courtroom. Finally,
the use of peremptory challenges as a means of eliminating biased jurors is discussed,
and the effectiveness of these challenges is considered. An overview of these theories
and related research literature will form the conceptual basis for the new empirical
studies described in this thesis, and provide theoretical context for the discussions of
the results and their implications.

1.1 Jury Decision Making

In the area of legal decision making it comes as no surprise that the majority of
research seeking to model cognitive processes has focused on the jury (Hastie, 1993).
The unique situation in which jurors are placed in the courtroom attracts researchers from numerous theoretical backgrounds, all seeking to explain how jurors reach decisions during trials. The focus of jury-based studies has varied widely, with the earliest research looking mainly at how jurors’ verdicts compared to those of judges (Kalven & Zeisel, 1966). Later research throughout the 1970s and 1980s branched out to include many more variables of interest such as jury size (Saks, 1977; Velasco, 1995), demographics of jurors (Nagel & Weitzman, 1972; Bray, Struckman-Johnson, Osborne, McFarlane, & Scott, 1978; Mills & Bohannon, 1980), disposition of jurors (Bray & Noble, 1978), and the effects of the overall trial structure and group deliberation (Devine, Clayton, Dunford, Seying, & Pryce, 2001).

Traditionally, research which has attempted to model the juror’s task of integrating many pieces of evidential information during a trial has classified emerging models into categories based on the overall approach (e.g. mathematical model versus cognitive explanation) and goal of the model (e.g. prescriptive versus descriptive capabilities) (Hastie, 1993; Devine, et al., 2001). No single approach has been credited with providing a wholly adequate account of the juror decision making process, however each model accounts for some of the observed and reported aspects of the juror’s task. It is therefore important to provide an overview of some popular models, in order to understand the various approaches presented in the research literature (Pennington & Hastie, 1986).

Two of the most frequently encountered, and empirically supported, models are discussed in the following sections, and the importance of juror understanding of forensic evidence is discussed in relation to each theoretical approach. The probability
theory models and narrative-based models have been selected as these are the most commonly used frameworks in the research literature concerned with evaluating how jurors integrate evidence into their judgments of guilt.

1.1.2 NARRATIVE-BASED MODELS OF JUROR DECISION MAKING

In what is arguably the most widely accepted descriptive model of juror decision making, the story model was proposed by Pennington and Hastie (1986) to explain the cognitive process undertaken by individual jurors when deciding on a pre-deliberation verdict. Unlike previous models such as integration theory approaches and averaging models (Ostrom, Werner, & Saks, 1978), the key feature of the story model is the organisation of evidence and trial information as a narrative representation, which enables jurors to make sense of complex information and reach a verdict decision which correctly applies the relevant law to the case (Pennington & Hastie, 1986).

The story model proposes that jurors utilise a three-stage process for reaching a pre-deliberation verdict decision. The first stage of the process involves the construction of a story through which all subsequent evidence presented is then evaluated (Pennington & Hastie, 1988). This reliance on a narrative representation of trial information is a robust finding in the research literature, and occurs despite the fact that the information presented at trial is most often not communicated in this format (Huntley & Costanzo, 2003; Devine, et al., 2001). In their initial study, Pennington and Hastie (1986) presented participants sampled from a jury pool with a three hour re-enactment of a murder trial. Using a verbal protocol methodology, participants were asked to describe aloud their decision making strategy for reaching a
verdict, and the findings revealed that the trial information was consistently recalled in a story format. In addition, participants were observed to have inferred details which were missing from their stories when these facts were not part of the trial testimony. These findings highlighted the importance of jurors’ prior experiences and knowledge in shaping their individual version of the trial story, which is an important and innovative feature of the story model (Devine, et al., 2001; Levett, Danielson, Kovera, & Cutler, 2005).

During the story construction process three sources of information are thought to contribute to an individual juror’s story (Pennington & Hastie, 1992). The first, and only legally relevant, sources of information are the case-specific details. This includes the evidence presented at trial, and the testimony presented by both the prosecution and the defence and their witnesses. In addition to the trial evidence, jurors are also relying on their general knowledge about similar events as well as their overall expectations about what aspects of the trial contribute to a complete and reasonable story (Pennington & Hastie, 1988; Smith, 1993; Smith & Studebaker, 1996). This emphasis on individual jurors’ prior knowledge and experiences offers an explanation for how different jurors can construct conflicting stories to represent the same trial evidence, and ultimately reach different pre-deliberation verdict decisions.

In addition to variations in stories between jurors, it has also been noted in the literature that each individual juror may construct multiple, alternative stories throughout the course of a trial (Huntley & Costanzo, 2003). In order to decide which story is most applicable, the alternatives are assessed on four dimensions; its coverage (or how well it accounts for the evidence presented), coherence (representing the
consistency of the story), how unique the story is, and how well it fits the most appropriate verdict category (Groscup & Tallon, 2009). As well as ultimately determining the juror’s verdict preference, the choice of story has also been demonstrated to influence jurors’ memories for the evidence presented at trial, as Pennington and Hastie (1988) found that jurors tended to recognise trial evidence which supported their chosen story better than contradictory evidence. As well as influencing memory for evidence, the selected story can also influence how new evidence is perceived and how much attention this new evidence is given, which is referred to as predecisional distortion (Carlson & Russo, 2001).

The second stage of the story model process requires the juror to determine the decision alternatives which are available to them, and stage three involves a combination of the information from the first two stages as the juror reaches a decision based on which verdict alternative the preferred story best corresponds to (Groscup & Tallon, 2009). This requires that jurors understand the available verdict categories and legally relevant concepts such as mens rea, and actus reus, which are often communicated to them through judge’s instructions. The final element of this process is the juror’s pre-deliberation decision of which verdict alternative best corresponds to the particular story they have constructed.

In terms of discussing how jurors understand and evaluate different types of evidence, the first stage of the story model is particularly important, and will therefore be the focus of the remainder of this review. The latter stages in the process are equally important to the overall picture of juror decision making. However, it seems reasonable to assume that if jurors do not understand the probative value of a certain
item of evidence, and therefore incorporate it into their narrative of the crime committed incorrectly this could lead to the incorrect application of the second and third stages. So it seems that the success of the story model in guiding jurors to a logical conclusion about a defendant’s guilt depends on their ability to recognise the value and weight of various types of evidence.

The most important feature of the story model is the acknowledgement of the importance of episode schemas in guiding a juror to construct a narrative of the events leading up to the crime in question (Pennington & Hastie, 1993; Groscup & Tallon, 2009). The use of schemas and other heuristics in decision making is not limited to that of jurors making judgements of guilt or innocence, as people use these cognitive tools to assist in everyday decision making as well as more complex tasks (Gigerenzer, 2004). Dual process models of decision making predict that heuristics and prior knowledge are more likely to feature prominently in decisions when the relevant information is complex or there is substantial uncertainty associated with the information (Cooper & Neuhaus, 2000). Complex evidence and uncertainty are common (and perhaps even necessary) features of the courtroom, and the story model accounts for this by highlighting the importance of heuristics and prior knowledge within the model.

Schoemaker, (2004) describes the process of relying on heuristics and schemas as a means “to connect the new stimuli to the mental models in our heads” (p.278). This incorporation of the new stimuli (which is evidence in the case of trial information) into existing mental models can help jurors make sense of a large amount of information which is often quite ambiguous. A reliance on mental models can,
however, also result in sub-optimal reasoning and decision making. Jurors can become over-reliant on these schemas or frames, and this can result in an unwillingness to change an initial story in light of evidence that does not fit the preferred schema (Schoemaker, 2004; Carlson & Russo, 2001). It is also possible that schemas can “filter information, restrict our attention, become too engrained, and make our view seem more complete than it really is” (Schoemaker, p. 279).

The story model emerged in the early 1980s, which is prior to the discovery of sophisticated forensic technology such as DNA analysis. It is, therefore, not surprising that the majority of studies which have tested and applied the story model used trial stimuli which contained evidence mainly from witness and victim statements (Pennington & Hastie, 1988). More recent developments in forensic technology have resulted in techniques and evidence types which are increasingly complex and difficult to adequately explain to jurors (Caddy & Cobb, 2004) and the average juror is not likely to have an extensive amount of previous knowledge about these techniques or what these types of evidence are capable of asserting about the case in question (Wheate, 2006). This lack of prior knowledge and understanding could lead to uninformed stories being constructed and thus potentially inaccurate decisions being made. This potential source of error may also be exacerbated by the claims that inaccurate popular media portrayals of forensic science practices may result in jurors consulting these fictional schemas when constructing a narrative about a particular case (Podlas, 2006; Thomas, 2006).

A second issue which relates the story model directly to the interpretation of forensic evidence concerns the juror’s ability to construct and consider story
alternatives which might account for the physical evidence recovered, in the event that the defendant is innocent. For example referring back to the example at the beginning of this chapter, the ‘touch DNA’ version of the crime scene evidence requires consideration of alternative explanations for its presence as it does not support a direct inference of guilt. How would evidence such as this be represented by a juror according to the story model? Would the juror construct only a story that supports the guilt of the defendant, because DNA evidence entered by the prosecution is perceived to be indicative of guilt in most cases (based on prior knowledge or episode schemas)? Or, would the juror also construct an alternative story which accounts for the DNA sample without necessarily concluding that the defendant is guilty of the crime in question? Some decision making research literature suggests that the construction of alternative stories is not likely, as people generally stop searching for solutions to their decision problem as soon as a satisfactory answer is found (Payne & Bettman, 2004). It is not clear how the evaluation of ambiguous forensic evidence may be represented within the story model’s theoretical framework, which illustrates some further research questions which the story model could be used to investigate.

Another theory of legal reasoning which places emphasis on background knowledge and prior beliefs was proposed by Wagenaar, van Koppen, and Crombag (1993). The theory of anchored narratives places a similar importance on the narratives constructed during criminal trials, and highlights the issue of how pieces of trial evidence are required to be ‘anchored’ in common knowledge about the world, which is believed by the trier of fact. Wagenaar et al. based much of their analysis,
and indeed the original study, in the context of Dutch criminal cases which are decided solely by judges, however they raise the issue of the lack of jury accountability and the risk that this results in jurors relying on evidence which is not safely anchored in reliable fact. Indeed it could be argued that if the theory of anchored narratives explains questionable verdicts reached by Dutch judges, the impact of this heuristic on laypersons (e.g. jurors) could be greater due to their lack of specialised knowledge about legally relevant issues.

Although research applying narrative-based models to decisions made about forensic science evidence is relatively scarce, another theoretical basis for modelling decision making has been proposed which is particularly useful for conceptualising decisions about certain types of physical evidence. Often forensic evidence is presented in the courtroom using probabilities and frequencies, and so it is important to consider how well people understand this type of information and how this impacts on judgements of probative value. Therefore the next section reviews the literature which has applied probability theory models to legal decision making.

1.1.3 Probability theory models

The goal of juror decision making models based on probability theory is to represent information and decision making processes in terms of formal mathematical models. If this can be achieved, it is believed that the various elements of the decision, or the individual pieces of evidence, can be combined using formal probability formulae in order to ensure that the decision reached follows rules of consistent and logical reasoning (Aitken & Stoney, 1991; Hastie, 1993). The usefulness of probability models is not to describe how jurors are actually reaching decisions, but
rather they provide a normative calculation of the probability of a defendant’s guilt which can be used to compare with a juror’s calculation to determine the presence of any errors or biases in the decision making process (Baron, 2000).

Probability theory models are considered by some authors to be especially applicable for jury research, particularly in cases where jurors are being presented with statistical or probabilistic evidence in the form of expert witness testimony (Thompson & Schumann, 1987; Smith, Penrod, Otto, & Park, 1996; Goodman, 1992). This commonly occurs in trials when an analysis of some form of physical evidence is presented to the jury by an expert witness. Typically the crime scene evidence is associative in nature, which refers to its ability to be matched to a comparison sample originating from the defendant (Thompson & Schumann, 1987). A jury is therefore presented with the physical evidence and also hears testimony from an expert which provides the court with a number of probabilities meant to assist jurors with evaluating the probative value of the evidence (Nance & Morris, 2005; Schklar & Diamond, 1999). An example is in the case of DNA evidence presented at trial, which is typically accompanied by the probability of the sample matching the defendant as a result of coincidence (the random match probability – RMP) and in some cases the lab error rate (LE) associated with the analysis carried out (Aitken & Stoney, 1991).

One way that probability theory has been applied to research on juror decision making is the use of Bayes’ Theorem as a normative model which can be used to both evaluate jurors’ decisions as well as potentially educate decision makers about how to reach rational conclusions given some set of evidence. Bayes’ Theorem is considered by some to be ideal for use in legal decision making research, as it is well suited for
judgment tasks in which some initial probability value (e.g. the perceived probability of the defendant being guilty before hearing any evidence) is required to be updated by the presentation of a series of subsequent probabilities (e.g. probabilities associated with pieces of evidence presented during a trial) (Aitken & Stoney, 1991; Faigman & Baglioni, 1988). Although the use of Bayesian reasoning as a training technique to help jurors reach more consistent and logical decisions is supported by some authors (Faigman & Baglioni, 1988), others have objected to this over concerns that it is too complex a method for jurors to understand and utilise (Williams & Johnson, 2008).

The argument against the use of Bayesian calculations in the courtroom was articulated in the court’s judgment made during the case of *R v Adams* (1998):

“To introduce Bayes’ Theorem, or any similar method, into a criminal trial plunged the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task” (p. 482).

The usefulness of Bayes’ Theorem in relation to the logical combination of criminal evidence to determine guilt is, however, limited to cases where an objective probability can be calculated and expressed for the evidence in question. For instance, testimony concerning blood type or DNA analysis can be accompanied by probabilistic testimony based on the relative frequency of the evidence in a relevant population (e.g. genetic frequencies). In this case, the probability of the evidence being found to match the defendant in the event that the defendant is not the source can be calculated with some degree of certainty, and subsequently integrated into the calculation defined by Bayes’ Theorem (Aitken & Stoney, 1991). Probability theory is not a useful a concept when the evidence being considered cannot be associated with
probabilistic testimony, as in the case of eye-witness testimony or alibi evidence, for example. It is for this reason that jury research which uses normative models such as probability theory has mainly focused on the evaluation of physical evidence.

A body of research literature has emerged which uses Bayes’ Theorem as a normative calculation to be compared to mock jurors’ decision making concerning the probability of guilt of a defendant. These studies typically utilise a mock juror methodology, and the goal is to investigate the relative weights assigned by participants to various pieces of evidence, and the subsequent impact that this has on their final perception of defendant guilt. The most common finding is that jurors tend to under-value probabilistic evidence as compared to a Bayesian calculation (Faigman & Baglioni, 1988; Goodman, 1992; Smith, et al., 1996; Thompson & Schumann, 1987), although this is not a consistent finding across all studies (Goodman-Delahunt & Hewson, 2009; Koehler, 2001).

A number of studies have focused on jurors’ evaluations of DNA evidence specifically in determining the guilt of a defendant using Bayesian normative comparisons. A typical example of the methodology used is a study by Nance and Morris (2005) which presented mock jurors with a case summary of a rape trial and manipulated the ways in which the relevant DNA evidence was explained to the participants. As is often the case in real trials with DNA evidence, the mock jurors in this study were presented with the random match probability and the probability of laboratory error associated with the DNA match presented in the trial. The format with which these probabilities were explained to the jurors was varied in each condition (presented as a frequency, likelihood ratio, or in a chart format) in order to
determine how this manipulation affected the weight assigned to the evidence by the participants. As a measure of the baseline estimate of guilt, participants estimated the probability that the defendant was guilty prior to hearing any evidence. As described by the story model literature, this initial perception of guilt can be informed by any number of prior beliefs, experience, and pre-trial information and can be used to calculate the rational conclusion which jurors should reach if they are using the Bayesian rules of logical reasoning.

Nance and Morris (2005) reported that jurors in all conditions (that is, all presentation formats) under-valued the DNA evidence when compared to Bayesian calculations. This finding is not exclusive to this study (Faigman & Baglioni, 1988; Goodman, 1992; Smith, et al., 1996; Thompson & Schumann, 1987) and a number of possible biases or fallacies are suggested by various authors in an attempt to explain this under-valuation trend in the research literature. These explanations include the defense attorney’s fallacy which results in jurors disregarding match evidence due to the fact that they perceive there to be too many other people in the relevant population who share the DNA profile in question (Thomson & Schumann, 1987). Goodman (1992) also describes the prosecutor’s fallacy which occurs when the match probability of a piece of trial evidence is very small if the defendant is innocent, and this is mistakenly equated with the probability of the defendant’s innocence. These types of errors in the logical interpretation of the statistical information presented in court are likely to result from the ambiguity between the probability that a DNA sample from the defendant produced a match with a crime scene sample, and the
probability that the defendant was actually the person who left the DNA sample at the crime scene (Williams & Johnson, 2008).

The significance of the Bayesian normative model in studies such as that by Nance and Morris (2005) is that it gives researchers a formal calculation with which to compare the actual decision strategy used by jurors when deciding on guilt based on evidence presented at trial. Although this provides a useful tool for the evaluation of juror decision making, the probability model approach does not sufficiently model all aspects of the process of jury decision making. As mentioned earlier, the use of calculations such as those based on Bayes’ Theorem require that each piece of evidence can be objectively assigned a probability value. This is not always possible as some forms of evidence, particularly witness or victim statements, are more subjective in nature. These types of evidence are problematic for this model because one of the essential calculations in the Bayesian formula is the probability of an item of evidence occurring in the event that the defendant is innocent (Schum & Martin, 1982). If the item of evidence in question is an eye-witness account identifying the defendant, the probability of that evidence occurring if the defendant is innocent is not easily quantifiable and therefore cannot be incorporated into the calculation.

A further limitation of probability models, and in particular the use of Bayesian calculations to describe the correct interpretation and integration of evidence, is that when using this method each piece of evidence presented during the decision making process is typically treated as independent of all other pieces of evidence (Kaplan, 1982). This is not likely to reflect the reality of the situation in a real case of juror decision making, and it has therefore been suggested by Schum (as cited in Kaplan,
1982) that conditional probability should be considered when conducting research using Bayesian normative models. This is due to the fact that the appropriate weight assigned to any one piece of evidence needs to be considered in light of the previous evidence presented – in other words not treated as independent of all other evidence.

Despite the limitations, probability theory models of evidence evaluation can be a useful way to conceptualise how evidence can (and perhaps should) be integrated into the decision making task of jurors. This type of approach may be particularly relevant as more and more advanced techniques for physical evidence analysis are developed and introduced into the courtroom. As the technology advances, so too must the ability of the fact finder to understand precisely what the evidence is capable of telling them in each particular case. Although the probability theory models do not describe the process that jurors actually engage in, they do give researchers a benchmark with which to compare juror behaviour and decision making strategies. These comparisons can help to detect biases in decision making which can then be targeted as areas to improve upon within the legal system. Research focusing on juror bias has been of interest to legal psychologists for decades, and is therefore the focus of a later section of this review.

1.1.4 Decision making about physical evidence

As stated above, early research investigating how jurors interpret physical evidence in the courtroom was conducted before the introduction of DNA technology, so many of these initial studies used blood typing evidence in the trial scenarios. Although blood typing and DNA testing vary considerably in terms of sensitivity and the ability to identify individuals, they are similar in terms of how they are presented
to jurors as both include testimony which provides probabilistic information based on
the frequency of the blood type or DNA profile in the relevant population. The
previous section, which described probabilistic models of juror decision making,
discussed how Bayes’ Theorem can be used to evaluate mock jurors’ interpretations of
probabilistic testimony. The common finding in these types of studies is that mock
jurors do not interpret probabilistic testimony in ways dictated by formal
mathematical logic, and this section elaborates on these findings further by focusing
on the interpretation of physical evidence specifically.

One of the earliest studies to investigate jurors’ interpretations of physical
evidence was conducted by Faigman and Baglioni (1988). In this study mock jurors
(187 student participants) read a transcript of a commercial burglary trial in which
blood typing evidence was presented as part of the prosecution’s case. The
perpetrator was alleged to have cut their arm on a broken window while gaining entry
to the premises, and blood found at the scene matched the blood type of the
defendant. Three versions of the blood typing evidence were presented, which
manipulated the frequency of the population expected to have the same blood type as
the crime scene sample (40%, 20%, or 5%).

The results reported by Faigman and Baglioni (1988) indicated that participants
recognised that the 5% frequency evidence provided the strongest indication of the
defendant’s guilt, however they did not distinguish between the probative value of the
other two evidence frequencies. Demographic characteristics such as age, gender,
education, and mathematics background were not related to the participants’
responses, and in all conditions the blood grouping testimony was rated as more
significant than any of the other testimony (eyewitness, arresting officer, or the defendant). One of the dependent measures in this study, which seems unlikely to yield reliable information, was in a post-trial questionnaire where participants were asked to indicate how strong they felt the blood evidence was before they heard the frequency testimony. This type of question seems an unreasonable estimate to require participants to make retrospectively, as they are unlikely to be able to make this estimate accurately after completing the trial task due to the hindsight bias (Baron, 2000; Smith & Greene, 2005). Overall, the authors concluded that participants undervalued the blood typing evidence, except in the 5% frequency condition.

Similar results were reported in a study by Goodman (1992), in which participants (233 students) heard blood typing evidence associated with a murder trial. The frequency of the blood type occurring was presented as either 0.1%, 1%, 5%, or 10% of the population, and the relevant population was varied from 100, 500, 1000, to 10,000 people. The results indicated that although participants were able to recognise that the blood evidence was incriminating, they did not differentiate between the different frequencies and population sizes, as the probability of guilt estimates clustered around the same values regardless of the frequency information provided. The participants’ responses were also compared to a Bayesian calculation, and this revealed that the physical evidence was undervalued in all conditions.

With the forensic use of DNA analysis increasing through the 1990s, similar studies began to emerge in which jurors’ understanding of DNA evidence was evaluated. Schklar and Diamond (1999) described two possible sources of error which jurors are susceptible to when confronted with DNA evidence. First, jurors may make
mathematical or logical errors in their evaluation of DNA testimony, which has been demonstrated previously in studies using blood typing evidence. These errors were further categorised into two types; misaggregation errors in which the subjective belief of a defendant’s guilt is not appropriately updated in the light of new evidence, and misperception errors which occur when a juror incorrectly judges the probative value of the evidence. The misperception error is described as an expectancy-based error, which occurs when jurors rely on misinformed background beliefs, which conceptually links this type of error with narrative models of jury decision making (Pennington & Hastie, 1986; Wagenaar, et al., 1993).

Another proposed source of error is referred to as Exemplar Cueing Theory which is closely related to the availability heuristic (Koehler, 2001). This refers to the fact that people rely on how easily they can imagine examples of events when evaluating the impact of low probability occurrences. This concept explains how the presentation format of physical evidence impacts on the perceived strength of the evidence by jurors, which has been demonstrated in a number of studies (Koehler, 1993; Nance & Morris, 2005). When evidence is presented in a way that highlights the uniqueness of that evidence (e.g. very small random match probabilities) this evidence will be perceived as compelling because it is difficult to conceive of another explanation (Saks & Koehler, 2008). However, when evidence is framed in a way that supports the existence of other explanations (e.g. stating that 100 other people in the relevant population would also match the crime scene blood type) this evidence will be perceived as weaker (McQuiston-Surrett & Saks, 2009).
In addition to investigating the various errors associated with reasoning about physical evidence, some authors have also considered how the perceived strength of physical evidence compares to that of nonphysical evidence, such as eyewitness testimony. Skolnick and Shaw (2001) presented 90 students with a murder trial scenario in which the strength and type of evidence (eyewitness testimony or physical evidence in the form of footwear analysis) was manipulated to create a 2x2 factorial design. Their analysis indicated a significant effect of evidence type and strength, with the highest number of guilty verdicts in the strong physical evidence condition. They concluded that physical evidence was more persuasive than eyewitness testimony, and also highlighted that participants perceived weak physical evidence as more probative than strong eyewitness testimony.

A more recent study comparing mock juror perceptions of eyewitness testimony and DNA evidence was conducted by Pozzulo, Lemieux, Wilson, Crescini, & Girardi (2009), and the results mirror those of Skolnick and Shaw (2001). Pozzulo, et al. presented 391 student mock jurors with a fictional murder trial scenario, in which an eyewitness identification decision (positive ID, non-ID, or foil ID) and DNA evidence (consistent or inconsistent with the defendant’s guilt) was included. In addition, the DNA evidence was presented either statistically or in general (non-mathematical) terms, resulting in a 3 x 4 factorial design. The dependent variables measured in this study were not the perceived probative value of the evidence, but rather the reliability and credibility of the eyewitness and the expert witness. This selection of dependent measures makes it difficult to determine the relative strengths attributed to each piece of evidence by participants, but the results did indicate that the type of
identification and DNA evidence influenced the perceived reliability and credibility of the witnesses as well as verdict preference.

Regardless of how the DNA evidence was presented, Pozzulo et al. (2009) found that it had a stronger influence on participants’ verdict decisions than the eyewitness testimony, and this is consistent with previous findings (Skolnick & Shaw, 2001). One aspect of the trial materials which was not discussed by Pozzulo et al. is the source of the DNA evidence, and how this might impact on the interpretation of their results. In the Pozzulo et al. study the murder trial was a fatal stabbing outside of a pub, and the DNA evidence was extracted from saliva on discarded cigarette ends found outside the pub. Considering the fact that cigarette ends discarded outside a pub are a likely occurrence, and therefore not necessarily directly linked to the crime in question, it is particularly interesting that this evidence was found to be so compelling to their participants.

One of the methodological issues raised in these types of study is the manipulation of evidence strength. Categorising independent variables such as physical or eyewitness evidence as strong or weak is usually a subjective judgment, based on the researcher’s perception of probative value. There is no objective manipulation of strength used, and therefore it it difficult to evaluate whether participants are making reasonable decisions and even more difficult to compare findings across studies (Anderson, Schum, & Twinning, 2005). The studies conducted within this thesis have attempted to overcome this issue through the use of more legally relevant definitions of physical evidence strength, based on two important features of forensic evidence; mobility and relevance (Kraemer & Weber, 2004;
Roberts & Zuckerman, 2004). These concepts will be explained in more detail in Chapter two.

Early research focusing on jurors’ interpretations of forensic evidence largely ignored the range of other decision making about forensic science evidence which takes place in the Criminal Justice System. Investigating police officers make decisions about what physical evidence to process, forensic laboratory analysts need to interpret the results of their analyses, the Crown Prosecution Service decides whether there is sufficient probative evidence to proceed to prosecute a case, and the Judge is required to decide what evidence is admissible for a particular case. There is a limited amount of research into these various decisions, however recently there have been questions raised about the extent to which forensic experts are capable of objectively analysing and interpreting evidence, which has led to a new trend of research questions in the literature.

Previous literature has assumed a certain amount of objectivity associated with physical evidence (Cooley, 2007), and only recently has research challenged this assumption by investigating the possibility that forensic evidence is in fact subjectively interpreted (Saks & Koehler, 2005; Ungvarsky, 2006). The main focus of such research has been on the impact of contextual information on experts’ objective analysis of forensic evidence. The importance of these issues, and the need for research in this area, was highlighted in the recent National Academy of Science (NAS) report entitled ‘Strengthening Forensic Science in the United States: A Path Forward’ (2009). This report cautioned against the common perception of forensic science as objective evidence, and highlighted several human biases that may have profound effects on the
reliability and admissibility of physical evidence. While the Law Commission in the UK takes a different approach focusing on the admissibility criteria for forensic science evidence (The Law Commission, 2009), the position taken in the NAS report is supported by Wheate and Jamieson (2009) who argue that issues of admissibility should come second to careful consideration of the reliability of forensic science techniques and the interpretation of evidence.

Dror, Charlton, and Peron (2006) were among the first researchers to advocate the importance of investigating what impact the context of a case may have on experts’ ability to accurately analyse physical evidence. In their first study fingerprint experts (N = 5), with an average of 17 years of experience, were presented with fingerprints which (unknown to them) they had previously examined in their career and had found to match one another. When presented with these fingerprints, accompanied by case context suggesting that they were not a match, three of the five experts concluded that they were not a match (which was contrary to their initial opinion of the fingerprints), one considered the comparison to be inconclusive, and one maintained their original opinion that the fingerprints matched. Despite the very small sample size, this was the first study to demonstrate the potential impact of case context and associated pressures on experts’ interpretations of physical evidence.

Dror and Cole (2010) suggest that this type of research is crucial, particularly given that fingerprint examination is becoming even more cognitively difficult with the ever increasing number of entries in national databases. Indeed, this is true of any forensic comparative practices such as footwear, bite marks, and ballistics examinations.
The study described in Chapter four of this thesis investigated the importance of case context on participants’ perceptions of forensic evidence, and in particular their ability to differentiate between varying strengths of evidence presented with and without case context. The results from this study formed the basis for the remaining chapters which sought to explain wide variations in participant responses by attempting to devise a measure of juror bias for forensic evidence. Therefore the next section of this chapter reviews some of the previous research literature on measuring juror bias, and on the relationship between pre-trial attitudes and verdict preference.

1.2 MEASURING JUROR BIAS

The importance of jurors’ prior knowledge and attitudes, as highlighted by the story model, has provided theoretical support for research aimed at identifying and measuring juror bias. Lawyers and judges have been intuitively aware of the impact of individual differences on jurors’ decisions for decades, and there have been numerous efforts to identify personality and demographic variables which reliably predict juror bias (Ellsworth, 1993). The goal of identifying pre-trial bias in potential jurors is of particular interest in countries (e.g. the United States) where pre-trial hearings allow both the prosecution and defence to evaluate and eliminate jurors that they believe would not be able to render a fair and unbiased decision (Feinman, 2000). This practice of jury selection (by exclusion) seems contrary to the idealised concept of a jury being comprised of a random and representative selection of citizens. However, it does reflect another fundamental element of the adversarial justice system which requires juries to be neutral and unbiased decision makers (Van Dyke, 1977; Levesque, 2006). Although the practice of exercising peremptory challenges originated in the UK,
this process was almost entirely abolished by the Criminal Justice Act 1988 (Wheeler, 2006). However, current courtroom procedures in England and Wales do allow the prosecution and defence counsel to object to any number of jurors before each juror takes the oath as outlined in the Juries Act 1974 S.12 (1)(b), although this is rarely used (Wheeler, 2006).

Juror pre-trial bias is not thought to be necessarily strong enough to overcome basic common sense in all trial situations. An interactionist approach, adopted by most of the jury bias research, views juror behaviour as influenced by both the situation (e.g. the specific trial and accompanying evidence) and the individual (e.g. juror personality, beliefs, attitudes, etc.) (Kassin & Wrightsman, 1988). In the case of a trial with very strong evidence, the situational factors will be especially salient and therefore this approach would predict that individual juror differences would play a relatively insignificant role in the decisions made by the jury (Ruva, McEvoy, & Bryant, 2007). However, in cases where the trial evidence is weak or ambiguous the lack of clear situational cues will result in a greater influence of juror biases and heuristics (Kaplan & Miller, 1978; Kassin & Wrightsman, 1988; Bull Kovera, McAuliff, & Hebert, 1999). This has been referred to as the Liberation Hypothesis by Kalven and Zeisel (as cited in Devine, Buddenbaum, Houp, Studebaker, & Stolle, 2009), who describe situations in which trial evidence is unclear or ambiguous leading to jurors being ‘liberated’ from the constraints of the trial evidence, leaving them free to rely on their pre-trial attitudes, beliefs, and sentiments about the trial participants.

Early research investigating the link between juror characteristics and decision making focused on demographics (age, gender, race, SES, etc.) (Sealy, 1981) and
personality constructs such as Locus of Control (Rotter, 1990), Authoritarianism (Jensen, 1957), and Belief in a Just World (Rubin & Peplau, 1975). The goal of this research was to identify juror characteristics that could reliably predict verdict preference. Most of these measures were not consistently successful at predicting verdict choices (Sealy, 1981; Visher, 1987; Moran, Cutler, & De Lisa, 1994), with the exception of measures of authoritarianism which were useful for identifying jurors who would be prone to vote for harsher punishments (Mitchell & Byrne, 1973). However, for the most part authoritarianism does not predict verdict decisions, and indeed even its ability to predict punitive behaviour has seen mixed results (Kassin & Wrightsman, 1988; Narby, Cutler, & Moran, 1993).

The relationship between authoritarianism and a preference for punitive reactions to criminal cases was the bases for the earliest attitude scale measuring juror bias specifically. The Legal Attitudes Questionnaire (LAQ) was developed by Boehm (1968) in order to investigate whether “an individual’s liberalism-conservatism attitudes have a systematic effect on the way he behaves as a member of the jury” (p. 737). The LAQ conceptualised three attitude dimensions which were thought to contribute to the successful prediction of juror behaviour; Authoritarianism (A), Anti-authoritarianism (AA), and Equalitarianism (E). The authoritarianism dimension was associated with conservatism and rigidity, which could be a bias with a detrimental impact on the defence due to contributing to the confirmation bias, particularly as the prosecution presents its case first at trial. The anti-authoritarianism dimension was essentially the polar opposite of the authoritarian personality, which would result in a directly opposing bias favouring leniency towards a defendant. Finally, the
equalitarianism personality was considered to be the compromise between the A and AA categories, which was described as having ‘no bias’.

The original LAQ consisted of ten clusters of statements representing common beliefs or attitudes toward the Criminal Justice System. Each group contained three statements, each reflecting one of the personality constructs measured by the scale (A, AA, and E) (Boehm, 1968). Participants indicated which of the three statements they agreed with most, and which they agreed with least, for each of the ten clusters of statements. The original validation study conducted by Boehm (1968) presented two versions of a murder case to 151 undergraduate students. In one version the evidence presented indicated the defendant was innocent, and in the other version the evidence indicated guilt. Participants were asked to render a verdict of not-guilty, guilty of manslaughter, or guilty of second-degree murder, as well as completing the LAQ. The results of this study indicated that when the innocent version of the murder case was presented, participants who convicted the defendant displayed higher authoritarianism on the LAQ than those who voted not-guilty. In the guilty evidence condition participants who scored higher on anti-authoritarianism on the LAQ were more likely to not vote for second-degree murder, and instead showed a preference for manslaughter.

A significant difficulty with interpreting the results of the LAQ resulted from the unconventional scoring method due to the ranking of agreement to clusters of three statements by participants. In an effort to improve the psychometric properties of the original LAQ, Moran and Comfort (1982) reduced the LAQ to 24 items and changed the scoring in favour of responses on Likert scale levels of agreement. This revised version
of the LAQ resulted in a total legal authoritarianism score and was completed by 319 jurors from real felony trials. The results of this study suggested that high legal authoritarianism scores significantly predicted a guilty verdict decision, but only for female jurors. As this study used real jurors and their actual verdict decisions across a wide range of felony trials, it was not possible to control for trial-specific variables such as the strength of evidence, which would presumably have contributed to the jurors’ verdict decisions. It is also important to note that the participants in this study would have also participated in deliberation as part of their role as a real juror, and it is not clear what impact this may have had on their verdict preference, and therefore the results.

Kravtiz, Cutler, and Brock (1993) further evaluated the original and revised versions of the LAQ with both students and jury-eligible members of the public. In this research the revised version of the LAQ consisted of the original 30 items from the LAQ, scored with Likert scale responses as in the Moran and Comfort (1982) study. As well as testing the reliability and validity of the scales, this study also compared participants’ scores with other measures known to be related to legal decision making such as measures of social desirability, belief in a just world, and the balanced F-scale. The authors concluded that the complicated scoring of the original LAQ resulted in many participants making erroneous responses, ultimately leading to a measure with questionable psychometric merit. Their results suggested that the revised LAQ produced more reliable and valid results that the original LAQ, but also that a further revision to the scale resulted in a 23-item version which demonstrated even higher internal validity. The recommendations made by the authors supported the use of this
shorter version of the LAQ for use in jury selection procedures. These results suggest an alternate form of the LAQ might perform better in terms of reliability and validity, however this study only required participants to complete the various measures and no evaluation of a case was required. It is therefore not clear from this study which version of the LAQ would have better predictive ability in relation to juror verdict preferences.

In another effort to establish a more generalised measure of pre-trial juror bias, Kassin and Wrightsman (1983) developed the Juror Bias Scale (JBS), which is still in use in the American legal system and has been evaluated more recently in the research literature (Lecci & Myers, 2002; De La Fuente, De La Fuente, & Garcia, 2003). The theoretical basis for the JBS is the assumption that jurors reach a verdict by comparing the implicit outcome of two decisions, as described by the story model (Pennington & Hastie, 1986; 1988). The first decision is the probability that the defendant committed the crime in question, and the second is the threshold of reasonable doubt. These two separate decisions are then compared with one another, and if the probability that the defendant is guilty is greater than the threshold of reasonable doubt then the juror votes guilty. The rationale for the JBS was to identify whether people differ in their pre-trial perception of these two constructs, and if so a measure of these attitudes could potentially predict verdict preference.

The JBS consists of 17 items, designed to measure the dimensions of probability of commission (PC) and of reasonable doubt (RD), each of which were worded to reflect either a pro-prosecution (nine items) or pro-defence bias (eight items). Scores on the JBS can range from 17-85 (after scoring pro-defence items in
reverse), where a higher score indicates a generalised pro-prosecution bias. Five filler items were also included in the scale to make the purpose of the measure less obvious to participants. The initial validation studies carried out by Kassin and Wrightsman (1983) found that the JBS was uncorrelated with social desirability scores ($r = -.01$), moderately correlated with belief in a just world scores ($r = .24$, $p < .01$) and moderately correlated with authoritarianism ($r = .43$, $p < .001$). In order to test the scale’s ability to predict verdicts a series of studies were carried out using both student and real juror samples as well as a range of crime types represented in the fictional trial stimuli. The findings indicated that participants who scored high on the JBS (defined using a median split) were significantly more likely to vote guilty. When pro-prosecution and pro-defence biased participants were compared in terms of their probability of guilt estimates and threshold of reasonable doubt, less consistent results emerged. In their first study the two groups differed in their reasonable doubt criteria and not in the probability of guilt estimates. However in their second study the opposite relationships emerged.

The predictive validity of the JBS was reported to be significant in all but one of the validation studies by Kassin and Wrightsman (1983). One crime type used in that study was a rape case, and in this condition there were no significant differences in verdict between pro-prosecution and pro-defence participants. In fact, the results of the study suggested the opposite trend (although not statistically significant) of pro-defence participants being slightly more likely to vote guilty. The authors were not able to provide a definitive reason for this anomaly, however they suggested that this finding could reflect a separate rape victim sympathy bias which may have confounded
their results. In order to test this explanation a follow up study was reported in which the JBS and the Rape Empathy Scale (RES) (Deitz, Blackwell, Daley, & Bentley, 1982) were administered to participants. The results of a correlation analysis indicated that pro-prosecution biased jurors were less sympathetic toward rape victims, however this was only statistically significant for male participants ($r = -0.44, p < .05$).

Despite some inconsistent and still yet unexplained results, the overall conclusion was that the JBS is a reliable predictor of verdict preference in jurors, and it is still the most commonly referenced measure of pre-trial juror bias in the literature. The original JBS was not developed using a factor analytic approach, and Myers and Lecci (1998) proposed that identifying the factor structure of the JBS would help to further understand the underlying constructs of the scale and explain its predictive ability. The results of their factor analysis of the JBS did not support the existence of two distinct theoretical constructs as described in the original Kassin and Wrightsman paper (1983). In order to further investigate the possible factor structure of the JBS, exploratory factor analysis was conducted on the PC and RD subscales separately. The RD subscale was found to have a three factor solution, with two items removed due to the fact that they did not load onto any of these three factors. The PC subscale yielded a two factor solution with three items not loading on the factors, which were therefore removed. The authors then used this revised version of the JBS, consisting of a six-item, two-factor PC scale and a six-item, three-factor RD scale in a subsequent validation study. The results of this follow up study confirmed the findings of the factor analysis.
In order to test this revised structure of the JBS in relation to predicting juror verdict, Myers and Lecci (1998) conducted a study with 406 undergraduate students in which participants completed the original and revised versions of the JBS and then considered a trial scenario involving a homicide. Although the authors indicated that their sample consisted of jury-eligible students, the age range included participants aged 17 years, which would in fact not make some of the participants eligible to be selected as members of a jury. The results suggest that the predictive ability of the revised 12-item JBS was slightly better than the original 17-item scale, and in particular the new version of the JBS improved the reliability of the PC scale which was found to be inconsistent at predicting verdicts in earlier research.

In a later study by Lecci and Myers (2002) the revised and original JBS scales were again tested, this time using a sample of jury-eligible members of the public and using three different trial scenarios (murder, rape and armed robbery). The participants in this study had a similar inconsistency (as the previously described student sample) as the age range included participants up to age 84 years which would make them ineligible to be selected for jury duty. The findings of this study replicated Myers and Lecci (1998) and found that the revised JBS had slightly better predictive ability than the original JBS, which lends further support for the reduced number of items. Similar to the findings in the original Kassin and Wrightsman (1983) study, the rape trial showed the weakest association between verdict preference and JBS scores.

The research described above focused exclusively on the prediction of verdict preference in jurors before group deliberation procedures. However, what is not considered in these studies is how juror bias operates within the deliberation process.
and the impact on post-deliberation verdict decisions. A study by De La Fuente, De La Fuente, and Garcia (2003) investigated the predictive ability of the original JBS adapted and translated for use with a Spanish population. They presented 153 undergraduate students with a homicide trial re-enactment (with either pro-prosecution or ambiguous evidence) and a questionnaire which included the JBS and items relating to decisions about the trial verdict. In a previous session participants had completed the JBS and were labelled as either pro-prosecution biased or pro-defence biased (based on a median split). Participants were randomly assigned to hear one version of the murder trial and deliberate in groups of nine participants which were homogeneous in terms of the JBS classification of the members (e.g. the juries consisted of entirely pro-prosecution or pro-defence members).

The results indicated that the JBS significantly predicted verdict preference, but only when the evidence presented in the trial scenario was ambiguous. It was also concluded that the deliberation process actually enhanced the effect of jury bias, as the prevailing bias in each jury group resulted in an increase in guilty verdicts (for the pro-prosecution juries) and a decrease in guilty verdicts (for the pro-defence juries) after the deliberation compared with the pre-deliberation decisions by each juror. The authors point out that this is of particular concern if the likelihood of homogeneous juries (with regard to various biased tendencies) is increased due to the use of voir dire hearings (Sommers & Norton, 2008). However, it is not clear from this research how heterogeneous juries (with respect to JBS scores) would deliberate, and what impact this would have on verdict decisions and the presence of bias in the deliberation room.
In a much more recent attempt at developing a measure of juror bias, Lecci and Myers (2008) developed the Pretrial Juror Attitude Questionnaire (PJAQ). The design of this attitude measure was based on the view that general attitudes are not likely to predict specific decision making, and therefore juror bias measures should be designed to measure attitudes which are directly relevant to legal decision making (Ajzen & Fishbein, 1977; Skeem, Louden, & Evans, 2004). The PJAQ consists of 30 items and was administered along with the JBS and the revised LAQ (23 item version described earlier) to 601 members of the public. The results of an exploratory factor analysis (conducted on the 30-item PJAQ and 17-item JBS) identified six factors, referred to as; conviction proneness, system confidence, cynicism towards the defence, social justice, racial bias, and innate criminality. After item analyses, 29 items were retained to form the PJAQ, which included 12 items which appeared in the original JBS.

Lecci and Myers (2009) tested the predictive ability of the PJAQ using trial scenarios for murder, rape, and robbery and found that five of the six factors were significantly associated with verdict preference. This was true for all crime types except the rape trial in which none of the PJAQ subscales were able to predict verdict preference. The authors concluded that the PJAQ performed better than the JBS and LAQ-revised scales at predicting mock juror verdicts, and highlighted that this is likely due to the fact that there were additional, legally relevant constructs included in this measure which were not accounted for previously. It was also concluded that “it is expected that indirect effects, particularly as mediated by the evaluation of evidence, would carry the greatest influence of pretrial bias” (p. 2029), which is an issue addressed by the studies conducted in this thesis.
There are some methodological issues which should be considered alongside the findings presented in the Lecci and Myers (2009) studies. Firstly, in order to generate an initial list of items to include in the PJAQ, 42 students were asked to generate items that they felt would indicate a bias that was likely to affect jurors’ verdicts. This resulted in 95 unique items, which were then judged by 110 students who rated the degree to which each item was an indicator of juror bias. It is not clear why the authors relied solely on the judgments of students for the generation of their scale items, and it could be argued that generating items based on the literature would have resulted in more robust scale items. In addition, by only including verdict as the dependent variable it was not possible to reach conclusions about participants’ perceptions of the probability of guilt of the defendant. Finally, there is no explanation for the predictive ability of some of the PJAQ subscales over others, which raises questions about the validity of the scale. For instance, the racial bias subscale predicted verdict preference even though there was no racial information in the trial scenario. Perhaps racial bias is operating as a proxy measure for some other relevant bias, which is why it demonstrated predictive ability in these scenarios.

Attempts to devise measures of individual differences which possess the ability to reliably predict juror verdict preference have demonstrated varying degrees of success. While general personality characteristics and demographic variables have not shown very consistent results (with perhaps the exception of authoritarianism) other more legally specific individual attitude measures have been able to predict verdict preference under some circumstances. One aspect of the research methodology which may partially explain the somewhat inconsistent results is the use of a strictly
dichotomous outcome decision (e.g. guilty/not-guilty verdict) in many of the studies. 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 (De La Fuente et al., 2003).

Since the LAQ and JBS were both developed and initially tested before the use of forensic science evidence in the courtroom, which is now commonplace, there is no reference to the evaluation of physical evidence in any of the early studies. Indeed, even the more recent validation studies described above, and the newer attitude measures (e.g. PJAQ) have used exclusively non-physical evidence in the trial stimuli (e.g. witness testimony, circumstantial evidence, etc.). This could be due to the perceived objectivity of physical evidence previously discussed, however the argument put forth in this thesis is that specific attitudes may exist which potentially bias jurors’ interpretation of forensic evidence. By developing a specific attitude measure it may be possible to identify the attitudes which indirectly lead to a particular verdict preference in trials with ambiguous or weak physical evidence.

1.3 Attitudes Toward Forensic Science and Legal Decision Making

A recent development in the research literature concerning jurors’ understanding and evaluation of forensic evidence is in response to anecdotal claims made by attorneys and judges regarding the impact of fictional representations of forensic science on jurors’ beliefs about the evidence presented in court. The existence of such an effect has been written about quite widely in the media, and has often been referred to as the *CSI Effect* (Tyler, 2006). The belief that popular fiction
can have an impact on jurors during real life trials is not a new one. Prior to the popularity of the forensic science-based programmes, other law-related television dramas were thought to be having profound effects on jurors’ expectations of the criminal justice system (Graham, as cited in Mann, 2006). An example was the so-called *Perry Mason Effect*, which was thought to result in jurors expecting that attorneys would manage to secure surprise confessions from suspects during cross examination (Brickell, 2008). Although these surprise confessions were depicted frequently on television, this rarely happens in actual courtrooms.

The *CSI Effect* refers directly to the enormously popular television drama of the same name (CSI – Crime Scene Investigation). The show, which first aired in 2000, is reported by Nielsen Ratings Research to be the number one watched television programme in the world for viewers over the age of 50 years, and the number five programme for viewers aged 18-49 years. In addition, its spinoff programmes (CSI Miami and CSI New York) also feature in the top 20 programmes for both age groups. CSI has also been awarded the ‘most watched drama series in the world’ title in 2007, 2008, and 2010 with an average of 73.8 million viewers each new episode (Nielsen Media Research, 2010). It is clear that the programme reaches a wide audience which is one of the reasons for reference to this particular programme to describe the potential effect of the popular media on jury behaviour.

In addition to the fact that CSI is watched by a massive audience each week, another important issue is how realistic the science portrayed on the show actually is. Houck (2006), a forensic consultant for the producers of CSI, estimates that 40 percent of the science portrayed in episodes of CSI does not exist, and Schweitzer and Saks
(2007) reported that even the fact-based scientific techniques appearing on the show are conducted in ways which are exaggerated and unrealistic. In particular, investigative timeframes are collapsed (Mann, 2006), forensic evidence always reaches a conclusive and unbiased conclusion (Stephens, 2006; Schweitzer & Saks, 2007), and court proceedings are not shown, resulting in trials being depicted as merely a formality (Nolan, 2006; Tyler, 2006). So it does seem to be the case that large numbers of people, in various countries, are tuning in to watch inaccurate portrayals of forensic science on fictional shows such as CSI. Just how this information might be affecting potential jurors in actual trials has only recently become the focus of empirical research.

Initially, two possible main effects were described in the literature to explain how the popularity of forensic fiction might influence jurors, both of which can have an impact on how jurors perceive and evaluate evidence (Schweitzer & Saks, 2007). The first effect is said to result in an increased burden on the prosecution during a trial, as the jurors have unrealistically high expectations for the presence of forensic evidence, which often cannot be met by actual methods of investigation and analysis (Marquis, 2007). A second effect which has been proposed is one in which the defence team is disadvantaged by the fact that jurors display an exaggerated belief in the ability of forensic science analysis to reliably identify the perpetrator of a crime (Tyler, 2006). This juror over-reliance on forensic evidence makes it much more difficult for the defence to argue against the evidence, which may be crucial in cases where the evidence is ambiguous, or of questionable probative value (DiFonzo & Stern, 2006). These two potential outcomes of the CSI effect have both been said to
be operating in court rooms, according to some judges and lawyers. However, until recently there were no empirical studies designed to measure any actual effects in potential or mock jurors.

A more comprehensive consideration of the possible effects of television programmes on juror behaviour was described by Cole and Dioso-Villa (2007). This paper proposed an additional four potential effects, although they argue that the most problematic impacts are the effects described above in Schweitzer and Saks (2007). The additional hypothesised effects include changes in prosecutor behaviour, in which a lack of forensic evidence is thought to require an explanation to avoid disappointment in juries. It is also proposed that jurors are actually becoming more educated about forensic science through the media, which might make them better at assessing expert testimony. A similar effect is evident in the dramatic rise in student numbers on forensic-related courses in higher education, as well as the possibility that offenders are becoming more forensically aware as a result of the popular media.

In terms of a theoretical basis for the CSI effect, Podlas (2006) describes the possible impact with reference to the story model of juror decision making. As described earlier, the story model posits that jurors do not arrive at a trial as a ‘blank slate’ in terms of their expectations of the process and the evidence that they will be asked to evaluate (Pennington & Hastie, 1986). Instead, jurors use their pre-existing beliefs about the law, crime, and evidence to organise the information presented to them as a narrative describing the events in the case (Pennington & Hastie, 1993). In many cases members of the public have had no previous experience of the criminal justice system therefore their only cognitive reference for a trial situation may be
based on information gathered from media accounts (Elffers, de Keijser, van Koppen, & van Haeringen, 2007; Daftary-Kapur, Dumas, & Penrod, 2010). If these pre-existing beliefs are shaped by the inaccurate portrayal of forensic science and the criminal justice system in the popular media, this could affect jurors’ abilities to accurately evaluate the evidence.

Although it has not been proposed in the existing CSI Effect literature, the theory of anchored narratives (Wagenaar et al., 1993) may also provide some theoretical support for the importance of attitudes about forensic science on legal decision making. Beliefs about the world held by the trier of fact form the basis of anchors for pieces of evidence, however if these beliefs are false then the reliance on certain pieces of evidence (e.g. forensic science) may be incorrect and have the potential to result in faulty verdict decisions. For instance, if weak forensic evidence (or a lack of forensic evidence) is considered to be compelling evidence by a juror based on beliefs about forensic science which are based on fictional representations, then this heuristic can result in reasoning errors and poorly supported verdicts.

Prior to 2006, reports of the CSI effect were only supported by anecdotal evidence, mainly from lawyers and judges who claimed to have experienced changes in juror behaviour in the courtroom. In an effort to quantify some of these opinions, a few studies have conducted surveys and interviews of members of the judiciary in order to determine to what extent they may have experienced the CSI effect. Robbers (2008) surveyed 290 American lawyers (prosecution and defence) and judges about their experiences and opinions and found that 79% cited specific instances in which they felt they had experienced juries who were influenced by forensic fiction on
television. In addition, 85% of participants reported that the way they conduct their job had changed in recent years, in order to counter jurors’ unrealistic perceptions of forensic science. Some of these changes included spending more time addressing these issues in \textit{voir dire} hearings (36%), use of negative expert testimony to explain missing evidence (23%) (also reported in Stevens, 2008), as well as specifically explaining to jurors the difference between television and real-life forensics (31%). Marquis (2007) reported similar opinions, and highlighted the significant changes to prosecutor behaviour, pointing out that these changes have been necessary because defence attorneys are exploiting the CSI effect to their advantage (e.g. arguing that a lack of forensic evidence requires an acquittal).

In the last five years, a small number of empirical studies have been conducted in order to investigate whether or not a \textit{CSI effect}, as described above, exists. The first of these studies was conducted by Podlas (2006) and focused on determining whether there was an anti-prosecution bias held by jurors who reported viewing CSI. The participants in this study were classified as frequent viewers of CSI (FV) or non-frequent viewers (NFV) based on self reported viewing habits, and the trial summary used was that of a rape case in which the only issue was consent, therefore forensic evidence was not presented, or legally useful in determining the verdict. The aim was to analyse the reasons given for verdicts of not-guilty, in order to determine if CSI viewers were more likely to attribute a decision to acquit to the absence of forensic evidence, even though this type of evidence would not have been useful to the case. The results of this study found that there were no significant difference between the reasons given for not-guilty verdicts between FV and NFV of CSI. The author
concluded that the findings did not support the claim that the *CSI effect* operates in such a way that jurors acquit due to substantially different reasons depending on their viewing habits (e.g. that jurors who view CSI are more likely to acquit on the basis of there being no forensic evidence presented).

There are some methodological issues with the Podlas study which may have affected the results. First, the viewing habits investigated in this study focused solely on one television programme (e.g. CSI), and ignored any other popular sources of forensic fiction that participants may have been exposed to (e.g. other television programmes, crime fiction novels, movies, etc.). Potentially more problematic is the fact that the researchers chose to use a case summary which did not rely on any forensic evidence, and therefore did not resemble a typical CSI-type scenario. If the mock jurors in the study were under the influence of the *CSI effect*, it would have more likely manifested itself in a scenario where they were being asked to evaluate forensic evidence (or lack thereof) which is relevant to a case, and which is likely to activate any pre-existing schemas which have resulted from viewing typical CSI episodes (Cavender & Deutsch, 2007; Winter & York, 2007).

A different approach was taken by Shelton, Kim, and Barak (2006) in which mock jurors were surveyed about their expectations of forensic evidence before being presented with a trial transcript, as well as their interpretations of the evidence contained in the case summary. Importantly, the participants in this study were drawn from a list of people who had been summoned for jury duty (*N* = 1027), prior to their attendance at a trial. In the first part of the study participants were asked to report their viewing habits relating to “law related television programs” (p. 340). This was
followed by a second stage which asked participants about their expectations of what forensic evidence they would expect to be presented within a variety of types of criminal trial. The final stage of the study presented participants with a series of scenarios, representing a range of criminal offences, and asked them to respond on a five point scale indicating their opinion of the likelihood of the defendant’s guilt.

The results of the Shelton et al. (2006) study indicated that the most frequently viewed law-related television program was CSI, with less than half (44%) of participants indicating that they never watch it. Correlation analysis also indicated that the more frequently participants reported watching a particular program, the more accurate they believed the portrayal of forensic information in the program to be. Participants also indicated a rather high expectation of the recovery of forensic evidence in ‘every criminal case’ (46.3%). In an analysis of the expectations of participants by crime type (e.g. murder, burglary, rape, etc.) CSI viewers had higher expectations for forensic evidence of all types (DNA, fingerprint, ballistics) across all different crimes compared with non-viewers. However, when participants were asked to indicate a verdict in each of the crime scenarios, this high expectation of forensic evidence did not result in CSI viewers requiring this type of evidence as necessary for returning a guilty verdict in every case type.

The Shelton, et al. (2006) study approached the potential for an effect of popular media on jurors from a wider angle than the previous study by Podlas (2006). Rather than defining the independent variable of ‘viewing habits’ as the frequency of watching specifically CSI and its spin-off programs, the Shelton et al. study widened the definition to include a number of popular law-related television programs. This is
likely to be a more realistic approach to measuring any potential effect that the media might be having on jurors, due to the fact that although CSI may be the most popular program it is not the only possible source of pre-trial information available to the public. The conclusion reached by Shelton et al. was that any elevated expectations of jurors demonstrated in this study are more likely the result of public interest in recent technological advances in forensic techniques rather than a result of jurors watching one particular forensic-related program. This more general “tech effect” (p. 362) is to be expected, and Shelton et al. suggest that the criminal justice system must be prepared to attempt wherever possible to meet jurors’ expectations during investigations and trials.

A subsequent paper by Kim, Barak and Shelton (2009) used the same data (N = 1027, from Shelton et al., 2006) to analyse the indirect relationship between viewing habits and verdict preference using path analysis. The authors reported that although no direct effect between exposure to CSI-type dramas and willingness to convict were found, the exposure to crime fiction did have an indirect effect by increasing participants’ expectations of forensic science. These raised expectations associated with viewing CSI-type shows were found to significantly lower the willingness to convict in the absence of forensic evidence. Unfortunately trial material used in this study did not contain any detail of the case being considered, but rather only presented participants with isolated statements about the cases. This makes it difficult to determine whether these findings would be replicated using a more realistic trial stimulus. However, the strength of this study was in the use of path
analysis as a way of determining the indirect effect of television viewing, and a similar approach is taken in the studies conducted within this thesis.

In another study investigating the effect of television viewing habits on the interpretation of forensic evidence, Schweitzer and Saks (2007) sampled 48 jury-eligible college students who were asked to read a transcript of a simulated trial involving one source of identification evidence. The forensic evidence presented in the scenario was in the form of expert testimony by a hair and fibre analyst who testified that it was his professional opinion that a hair recovered from the perpetrator’s ski mask originated from the defendant. Participants were asked to evaluate the usefulness of the trial evidence through a series of questions and they were divided into two categories based on their viewing habits for both “forensic related” and “general crime-themed” programs (p. 362).

The results of the Schweitzer and Saks (2007) study indicated that viewers of both forensic-related and general crime related programs rated themselves as having a higher level of understanding of the forensic technique presented than did non-viewers. Subsequently, forensic-related viewers also indicated significantly more confidence in their verdicts than did non-viewers. However, the difference between the groups in terms of their verdicts was not significant. One significant difference between forensic-related viewers and both non-viewers and general crime-related viewers, was that the forensic-related viewers were more critical of the forensic technique presented in the case. The authors therefore concluded that these findings indicated a potential bias in jurors who are viewers of forensic-related programs (such as CSI) to be overly critical of forensic evidence – a manifestation of the CSI Effect.
which they argue burdens the prosecution due to potentially unrealistic expectations to produce increasing amounts and quality of evidence.

Unfortunately Schweitzer and Saks (2007) did not consider some of their findings in relation to the potential impact on jury deliberations, and an area for future research might include determining whether the increased confidence and self-reported understanding of forensic science might influence the deliberation process. It is possible that these types of jurors may have a more dominant influence on deliberations, and indeed may even be perceived by other jurors to be more knowledgeable than they actually are (Devine, et al., 2001).

In an effort to avoid using television viewing habits as the independent variable of interest, Lieberman, Carrell, Miethe, and Krauss (2008) reported the results of three studies which extended the previously used methodology in the CSI effect literature to include more explicit attitude measures, investigating how mock jurors’ attitudes were related to their perceptions of a criminal case. In the first study they surveyed participants (383 students and 233 jurors) about their perceptions of the accuracy and persuasiveness of different types of forensic evidence (e.g. DNA, fingerprints, hair, fiber, etc.). It was reported that DNA evidence was perceived to be most accurate and persuasive, and they found no relationship between television viewing habits and perceptions of evidence. Participants also read either a rape or murder scenario in which evidence type was manipulated, and was either incriminating or exculpatory in nature. For both crime types, incriminating DNA evidence lead to 100% conviction rates, which was higher than any other evidence type.
This does suggest that DNA evidence is particularly convincing to jurors, however there were some discrepancies in juror decisions between crime types which were not addressed by the authors of this study. In particular, the authors asserted that there was no reason to suggest jurors would perceive the crime types differently, despite the fact that previous research has found rape trials scenarios to differ significantly from non-sexual offences, due to different biases operating in these cases (Kassin & Wrightsman, 1983; Lecci & Myers, 2008). In fact the participants in this study appear to have perceived the evidence differently depending on the crime type. In particular, incriminating fingerprint evidence resulted in 100% conviction in the rape scenario, but only 67% conviction in the murder trial scenario. The same was true for victim testimony evidence, although it is not clear how there was victim testimony included in the murder trial scenario (as the victim is deceased and could therefore not have provided testimony in the murder trial).

The second study reported by Lieberman et al. (2008) presented student participants \( N = 114 \) with a sexual assault trial which included multiple types of forensic evidence (semen, blood, hair, fingerprints). In each condition participants were told that only one of the sources of physical evidence was incriminating, while the remaining types of evidence yielded inconclusive results. By including both incriminating and inconclusive pieces of physical evidence within the same trial scenario, this study was able to determine which forensic evidence types led to conviction of the defendant, despite other physical evidence to the contrary (McAllister & Bregman, 1986). The results confirmed that there were significantly more convictions and higher probability of guilt estimates when the incriminating
evidence type was a DNA or blood type match, even when accompanied by damaging cross-examination within the scenario.

The third study in the Lieberman et al. (2008) paper was the most innovative, as this was the first paper to approach the CSI effect by attempting to measure participants’ attitudes about forensic science, and using these attitudes to predict their verdict preference. The participants in this study were students ($N = 130$), and they completed a pre-trial DNA beliefs assessment before reading a murder trial scenario. This pre-trial assessment included items from Kassin and Wrightsman’s (1983) Juror Bias Scale, and an unpublished master’s thesis scale (Martin & Cohn, 2002, as cited in Lieberman et al. 2008) as well as four additional questions specifically about DNA. Participants indicated their level of agreement with the items on a 10-point Likert scale, and these scores were analysed using logistic regression as predictors of verdict. The scores on the four DNA beliefs items were found to significantly predict verdict preference and probability of guilt estimates.

The main contribution of the Lieberman et al. (2008) findings to the previous literature is in the measurement of attitudes which successfully predicted mock juror preference, as the previous research had been focusing on television viewing habits as the independent variable of interest. This reliance on measures of viewing habits is likely one of the reasons that previous research had not identified any proof of the CSI effect existing, despite the seemingly overwhelming anecdotal support in earlier papers (Trask, 2007). The use of an attitude measure highlights the importance of relevant beliefs held by participants, and reduces the need to identify the specific source of these attitudes in the popular media. The studies conducted in this thesis
also use this methodological approach for investigating more general attitudes about forensic science (as opposed to attitudes about specific forensic techniques).

The findings described in Lieberman et al. (2008) provide some support for the impact of pre-trial attitudes about DNA evidence on mock juror verdicts, though there are some other aspects of their studies which may have influenced the results. As previously mentioned, these authors’ assumption that rape and murder scenarios would be perceived similarly by participants does not take into account the previous literature which suggests this is not true. Unfortunately this assumption resulted in a lack of contrasts between crime types for the different evidence types in this paper. In addition, the evidence presented in the trial scenarios was always of high probative value, whereas varying the strength of the evidence would have allowed interesting comparisons to be made. Interactionist models of juror decision making would predict that the impact of participants’ prior beliefs would have been more pronounced if the evidence had been less probative (Kassin & Wrightsman, 1988), however the design of this study did not allow for this to be tested.

Due to very limited empirical research to date, and limitations accessing real juries during trials, it is not entirely clear whether any version of the CSI Effect actually exists and is in fact influencing jurors’ decision making processes. The studies discussed above had varying approaches to addressing this research question, and therefore different goals and findings. One common finding in these studies, however, is that there was no significant difference between jurors who were classified as CSI-viewers and those who were non-viewers in the verdicts that they reached at the end of their mock juror task. This finding was consistent despite viewers and non-viewers
being defined differently (referring to different television programs and various viewing frequencies), and varying trial scenarios across the studies. This finding seems to suggest that although jurors may be influenced by inaccurate media representations of forensic science techniques, this influence is not having a direct effect leading them to reach different verdict decisions from those jurors who are not exposed to the same media influence. However more recent studies (Kim et al., 2009) suggest that treating the CSI effect as an indirect influence on juror verdict preference shows some promise, and is therefore the approach taken in this thesis.

Although there has been no definitive proof of a CSI effect in the literature, some authors have written about the general plausibility of such an effect existing and having an impact on the legal system. Tyler (2006) suggests that there are two issues which need to be considered when discussing the possibility of jurors being influenced by the viewing of forensic science related media. The first is the research literature concerned with the extent to which jurors are influenced by the mass media in general. The past few decades have seen substantial research in this area, and much of this literature has concluded that pre-trial publicity (PTP) can have a prejudicial effect on jurors’ decision making processes (Hope, Memon, & McGeorge, 2004). If it is true that general PTP from a number of media sources can influence the way that jurors perceive and understand trial evidence, then Tyler (2006) concludes that it is plausible that exposure to television programs such as CSI might have an impact on the way jurors interpret evidence in court. One finding in the research which may be especially relevant to any CSI effect is that an effect of PTP on juror decision making may be particularly likely when the PTP contains information concerning physical
evidence presented during the trial (Shaw & Skolnick, 2004) when compared to PTP relating to witness evidence.

The second issue identified by Tyler (2006) as being important in determining the plausibility of the CSI Effect is whether jurors are able to consciously ignore their preconceptions during trial deliberation. This question is relevant because PTP is only problematic if jurors are not able to ignore this information when making their verdict decision. Much of the psychological literature on this matter has focused on the ability of jurors to ignore evidence that has been presented to them, following instructions from a judge that the evidence is inadmissible. The majority of findings in this research area have concluded that jurors are not able to ignore information, even when instructed to do so by a judge (Casper & Benedict, 1993; Kaplan, 1982). It is therefore unlikely that jurors are able to put aside their prior beliefs or pre-trial information when interpreting trial evidence, particularly if that evidence is ambiguous.

By considering the possibility of a CSI Effect in the context of research relating to the more general area of pre-trial publicity, Tyler (2006) concludes that it is likely that television programs such as CSI are capable of having some effect on juror decision making. What exactly the effect is remains unclear, and requires further empirical investigation. Whether jurors are likely to be biased for or against the prosecution in a case has not been demonstrated clearly by the existing research. However, some authors believe that either type of bias is possible due to exposure to fictional representations of forensic science (Huey, 2010; Tyler, 2006; Mann, 2006; Shelton et al., 2006). It seems reasonable to move forward with research which does
not focus too narrowly on the viewing of particular television dramas in search of biases, but rather treat the *CSI Effect* as a wider pre-trial publicity bias and shifting focus to the attitudes created by the media. This approach to future studies may uncover more about this particular source of bias, when it is likely to operate, and the effects that it may be having on juror decision making.

1.4 **OVERCOMING JUROR BIAS IN THE COURTROOM**

In the previous sections it has been suggested that juror bias should be considered the norm in the courtroom, rather than the exception (Myers & Arena, 2001), which may be in direct conflict with the fundamental requirement for a defendant to have their case decided by an impartial jury. Kassin and Wrightsman (1988) highlight the assumptions underlying the concept of an impartial jury, which require that either jurors arrive at court as a ‘blank slate’, that jurors can effectively put aside pre-trial information or biases, or that biased prospective jurors will be identified and removed from the jury venire. It is clear from the research described in the previous sections of this review that jurors do not arrive at court without previous beliefs, knowledge, and attitudes which affect their perceptions of criminal cases (Pennington & Hastie, 1986), and it has also been suggested that these pre-trial biases are difficult, if not impossible, to ignore (Crocker & Bull Kovera, 2010). It has therefore been suggested that the most effective way to ensure jury impartiality is to exclude potential jurors who are found to have biases which would affect their interpretation of the evidence and the case (Brooks, 2004). The use of pre-trial hearings for the purpose of identifying bias in potential jurors, and the effectiveness of these hearings, is therefore the subject of this section.
In countries with adversarial legal systems such as England and Wales, Australia, Canada, and the United States, the use of pre-trial hearings (or *voir dire* hearings) for identifying potential juror bias is relatively commonplace although how these hearings are conducted varies greatly between jurisdictions (Brooks, 2004; Kaplin & Martin, 2006). There are two main processes which provide the means for excluding jurors presumed to have biases relevant to a particular case; challenges for cause, and the peremptory challenge (Norton, Sommers, & Brauner, 2007).

Challenges for cause require the judge to be convinced, by the prosecution or defence, that a particular juror holds some bias or previous knowledge which would disable an impartial judgement. Typically the court allows an unlimited number of challenges for cause, however there needs to be a justifiable reason for excluding a juror on these grounds (e.g. they have a relationship with one of the parties, etc.) (Hastie, 1991).

Alternatively, peremptory challenges are limited in number, but allow both the prosecution and defence to exclude potential jurors without having to provide a reason to the court (Van Dyke, 1977; Crocker & Bull Kovera, 2010). Due to the subjective and unexplained nature of the peremptory challenge, and therefore the focus on this type of challenge in the research literature, it will be the focus of the remainder of this section.

The use of peremptory challenges for excluding jurors has historically been a contentious issue, and the use of this procedure differs widely between countries (Kerr, Kramer, Carroll, & Alfini, 1991). Although the concept of the peremptory challenge originated in England, they are rarely used in courts in the UK at present (Wheeler, 2006) although the right for the prosecution or defence to challenge jurors
still exists (Juries Act, 1974 s. 12(1)(b)). Australian and Canadian criminal courts both also allow for the use of peremptory challenges however, the United States exercises the most stringent jury vetting procedures through pre-trial hearings (Brooks, 2004).

The exclusion of jurors on the basis of pre-trial hearings and peremptory challenges are thought, in theory, to result in an impartial jury, however this not the only function served by these challenges (Gobert, 1997). Proponents of these challenges argue that they also improve the appearance of the fairness of the court process by allowing some control over the jury selection process, and allow for the elimination of incompetent jurors which is thought to result in more effective deliberation processes and reduces the likelihood of a hung jury (Hastie, 1991; Rose, 1999). However some argue that the use of challenges undermines the purpose of the jury system and is contrary to the concept of a representative jury (Rose, 1999). In addition there have been concerns that the use of pre-trial hearings may give jurors unfavourable views of the trial attorneys and the evidence, and that the motivation for lawyers to select jurors who are likely to favour their side of the trial renders hearings inherently impartial themselves (Gobert, 1997; Greene, Heilbrun, Fortune, & Nietzel, 2007). For the purposes of this thesis, the most important argument against the use of peremptory challenges refers to the court’s ability to correctly and reliably identify bias in potential jurors, as juror exclusions made on the grounds of faulty assumptions of bias ought to be avoided (Fulero & Penrod, 1990).

There has been a relatively small body of research investigating the tactics used by judges and lawyers in voir dire hearings, as well as the effectiveness of these tactics. The assumption held by some proponents of jury selection is that lawyers are well
equipped to detect bias in potential jurors, as articulated by Bigam (as cited in Olczak, Kaplan, & Penrod, 1991)

“Trial attorneys are acutely attuned to the nuances of human behaviour, which enables them to detect the minutest traces of bias or inability to reach an appropriate decision” (p. 432)

However, for the most part it seems that lawyers rely on what Fulero and Penrod call “jury selection folklore” (1990, p. 229) which refers to longstanding beliefs held by members of the judiciary about the relationship between juror demographics and verdict preferences. However, as we have seen in previous sections of this chapter, there is not a great deal of empirical support for links between juror characteristics and verdict decisions (Kassin & Wrightsman, 1988), so the reliance on these methods for jury selection is unlikely to be effective (Hastie, 1991), and in fact suggests that lawyers may not perform any better than laypersons at detecting bias in jurors (Hepburn, 1980). A study by Olczak, Kaplan, and Penrod (1991) demonstrated this by comparing strategies used by lawyers and students in selecting jurors for a mock trial. The results showed that student and lawyer participants followed similar reasoning when deciding which jurors to exclude, so the authors concluded that lawyers rely on the same stereotypes held by laypersons when making assumptions about juror bias.

Given the uncertainty concerning the strategies used by lawyers for identifying juror bias, there is wide variability in pre-trial procedures with some courts having the Judge conduct the juror pre-trial questioning. The success of attempts to identify juror bias relies on the honesty of potential jurors, which may be problematic particularly as
social desirability may play a role in responses to questions about attitudes (Gobert, 1997). Although some argue that the social desirability effect is greater when judges conduct juror questioning (due to the elevated status of the judge), it is also argued that judges should be more neutral in their juror selection compared with lawyers who are inclined to select jurors who will be sympathetic to their side of the case (Jones, 1987). However research into self disclosure has also suggested that people tend to be more honest when they disclose personal information if they have built up rapport with the questioner, which is something that may occur more so when lawyers conduct voir dire hearings (Greene, Heilbrun, Fortune, & Nietzel, 2007).

Another area of particular concern is the potential for peremptory challenges to be used as a discriminatory tool, which is prohibited by the law but nonetheless is believed to occur in some trials (Norton, Sommers, & Brauner, 2007). In the U.S., the Batson Decision, made in the case of Batson v. Kentucky (1986), prohibited the exclusion of jurors based on ethnicity and this was extended to include gender (Greene, Heilbrun, Fortune, & Nietzel, 2007). In an attempt to determine whether gender and race are used as a basis for juror exclusion, a study by Rose (1999) observed 13 criminal trials and their associated voir dire hearings. The study found no relationship between gender and the likelihood of a juror being excluded by either the prosecution or defence. In addition it found that black jurors were not excused more frequently than white jurors overall, though it did conclude that the prosecution was more likely to exclude black jurors while the defence was more likely to exclude white jurors. The small sample of trials in this study (N = 13) limit the generalisability of the
findings, and the fact that the trials observed did not contain any gender or racially relevant issues may have also contributed to the null results.

A common methodology for investigating whether gender and race are used as grounds for challenges is described in a study by Norton, Sommers, and Brauner (2007). Participants were given a summary of a criminal case (involving a woman charged with murdering her abusive husband) and told to act as the prosecuting attorney when reading the profiles of two prospective jurors (differing only in gender) and were asked to fill the final spot on the jury by excluding one of these jurors. The majority of participants (71%) chose to exclude the female juror, and when asked to provide reasons for their decision only 10% cited gender-related explanations. The same results occurred when participants were given explicit instructions not to eliminate on the basis of gender. These findings suggest that gender does play a role in exclusion decisions, and that this can be easily justified using gender-neutral reasons, although the use of students for participants in this study may limit its ecological validity.

In a somewhat dated, but methodologically superior study, by Zeisel and Diamond (1978) a comparison was made between excluded and selected jurors in twelve real criminal trials. Excluded jurors were asked to remain in the courtroom and act as shadow jurors in the twelve cases, so their pre-deliberation verdicts could be compared with that of the real jurors, who were interviewed after the trials. The authors found that in seven of the twelve cases there were no significant differences between the two juries in terms of their verdict decisions. However in the other five cases, there were substantial differences in their reported perceptions of the
defendant’s guilt, but in only two of these cases would these differences have likely resulted in a different verdict. The authors concluded that jury selection can have an impact on the verdict outcome in some cases, but overall the attorneys were not very effective in their selection of whom to exclude. It is not clear why this study only analysed participants’ pre-deliberation decisions, as it may have been more informative to have allowed the shadow juries to deliberate so their final verdict decisions could be compared more directly with the actual juries’ decisions.

In an effort to improve the effectiveness of pre-trial juror selection, and the reliability of bias detection, so-called scientific jury selection emerged as a consultancy field, primarily catering to the US trial system (Moran, Cutler, & De Lisa, 1994). The belief held by many lawyers in the USA that cases can be won or lost at the pre-trial stage (Van Dyke, 1977; Lecci, Snowden, & Morris, 2004) has created a market for consultants who claim to take a more scientific and empirically-supported approach to identifying bias in potential jurors (Myers & Arena, 2001). Trial consultants (a more popular term, recently replacing ‘scientific jury selection’) typically possess a wide range of experience and backgrounds, including those who previously practised law themselves and those with psychology qualifications, and are hired to provide ‘expert’ guidance to the prosecution and/or defence attorneys during the jury selection phase of trials (Lecci, et al., 2004).

There are a number of different research strategies employed by jury consultants, such as surveys of relevant communities, mock trial simulations, focus groups and post-trial interviewing of jurors (Seltzer, 2006). Considering the empirical literature which suggests that attorneys are no better than laypersons or a random
selection of jury members, it seems reasonable to assume that more empirically-based strategies would be superior. However, some argue that professional trial consultant services are not supported by empirical research (Wallendael & Cutler, 2004). Those who argue against the commercialisation of such trial consultancy services worry that any empirical research carried out by these services is not likely to be published as consultants do not want their methods to be publicly available (Fulero & Penrod, 1990). This limits the peer review of such research, and therefore its reliability and validity is not discernable. However those who argue for the use of jury consultants cite the need for voir dire hearings to move away from juror bias folklore and become increasingly based in social psychological theory (Myers & Arena, 2001; Lecci, Snowden, & Morris, 2004).

Evaluating the effectiveness of trial consultation is extremely difficult and very little research has been dedicated to this question. One thing that makes comparisons very difficult is the lack of consistency with which voir dire hearings are conducted between different courtrooms (Sommers & Norton, 2008). Without a standard format for voir dire hearings, it is not possible to make generalisations about the comparative effectiveness of trial consultant advice. It is also important to note that consultants themselves are unlikely to provide reliable estimates of effectiveness, as they are likely to perceive a victory in a case as an indication of successful jury selection, while this might be more likely a function of the evidence presented in the case (Lecci, et al., 2004).

Lecci, et al. (2004) provide an argument for using social psychology research to improve jury consultancy and voir dire hearings generally, which in part provides
support for the new studies reported within this thesis. The questions asked of potential jurors in pre-trial hearings, or used by trial consultants, are typically not from empirically-derived, standardised measures, and thus the predictive validity and reliability of such questions is not known. Lecci et al. therefore argue that the use of standardised measures for predicting juror bias should be utilised to improve the jury selection process, and call for more research to develop such tools for use in voir dire hearings. While Lecci et al. suggest that existing measures such as the Juror Bias Scale (Kassin & Wrightsman, 1983) and the Legal Attitudes Questionnaire (Boehm, 1968) are more reliable than unsubstantiated practices of lawyers, judges, and trial consultants, they also note that research suggests more trial-specific measures perform better at predicting juror verdict preference.

Although the majority of the research literature concerning juror bias focuses on the court’s ability to identify problematic juror attitudes, which may form the basis of a peremptory challenge, an alternative is to provide training to jurors in order to improve their ability to evaluate physical evidence. DNA evidence in particular has been the focus of a recent research project conducted by Goodman-Delahunty and Hewson (2009), in which the effectiveness of providing mock jurors with a tutorial as part of the expert evidence was evaluated. At the outset of the study, participants’ knowledge about DNA evidence was very low (on average 24% of questions were answered correctly), regardless of their demographics or television viewing habits (a similar finding was also reported by Brewer & Ley, 2010). The results suggest that the use of a tutorial improved participants’ knowledge about DNA evidence by an average of 37%, which seems promising for overcoming difficulties with evidence
interpretation in the courtroom. However this study also concluded that participants who frequently watched CSI-type programmes and rated the content of such shows as realistic, learned significantly less from the DNA tutorial. Although the viewing of forensic television did not have a direct effect on verdict preference in this sample, it does appear to be the case that beliefs held by those who reported watching forensic programmes interfered with the effectiveness of the DNA tutorial. This preliminary finding suggests that jurors with strong pre-trial beliefs about forensic evidence may not respond to courtroom efforts to educate in the hopes of removing such bias.

The arguments made above for further contributions from the social psychology literature, and more attention to appropriate standardised attitude measures for identifying juror bias, form the basis for the approach taken in Chapters five through eight of this thesis. The development in this thesis of a measure, designed to identify pre-trial attitudes about forensic science evidence, combines the theoretical and empirical approaches summarised within this literature review. This provides a novel approach to evaluating the importance of pre-trial attitudes about forensic evidence on pre-deliberation verdict preferences, and results in opportunities for further research regarding the use of such a measure in pre-trial hearings.
CHAPTER TWO: METHODOLOGICAL CONSIDERATIONS

This chapter outlines aspects of the research methodologies utilised throughout this thesis, which require consideration and justification with reference to the existing body of jury decision making literature.

2.1 MOCK JUROR PARADIGM

Despite the legal and moral importance inherent in decisions made by juries, the vast majority of research investigating the decision making processes of jurors has been conducted through trial simulations. Legislation prohibiting observation of real jury deliberations and post-trial interrogation of jury members has led researchers to adopt alternative methodologies for jury research (e.g. Contempt of Court Act, 1981 in the UK). There are two main arguments for preserving the sanctity of jury deliberations, firstly such research might encourage doubt in the jury system and secondly that jurors might feel unable to express their true views for fear of observation and accountability (Auld, 2001). These concerns were summarised by Dr. Glanville Williams:

“The real reason for keeping the jury’s deliberations secret is to preserve confidence in a system which more intimate knowledge might destroy” (as cited in Auld, 2001, p. 165)

The use of mock juries in research is a contentious issue, with some authors arguing that the artificial nature of such studies and participants render them difficult (if not impossible) to generalise to real jury situations (Bray & Kerr, 1982; Diamond, 1997; Thomas, 2010). However other authors suggest that the use of carefully
designed and controlled mock jury experiments are the only way to fully understand the cognitive processes and trial variables which may contribute to verdict decisions (MacCoun, 1989), particularly as manipulation of these variables is not ethical in real trial situations.

In a review of jury simulation methods, Diamond (1997) identified a number of features of jury research which could impact the validity of results. The first of these features is the sampling of participants, and their similarity (or lack thereof) to real juries. The use of university students as participants in general psychology research is widespread, and this is no different in the jury research subset of the literature. Apart from the fact that students tend to be younger than real jury members (HMCS, 2009; Bray & Kerr, 1982), they have also been found to be more likely to succumb to normative pressures and are better at complex cognitive tasks than general members of the public (Diamond, 1997). However it may be the case that these differences between students and general members of the jury-eligible public do not significantly impact on verdicts or evaluation of trial evidence. In a review of research comparing student and non-student mock jurors, Bornstein (1999) found that only 19% of studies found a main effect of sample on verdict preferences. In order to address this sampling issue within this thesis, a range of samples have been utilised across the following chapters which included jury-eligible students, jury-eligible members of the public, and participants recruited from actual jury venires. This will allow some comparisons to be made across studies, and these findings will be reviewed in the final discussion (see Chapter 9).
Another feature of trial simulations which is commonly highlighted as potentially problematic is the use of vignettes, or written trial transcript materials as stimuli in jury studies. In the existing literature a range of media have been used to present trial information to participants including live trial re-enactments, videotaped trials, audio recorded testimony, and written transcript materials (DeMatteo & Anumba, 2009). Even the most elaborate and seemingly realistic trial simulation does not include the extensive and complex information given to juries in real trials, and so the validity of the over-simplified stimuli has been questioned (Bridgeman & Marlowe, 1979). Some research has directly compared the impact of various trial media on mock juror participants, and the majority of these studies have found that the presentation medium does not have an impact on verdict (Bornstein, 1999). In addition to these direct comparisons, Bornstein found that several studies have successfully replicated their findings using a variety of trial simulation formats.

The view taken in this thesis is that the use of brief written trial vignettes was preferred to more elaborate stimuli for a number of reasons. Firstly, the production of more comprehensive and realistic trial re-enactments are costly and time consuming, and given the lack of consensus regarding the impact of such material found in the literature it was not thought to be justified. Secondly, the studies described within this thesis aimed to isolate and manipulate particular aspects of the juror decision making task, and the use of brief written trial materials acts to minimise the possible effect of other extraneous information (e.g. the demeanour of the defendant, lawyers, etc.). A final consideration was the decision to use the internet to distribute the study materials for completion online, which further justified the use of written stimuli.
Although it would have been possible to present other types of media (e.g. videos or audio files) to participants through online administration of the study, this would potentially cause technical issues for participants due to software limitations and compatibility issues.

2.2 Online research

As mentioned in the previous section, the decision was made to present participants with trial stimuli consisting of written materials and a second methodological consideration was the way in which the material was delivered to and completed by participants. One of the reasons that the existing literature is dominated by studies in which students play the role of mock jurors is the fact that this population is easily accessible by academic researchers and therefore relatively large sample sizes can be obtained with little effort. However an alternative method for recruiting large samples of participants more representative of the jury-eligible public is the use of the internet. Studebaker et al. (2002) describe the potential benefits of research conducted online including lower costs and less time and travel commitment for participants compared to methods requiring attendance in a central location (e.g. the researcher’s laboratory).

As described in the previous section, the reliance on students as research participants is often criticised for resulting in mock jurors who are significantly different from actual jurors in age, gender distribution, and educational level (Bray & Kerr, 1982). Research carried out in an online environment has faced similar criticisms regarding sampling biases. However, studies comparing online and offline generated, non-student samples have found that there may be little difference between these
two types of samples in most demographic characteristics (Ballard & Prine, 2002). A study by Smith and Leigh (1997) compared internet and non-internet samples and concluded that participants did not differ significantly with respect to sexual orientation, marital status, ethnicity, education, or religion. Their samples did differ however in age and gender. The internet sample in the Smith and Leigh study was biased toward male respondents, and the non-internet sample showed a stronger opposite bias for females. However, other research has supported the view that internet-based samples of participants are increasingly diverse and are not likely to differ significantly from non-internet, non-student samples (Gosling & Johnson, 2010; Krantz, Ballard, & Scher, 1997; Stanton, 1998).

Although the use of internet recruitment and participation limits the sample to those with access to the internet, this was considered less problematic than confining the entire thesis to the use of students as participants. In addition, it could be argued that although some demographic groups might be underrepresented in research using internet samples there are also groups which are less likely to be recruited as jurors. It is well documented that certain segments of the population are underrepresented on juries due to the selection practices used by the courts (Hastie, 1993) and therefore it is clear that jury selection itself does not follow the ideal of random selection. The decision was therefore made for the purposes of this thesis that there were likely to be less sampling biases regarding age, gender and education using online techniques than there would be relying on student participants for every study.
2.3 Dependent Variables

A final methodological consideration relates to the design and measurement of the dependent variables of interest within this thesis. In a real courtroom juries are only required to reach a dichotomous verdict decision – guilty or not-guilty of the charges against the defendant. Although this ultimate decision has been recorded in the studies described in the following chapters, other variables were required in order to further understand the underlying processes involved in the participants’ decisions.

Existing research which has taken a Bayesian approach to juror decision processes (described in Chapter one) has utilised measures of estimated probability of guilt, which allowed some conclusions about the impact of various pieces of trial evidence to be made and compared against a Bayesian normative calculation (Goodman, 1992; Nance & Morris, 2005; Thompson & Schumann, 1987). This use of probability of guilt measures was adopted in this thesis as a means of quantifying the perceived strength of the prosecution’s case (pre- and post-evidence), and was shown to predict final verdict preference. In addition to probability of guilt estimates, the use of Likert scale ratings of probative value were used to quantify the perceived strengths of different evidence which could then be analysed to determine the relative contributions of different testimony to the perceived probability of the defendant’s guilt (Robson, 2002).

The decision to have participants quantify their perceptions of the defendant’s guilt in the trial scenarios used in this thesis requires further justification. Although previous literature applying Bayes’ theorem to juror decisions uses similar measures of prior and posterior probability of guilt (as described in Chapter one), some authors
question the validity of this approach. Schum (1994) argues that using the difference between prior and posterior probabilities of guilt as an indicator of the “inferential force” (p. 254) of the evidence presented may be problematic. This is due to the fact that probabilities are measured on a scale of 0 to 1 (or 0 to 100, if expressed as percentages), and therefore if the prior probability is close to 1, there is very little room for adjustment in the light of very strong evidence. For example, if a participant indicates that they perceive the probability of the defendant’s guilt to be 0.9 before hearing very compelling evidence, their estimate can only be adjusted by 0.1 regardless of the perceived weight of the evidence. The result is therefore an inaccurate quantification of how strong the participant considered the evidence to be.

A second issue raised by Schum (1994) is the fact that using probability estimates in this way assumes that participants are able to make rational, and mathematically correct judgments, and correctly express these decisions in a numerical format. Chapter one has already reviewed some research findings which suggest that participants are not savvy about probabilities and frequencies, so perhaps it is overly optimistic to assume that participants can correctly quantify their subjective opinions. However, Taroni, Aitken, Garbolino, and Biedermann (2006) argue that probabilities in the context of legal decisions are not entirely about the correctness of the numbers assigned. They support the use of probability of guilt estimates as it is not the absolute value or accuracy which is of interest, but rather the relative adjustments made by participants in light of new evidence.

Finally, Baron (2000) describes a number of different theoretical views for the use of probabilities as measures in decision making research, and defines probability in
these situations as “a numerical measure of the strength of a belief in a certain proposition” (p. 94). The ‘personal theory’ of probabilities, as described by Baron, emphasises the fact that probability judgements made in situations of uncertainty are expressions of personal belief and knowledge about a situation, rather than a measure of fact or precision. The methodology utilised in this thesis adopts the view that the mock juror paradigm is an ideal example of subjective decision making plagued with uncertainty, and therefore the personal theory perspective is the most appropriate view of probability judgements in these experimental conditions.

Having offered some justification for the main methodological considerations underlying the mock juror paradigm utilised throughout this thesis, the following chapters present the findings of the six studies comprising the thesis. Additional methodological and statistical issues arise in each chapter, and will be addressed in the corresponding method and discussion sections of each study. The final discussion chapter will once again address the issue of sampling raised in this chapter, as the different samples used across the six studies will be critically compared in that final chapter.
CHAPTER THREE: GENERAL FORENSIC AWARENESS AND EVIDENCE EVALUATION

3.1 ABSTRACT

Members of the public who are called for jury service are often required to make decisions regarding physical evidence presented in court. The story model of juror decision making highlights the importance of prior knowledge for juror decision making, and the current study investigated the source of people’s knowledge about forensic science and the accuracy of such knowledge. This study utilised a survey design to investigate the accuracy of participants’ beliefs about common forensic science techniques, in order to determine whether forensic awareness is related to educational level or reported source of forensic knowledge (e.g. movies/television, novels, news media, etc.). In addition, participants (N = 163) rated the probative value of forensic evidence, which was manipulated by varying the degree of mobility and relevance of the evidence. Results indicated that there is much variability in the accuracy of participants’ knowledge of various forensic techniques, and overall forensic awareness was not significantly related to educational level or reported source of knowledge. Participants performed surprisingly well on the evidence evaluation task, and did appear to be taking mobility and relevance of evidence into account when determining probative value. These findings are discussed with reference to the previous literature and theory, and the implications for real juries are considered.
3.2 Introduction

Since the early 1900s, when techniques for identifying people using fingerprint recovery and analysis were developed, forensic science principles and technologies have continued to develop new ways for investigators of criminal acts to identify perpetrators. In particular the last fifty years has produced advances in technology which have enabled an increasing number of types of forensic analyses, with the ability to identify many different materials and pieces of trace evidence recovered from crime scenes (Caddy & Cobb, 2004). The most notable recent development in the field of forensic science has been the application of DNA profiling to the identification of individuals through comparisons with crime scene samples. This DNA technology has been rapidly improving since its first forensic application in the 1980s, resulting in increasingly sensitive recovery techniques and testing procedures (Duncan, 2008).

Traditionally research which has focused on how jurors evaluate physical evidence has been interested in determining whether jurors understand the presentation of probabilistic information which often accompanies evidence such as blood typing or DNA. Since the usefulness of blood typing or DNA evidence is dependent upon population genetics and frequencies of certain genetic markers in the relevant population, these types of evidence are often accompanied by complicated statistical information which jurors have been found to have problems understanding (Thompson & Schumann, 1987; Smith, Penrod, Otto, & Park, 1996; Goodman, 1992).

Difficulties understanding the probabilistic nature of some sources of physical evidence are not the only potential sources of error which jurors face during criminal
trials. A similar, but conceptually different error could be made, resulting from a misunderstanding of the probative value of a piece of evidence in relation to what exactly the evidence is capable of telling the court about the events in question. This is referred to by Schklar and Diamond (1999) as the misperception error. In other words, correctly evaluating the ability of the evidence, and related technology, to identify a particular person is one part of the juror’s decision making task, but a second aspect of the task requires the juror to understand that the evidence may not support the inference that such an identification indicates guilt (Roberts & Zuckerman, 2004).

The focus of this thesis is on this second potential source of juror error, and this chapter investigated whether mock jurors were able to differentiate between pieces of physical evidence which are similar in nature and type but are of varying probative strengths. As it is well documented that jury decision making is informed by peoples’ prior knowledge and beliefs (Pennington & Hastie, 1986, 1988, 1993; Wagenaar et al., 1993), the study described in this chapter used a measure of general forensic awareness to determine whether knowledge about forensic science practices influenced participants’ evidence evaluation abilities.

The term ‘forensic awareness’ has been used quite extensively, both in the research literature and in news media coverage of forensic-related stories. This term has not been consistently defined in any of these sources, so for the purposes of this thesis ‘forensic awareness’ refers to the degree to which a person’s perceptions and beliefs about common forensic techniques match the reality of these techniques in practice (Magnussen, Melinder, Stridbeck, & Raja, 2009).
This study utilised a questionnaire to investigate participants’ perceptions of various forensic techniques. The first part of the questionnaire presented participants with a series of statements about a range of forensic techniques and asked for their level of agreement with each statement. The second part of the questionnaire asked participants to indicate the perceived strength (probative value) of various types of forensic evidence. Together, these two tasks provided an indication of how realistic and accurate participants’ beliefs about forensic techniques were, as well as how they perceived the varying strengths of common types of evidence.

3.3 Method

3.3.1 Participants

The target population for this study was members of the general public who are eligible for selection as jurors using the basic criteria described by the Criminal Justice System for England and Wales (www.cijonline.gov.uk). These criteria require jurors to be at least 18, but not more than 69 years of age, and be registered as a parliamentary or local government elector. They must also have lived in the UK for at least five years since they were 13 years of age and cannot have been convicted of a criminal offence for which they have served a prison sentence or community order. For the purposes of verifying basic jury eligibility for this study participants were asked to indicate their age and whether they have ever been convicted of a criminal offence. These details were not able to be confirmed independently and therefore relied on accurate self report by participants.

In total, 173 people completed the online questionnaire however some respondents had to be excluded from the analyses. One respondent indicated that
they did not give consent to their responses being included in the analyses, six respondents indicated that they had been convicted of criminal offences, and two participants were under the age of 18 and therefore did not meet the minimum criteria for jury eligibility for the purposes of this study. One additional participant did not appear to have been answering honestly because all of their responses to the items in the second section of the questionnaire were the same. Therefore ten respondents in total were excluded from the analyses resulting in a total sample size of 163. The final sample consisted of 118 female and 45 male respondents, aged between 18 and 65 years (\(M_{age} = 32.55, SD = 11.81\)). There was a range of highest educational level reported by participants, with 47% having completed high school, 23% with undergraduate degrees, and 30% with a postgraduate qualification.

3.3.2 MATERIALS

The online presentation of the questionnaire was developed using the Bristol Survey Tool, for which the University of Leicester holds a user license. This tool enables an online questionnaire to be designed and distributed, offers a number of options in question and answer format and also collects and stores the responses for analysis. The data can be exported directly into Excel/SPSS for analysis. Once the questionnaire was designed using this software, a web address was assigned for the project which participants could then be directed to in order to complete the questionnaire. The link to the questionnaire website was able be communicated to potential participants by advertising the address online, as well as by email link which was particularly useful for utilising a snowballing sampling technique.
The questionnaire began with the collection of demographic information about the participant. As indicated above, some of the information collected was necessary in order to determine whether the participant met the basic jury eligibility criteria for this study. Other items in this section of the questionnaire were used for further analyses as independent variables, such as gender and education. The final question in this section of the questionnaire asked participants to indicate where they felt the majority of their knowledge about forensic science comes from (television/movies, news media, true crime novels, formal education, career, involvement in a court case).

Following the collection of demographic details, the questionnaire was divided into two sections. Each is described below.

**Part One: General forensic awareness/knowledge.** This first section of the questionnaire presented participants with a series of statements about a variety of commonly used forensic techniques (DNA, fingerprints, ballistics, gunshot residue, footwear analysis, and glass/trace evidence). Respondents were asked to indicate their level of agreement with each statement using a five-point Likert Scale format, coded as strongly disagree (1) to strongly agree (5).

Some items in this section were negatively worded (e.g. statement about the forensic technique was false), and therefore these items were scored in reverse. This scoring resulted in a higher score being associated with a more accurate belief about the forensic technique (e.g. as compared with the use of the technique in practice). The following are two sample items in this section relating to DNA evidence:
• DNA can be used to positively identify someone. *(positively worded item, as correct answer in relation to the research literature is ‘yes’)*

• A DNA profile can only be recovered from large amounts of DNA – e.g. samples that are visible to the naked eye. *(negatively worded item, as correct answer is ‘no’)*

The items in this section of the questionnaire were developed with reference to a number of up-to-date, core forensic science texts, and the forensic techniques chosen for use in the questionnaire represent the most commonly used techniques in practice (White, 2004; Williams & Johnson, 2008). To ensure the most up-to-date information concerning practical uses of the techniques, the most recent fact-sheets produced by the Forensic Science Service (UK) were consulted (Forensic Science Service, 2003a, 2003b, 2004, 2005, 2006a, 2006b, 2006c, 2007a, 2007b, 2007c). These fact-sheets are circulated to police forces in the UK with the objective of keeping police personnel informed of the most current processing techniques for forensic evidence. The full list of items in this section of the questionnaire, with references to the relevant literature, is available in Appendix A.

**Part Two: Evaluating the strength of forensic evidence.** The second section of the online task presented participants with a series of statements about various types of physical evidence recovered under varying forensic circumstances. The participants indicated, using a Likert Scale, their perception of the strength of each type of evidence (in terms of usefulness for determining the guilt of a suspect) as described in each item. The Likert scale had five points and was coded from
irrelevant/useless (1) to moderate (3) to conclusive/definite (5) (therefore a higher score indicated evidence which was perceived to have high probative value).

This section was less accuracy-focused than the first task and more concerned with the relative strengths assigned to various types of evidence presented under different circumstances. This is especially true given that very little contextual information was provided in these items, which would normally be used by jurors to determine the overall evidential value of any particular piece of evidence. In the absence of this additional information it was not realistic to evaluate the absolute accuracy of participants’ responses, but rather the measure of interest was the perceived strengths of the items relative to one another.

In this section, the evidential items are not only varied in terms of the type of forensic evidence (fingerprints, DNA, footwear, etc.) but also in terms of the investigative relevance of the evidence and the mobility of the sample. The mobility of an evidence sample is an important factor to consider when evaluating evidence strength because the more mobile a sample is the less evidential value it has (Champod et al., 2004; Bond, 2007). This is due to the fact that mobile samples have more potential to be innocently explained, as it is plausible that they could have been deposited at a crime scene without requiring the associated person to be involved in the offence (e.g. a discarded cigarette end found at a scene which could have been innocently transported there from somewhere else).

The investigative relevance of a sample is determined by a number of factors, which are largely dependent on the nature of the sample. Issues which were considered to determine investigative relevance included the location of the sample
(e.g. on a suspect or victim’s body, or in a room, etc.) and the uniqueness of the sample (Champod et al., 2004). Generally, investigative relevance can be conceptualised as the extent to which guilt can be directly inferred from the evidence.

Manipulating the evidence in terms of mobility and investigative relevance allowed for analyses to determine whether these details were associated with the participants’ perceptions of the probative value of a particular item of evidence.

The manipulation of these two features of the evidence resulted in four relative strengths of evidence as illustrated in Figure 1:

<table>
<thead>
<tr>
<th></th>
<th>Low Mobility</th>
<th>High Mobility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Relevance</td>
<td>Moderate/weak evidence</td>
<td>Weakest evidence</td>
</tr>
<tr>
<td>High Relevance</td>
<td>Strongest evidence</td>
<td>Moderate/strong evidence</td>
</tr>
</tbody>
</table>

*Figure 1*. Relative strengths of evidence as a function of mobility and investigative relevance.

Although two manipulations were used for each item of evidence (mobility x relevance), it is important to understand the impact of these concepts individually in order to explain the relative strengths of each item of evidence. As illustrated in Figure 1, investigative relevance is treated as a more salient factor in the
determination of probative value, as the strongest categories of evidence must have high levels of investigative relevance. This is due to the fact that investigative relevance is generally a necessary condition for evidence to meet in order for it to be considered admissible in a court of law (under normal circumstances) (Roberts & Zuckerman, 2004). Typically, evidence which is of very low investigative relevance (regardless of its level of mobility) would not be considered admissible, however in the case of mobility both levels (high and low) can be considered admissible if they are of sufficient investigative relevance.

The following are examples of items in this section of the questionnaire (note that the information in brackets/italics was not given to participants):

i. DNA recovered from discarded chewing gum found on the pavement outside a burglary scene. *(high mobility / low relevance)*

ii. A suspect's fingerprint found on a table in a burglary scene. *(low mobility / high relevance)*

iii. Fingerprints recovered from an empty crisp packet in a stolen car. *(high mobility / high relevance)*

iv. Shoeprints in a garden outside a burglary scene match a suspect's footwear. *(low mobility / low relevance)*

A complete list of the items used for this section of the questionnaire is provided in Appendix B.
For each item in this section participants were also asked to indicate their level of confidence in their answer using a four-point scale ranging from ‘entirely unsure’ (1) to ‘extremely confident’ (4).

3.3.3 Procedure

Once the questionnaire was developed and adapted into an online version it was piloted on a group of students who were participating in a course on advanced online research methods (N = 28), at the University of Leicester. The feedback received from the pilot participants was incorporated into the final version where appropriate, for example some of the questions contained forensic-related language which was not well understood by pilot participants and in other items descriptions of evidence were considered to be too ambiguous and were subsequently made clearer in the final version of the questionnaire.

The recruitment of participants for this research project was achieved through two advertising approaches. The first involved posting a summary of the research project on the webpage for the Participant Panel at the University of Leicester School of Psychology. This webpage is accessed by members of the public who are interested in participating in research projects within the School of Psychology and who have voluntarily signed up to be members of the panel. A press release, which included a summary of the project and a call for volunteers, was also issued through the University Press Office in order to attract participants from the wider public. A snowballing technique was also employed by asking participants to forward the study to others after they had completed the online tasks. The use of a snowballing technique can encourage participants to forward the details of the online
questionnaire to other potential participants who meet the sampling criteria (Robson, 2002).

3.4 RESULTS

3.4.1 FORENSIC AWARENESS

Participants were asked to select from a list of sources where they felt the majority of their knowledge about forensic science comes from. Figure 2 illustrates the distribution of responses to this question.

![Figure 2. Source of forensic knowledge as indicated by participants (N = 163)](image)

Almost half of respondents (48%) indicated that most of their knowledge about forensic science comes from watching movies and television programmes, with the second most popular response being news media (21%). The remaining three
categories which contained responses were crime novels, career/professional experience, and education which had roughly equal numbers of responses (10%, 11%, and 9% respectively). No respondents indicated that their knowledge of forensic science comes mainly from personal involvement in the Criminal Justice System.

There were a total of 22 items in the first section of the questionnaire, in which respondents indicated their level of agreement with statements about the use of various forensic science techniques. The coding of these responses resulted in a score of five for highest level of agreement with an accurate response, and a score of one for the most disagreement with an accurate response. A score of three indicated that the respondent had no opinion about the questionnaire item.

The forensic awareness overall scores (the sum of scores for all items) ranged from 68 to 106 with a mean score of 81.66 ($SD = 5.88$). A histogram of these scores revealed that they are distributed relatively normally, and a Kolmogorov-Smirnov test confirmed normality, $D(163) = .07, p = .06$.

Table 1 shows the distribution of responses to each of the items in the first section of the questionnaire. The response which is most accurate (as determined by the relevant research literature) is highlighted in bold italics for each item. When determining the frequency of accurate responses in Table 1 the categories ‘agree’ and ‘strongly agree’ were combined, as were responses ‘disagree’ and ‘strongly disagree’. A series of one-sample t-tests were conducted to determine whether the mean score for each item differed significantly from a test score of three (which was the ‘no opinion’ scale response) in order to determine whether participants held opinions about the items on average (Phillipon, Cherryman, Bull, & Vrij, 2007). The results
indicated that only three items did not differ significantly from a score of three at \( p < .05 \) (these items are indicated in Table 1).

There is a substantial amount of variation in accuracy across the various items in this section of the questionnaire. Respondents were quite accurate in their beliefs regarding whether DNA and fingerprints could be used to positively identify a person (90.1% and 92.6% answered correctly respectively), and there were also a large percentage of correct responses to items referring to whether DNA can be recovered from blood (92.0%) and whether analysis of a bullet can determine the type of gun used (95.1%).

The items with the lowest percentages of correct answers both concerned gunshot residue (GSR) evidence with only 22.7% of respondents believing that GSR can be washed off a person’s hands and only 20.2% believed that GSR can identify the manufacturer of a firearm cartridge.
Table 1

*Percentage of responses for each category of agreement and overall item mean score*

<table>
<thead>
<tr>
<th>Statement</th>
<th>Agree</th>
<th>No opinion</th>
<th>Disagree</th>
<th>Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA can positively ID someone</td>
<td>90.1</td>
<td>0.61</td>
<td>9.2</td>
<td>4.28*</td>
</tr>
<tr>
<td>Fingerprints can be recovered from any surface</td>
<td>49.1</td>
<td>14.1</td>
<td>36.8</td>
<td>2.83</td>
</tr>
<tr>
<td>DNA can be recovered from chewing gum</td>
<td>80.4</td>
<td>12.9</td>
<td>6.7</td>
<td>3.96*</td>
</tr>
<tr>
<td>GSR stays on hands after firing a gun, even after washing</td>
<td>60.1</td>
<td>17.2</td>
<td>22.7</td>
<td>2.54*</td>
</tr>
<tr>
<td>Impossible to tell who is the regular wearer of a pair of shoes</td>
<td>19.6</td>
<td>25.1</td>
<td>55.2</td>
<td>3.45*</td>
</tr>
<tr>
<td>DNA profile can only come from visible stains</td>
<td>8.0</td>
<td>6.7</td>
<td>85.3</td>
<td>4.23*</td>
</tr>
<tr>
<td>DNA can be recovered from blood</td>
<td>92.0</td>
<td>6.1</td>
<td>1.8</td>
<td>4.34*</td>
</tr>
<tr>
<td>Fingerprints can positively identify someone</td>
<td>92.6</td>
<td>3.7</td>
<td>3.7</td>
<td>4.34*</td>
</tr>
<tr>
<td>DNA cannot be recovered from a surface someone has touched</td>
<td>24.5</td>
<td>19.6</td>
<td>55.8</td>
<td>3.50*</td>
</tr>
<tr>
<td>Fingerprints can be destroyed by an offender to avoid detection</td>
<td>74.2</td>
<td>12.3</td>
<td>13.5</td>
<td>3.75*</td>
</tr>
<tr>
<td>Footwear impressions can be used to link crimes</td>
<td>74.2</td>
<td>17.2</td>
<td>8.6</td>
<td>3.75*</td>
</tr>
<tr>
<td>Only complete fingerprints can be used to ID someone</td>
<td>15.3</td>
<td>14.7</td>
<td>69.9</td>
<td>3.64*</td>
</tr>
<tr>
<td>A bullet can determine the type of weapon used</td>
<td>95.1</td>
<td>2.4</td>
<td>2.4</td>
<td>4.37*</td>
</tr>
<tr>
<td>Footwear impressions can be recovered from clothing &amp; bodies</td>
<td>55.8</td>
<td>27.6</td>
<td>16.6</td>
<td>3.47*</td>
</tr>
<tr>
<td>A bullet is only useful for analysis if it is not damaged</td>
<td>17.2</td>
<td>25.8</td>
<td>57.1</td>
<td>3.50*</td>
</tr>
<tr>
<td>Detectable amounts of GSR can be transferred between suspect and armed police</td>
<td>33.7</td>
<td>46.0</td>
<td>20.2</td>
<td>3.13</td>
</tr>
<tr>
<td>Fragments of glass from suspect’s clothing can be matched to broken glass at scene</td>
<td>65.0</td>
<td>22.7</td>
<td>12.3</td>
<td>3.67*</td>
</tr>
<tr>
<td>Regardless of quality of footwear impression, a conclusive match is unlikely</td>
<td>25.1</td>
<td>23.9</td>
<td>50.9</td>
<td>3.29*</td>
</tr>
<tr>
<td>GSR can be used to determine the manufacturer of a firearm cartridge</td>
<td>20.2</td>
<td>42.9</td>
<td>36.8</td>
<td>2.80*</td>
</tr>
<tr>
<td>A bullet can be traced to the weapon which fired it</td>
<td>78.5</td>
<td>10.4</td>
<td>11.0</td>
<td>3.94*</td>
</tr>
<tr>
<td>If a suspect breaks a window, glass traces will remain on clothes until washed</td>
<td>41.7</td>
<td>31.3</td>
<td>27.0</td>
<td>2.83*</td>
</tr>
<tr>
<td>GSR can be linked to a particular fired bullet</td>
<td>28.2</td>
<td>29.4</td>
<td>42.3</td>
<td>3.15</td>
</tr>
</tbody>
</table>

*Note.* Most accurate response highlighted in bold italics. These are abbreviated versions of the items presented to participants (see Appendix A for the full wording).

*indicates that the mean score differs significantly from the ‘no opinion’ score of 3, at $p < .05$
3.4.2 Relationship between educational level and overall forensic awareness score

**Hypothesis:** Participants with higher educational achievement will have higher overall forensic awareness scores.

A one-way, between subjects ANOVA was conducted to investigate the relationship between participants’ educational level and their overall forensic awareness score. The descriptive statistics for mean forensic awareness scores by educational level are displayed in Table 2.

As the design of this study is non-experimental, it was not possible to allocate an equal number of participants to each category of the independent variables (e.g. educational level, source of forensic knowledge, etc.). This results in unequal numbers of participants in each category which is not ideal for statistical analyses. In some instances the number of participants in each category can be equalised by randomly deleting responses, however in the case of non-experimental research this is not always appropriate as it misrepresents the actual proportions of people in each category in the population from which the sample is drawn (Tabachnick & Fidell, 2007). For the purposes of the following analyses the unequal sample sizes are considered to be of little concern as the impacts of this on simple one-way ANOVA calculations are minimal, particularly as the assumption of homogeneity of variance is still met in each of the following cases (Tabachnick & Fidell).
### Table 2

**Forensic awareness scores by educational level**

<table>
<thead>
<tr>
<th>Educational level</th>
<th>N</th>
<th>Mean FA score</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary</td>
<td>77</td>
<td>81.26</td>
<td>6.42</td>
<td>68</td>
<td>106</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>37</td>
<td>83.54</td>
<td>5.51</td>
<td>73</td>
<td>99</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>49</td>
<td>80.88</td>
<td>5.03</td>
<td>68</td>
<td>94</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td>81.66</td>
<td>5.88</td>
<td>68</td>
<td>106</td>
</tr>
</tbody>
</table>

The results of the ANOVA indicated that there was no significant difference in mean forensic awareness scores between the educational level groups, $F(2, 160) = 2.55, p = .08$. Therefore the hypothesis that mean forensic awareness scores would differ between educational level groups was not supported.

### 3.4.3 The relationship between forensic awareness and source of forensic knowledge

**Hypothesis:** Participants who indicated that their forensic knowledge comes mainly from tv/movies would score lowest on forensic awareness (mainly due to the unrealistic nature of portrayals of forensic science in fictional programmes) and those participants who have had experience with forensic science through their professional or educational careers will have the highest forensic awareness scores.
The relationship between the source of participants’ forensic knowledge and overall forensic awareness score was investigated using a one-way, between-subjects ANOVA. Descriptive statistics are provided in Table 3.

Table 3

*Descriptive statistics for forensic awareness scores by source of forensic knowledge*

<table>
<thead>
<tr>
<th>Source of forensic knowledge</th>
<th>N</th>
<th>Mean FA score</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime novels</td>
<td>17</td>
<td>85.12</td>
<td>7.81</td>
<td>77</td>
<td>106</td>
</tr>
<tr>
<td>Education</td>
<td>14</td>
<td>84.21</td>
<td>4.89</td>
<td>76</td>
<td>97</td>
</tr>
<tr>
<td>Career</td>
<td>18</td>
<td>82.11</td>
<td>4.70</td>
<td>74</td>
<td>91</td>
</tr>
<tr>
<td>News Media</td>
<td>34</td>
<td>81.88</td>
<td>5.65</td>
<td>73</td>
<td>98</td>
</tr>
<tr>
<td>TV/Movies</td>
<td>80</td>
<td>80.29</td>
<td>5.56</td>
<td>68</td>
<td>94</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td>81.66</td>
<td>5.88</td>
<td>68</td>
<td>106</td>
</tr>
</tbody>
</table>

A one-way ANOVA revealed overall a significant difference in mean forensic awareness scores between the various sources of forensic knowledge, $F(4, 158) = 3.45, p < .05, r = .28$. The results of the ANOVA are summarised in Table 4. Post hoc tests using the Bonferroni correction found that the difference in means was only significant between the categories TV/movies and crime books ($p < .05$), therefore there are no statistically significant differences between the remaining forensic knowledge categories.
Table 4

Analysis of variance summary table for forensic awareness scores by forensic knowledge source

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>450.62</td>
<td>4</td>
<td>112.66</td>
<td>3.45</td>
<td>.01</td>
</tr>
<tr>
<td>Within groups</td>
<td>5155.82</td>
<td>158</td>
<td>32.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5606.44</td>
<td>162</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The hypothesis was only partially confirmed by this analysis as the category of tv/movies did in fact score the lowest on forensic awareness. However, this was only significantly different from the highest scoring category (crime books). These results need to be interpreted with caution as the ‘source of knowledge’ variable had unequal sample sizes in each group, due to the naturally occurring variations in participants’ responses to this item. Although the $F$ statistic is thought to be robust when sample sizes are unequal if the assumptions of normality and homogeneity of variance are met (as they are with these data), these results should be considered with the uneven sample sizes noted (Field, 2009; Tabachnick & Fidell, 2007).

3.4.4 Forensic awareness scores by evidence type

Hypothesis: Participants would be more knowledgeable about DNA and fingerprint techniques due to the significant media coverage given to these forensic procedures compared with the other evidence types.
The questionnaire items were grouped by evidence type producing six mean forensic awareness scores for each participant in the following evidence categories; DNA, fingerprints, footwear impressions, glass/trace evidence, ballistics, and gunshot residue (GSR).

A one-way within-subjects ANOVA was used to determine whether participants were more knowledgeable about certain types of evidence over others. The descriptive statistics for the various evidence types are displayed in Table 5:

Table 5

*Mean forensic awareness scores for each evidence type*

<table>
<thead>
<tr>
<th>Evidence type</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballistics</td>
<td>3.94ₐ</td>
<td>0.61</td>
</tr>
<tr>
<td>DNA</td>
<td>3.87ₐ</td>
<td>0.45</td>
</tr>
<tr>
<td>Fingerprints</td>
<td>3.64ₐ</td>
<td>0.46</td>
</tr>
<tr>
<td>Footwear</td>
<td>3.49ₐ</td>
<td>0.59</td>
</tr>
<tr>
<td>Glass</td>
<td>3.25ₜ</td>
<td>0.58</td>
</tr>
<tr>
<td>GSR</td>
<td>2.91ₜ</td>
<td>0.41</td>
</tr>
</tbody>
</table>

*Note.* Means with different subscripts differ significantly at *p* < .01, N = 163

The results of the one-way, within-subjects ANOVA are summarised in Table 6. Mauchly’s test of sphericity revealed that this assumption had been violated by these data (*χ²*(14) = 56.84, *p* < .05) and in order to address this violation the degrees of freedom used in the ANOVA were corrected using Greenhouse-Geisser estimate of
sphericity (ε = .87) (Field, 2009). The results of the ANOVA show that there was overall a significant difference in the mean forensic awareness scores between the six categories of evidence, $F(4.34, 703.74) = 99.14, p < 0.01, \eta^2 = .38$.

Table 6

*Repeated measures analysis of variance summary table for forensic awareness scores by evidence type*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence type (sphericity assumed)</td>
<td>124.21</td>
<td>5</td>
<td>24.84</td>
<td>99.14</td>
<td>.000</td>
</tr>
<tr>
<td>Greenhouse-Geisser correction</td>
<td>124.21</td>
<td>4.34</td>
<td>28.59</td>
<td>99.14</td>
<td>.000</td>
</tr>
<tr>
<td>Error (sphericity assumed)</td>
<td>202.96</td>
<td>810</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser correction</td>
<td>202.96</td>
<td>703.74</td>
<td>.29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Post hoc tests, using the Bonferroni corrections for multiple comparisons, revealed that the differences in means were significant between all pairs of evidence types with two exceptions; mean scores on DNA items did not differ significantly from scores on ballistics items ($p > .05$) and mean fingerprint item scores did not differ significantly from footwear evidence items ($p > .05$).

The hypothesis that DNA and fingerprint evidence would have the highest forensic awareness score was only partially confirmed. DNA evidence did indeed score
high on awareness however, so did ballistics evidence and these scores did not differ significantly from one another. Fingerprint evidence did not score significantly better than footwear identification techniques, although these two did score significantly better than GSR and glass/trace evidence, as predicted. Despite these apparent differences in knowledge about particular evidence types, it is not possible to directly compare participants’ scores between the evidence types due to the fact that the questions were not standardised across all evidence groups. It could be the case that the questions about some types of evidence were more difficult than other types, and therefore comparing responses between all items may be misleading.

3.4.5 Mobility and relevance as factors contributing to probative value

The evidence presented to participants in the second section of the questionnaire was manipulated on two factors, each of which had two levels; mobility of the evidence (low or high) and investigative relevance of the evidence (low or high). This manipulation resulted in a 2x2 design with four categories of evidence with varying relative strengths as highlighted in Figure 1 (p. 96).

Hypotheses:

i. Participants would rate the strongest evidence category (low mobility/high relevance) significantly higher than the other three categories.

ii. The weakest category of evidence would be rated significantly lowest in strength.
iii. The remaining two categories will not differ significantly in terms of strength ratings due to their ambiguity.

A factorial, within-subjects ANOVA was used to investigate whether there was a main effect of mobility and relevance, and whether there was a significant interaction between these independent variables.

Descriptive statistics for the mean strength ratings for each category of evidence are summarised in Table 7. The mean strength scores for each category follow the trend that would be expected from Figure 1, therefore it seems that participants are considering both mobility and relevance as important when rating strength of evidence.

Table 7

*Descriptive statistics for strength ratings by evidence category*

<table>
<thead>
<tr>
<th>Category of evidence</th>
<th>Mean strength rating</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>High mobility/Low relevance</td>
<td>7.49</td>
<td>1.69</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Low mobility/Low relevance</td>
<td>8.55</td>
<td>1.90</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>High mobility/High relevance</td>
<td>9.93</td>
<td>1.74</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Low mobility/High relevance</td>
<td>11.18</td>
<td>1.51</td>
<td>7</td>
<td>15</td>
</tr>
</tbody>
</table>

*Note.* All mean strength ratings differ significantly at $p < .05
The results of the factorial within-subjects ANOVA are summarised in Table 8.

The results indicated that there was a significant main effect of mobility of the evidence, $F(1, 162) = 135.69, p < .05, r = .46$ as well as of the level of investigative relevance, $F(1, 162) = 476.97, p < .05, r = .75$. No significant interaction between mobility and relevance was found, $F(1, 162) = 1.24, p = .27$. This was interpreted to mean that regardless of the level of mobility, high relevance evidence was rated stronger than low relevance evidence; and regardless of the degree of relevance, low mobility evidence was rated stronger than high mobility evidence (as illustrated in Figure 3). The larger effect size for the main effect of relevance supports the fact that this is the more salient feature of evidence when determining probative value.

Table 8

*Factorial, repeated measures analysis of variance summary table for strength ratings by evidence category*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>217.99</td>
<td>1</td>
<td>135.69</td>
<td>.000</td>
</tr>
<tr>
<td>Error (mobility)</td>
<td>260.26</td>
<td>162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance</td>
<td>1048.97</td>
<td>1</td>
<td>476.97</td>
<td>.000</td>
</tr>
<tr>
<td>Error (relevance)</td>
<td>356.28</td>
<td>162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobility * Relevance</td>
<td>1.47</td>
<td>1</td>
<td>1.24</td>
<td>.27</td>
</tr>
<tr>
<td>Error (mobility*relevance)</td>
<td>191.78</td>
<td>162</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These results only partially confirmed the hypotheses in that while the strongest evidence was rated significantly strongest and the weakest evidence was rated significantly weakest, participants also significantly differentiated between the middle two categories, which was not predicted.

### 3.4.6 Relationship between evidence strength and confidence

As well as rating the strengths of the various pieces of evidence, participants also indicated how confident they felt about their ratings on a four-point scale (with ‘entirely unsure’ coded a 1, and ‘extremely confident’ coded as 4). For each of the four categories of evidence strength there is a total confidence score for each
participant, which is the sum of the confidence ratings for the three items in each category.

**Hypothesis:** Participants will be most confident about their ratings of the highest and lowest strength categories of evidence, as these are quite obviously of a high and low quality respectively. The remaining two categories of evidence strength will have lower confidence scores as participants will be uncertain about these items due to their increased ambiguity.

The mean confidence ratings by evidence strength category are summarised in Table 9.

<table>
<thead>
<tr>
<th>Category of evidence</th>
<th>Mean confidence rating</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>High mobility/Low relevance</td>
<td>8.36\textsubscript{a}</td>
<td>1.58</td>
</tr>
<tr>
<td>Low mobility/Low relevance</td>
<td>8.85\textsubscript{b}</td>
<td>1.72</td>
</tr>
<tr>
<td>High mobility/High relevance</td>
<td>8.93\textsubscript{b}</td>
<td>1.45</td>
</tr>
<tr>
<td>Low mobility/High relevance</td>
<td>9.43\textsubscript{c}</td>
<td>1.41</td>
</tr>
</tbody>
</table>

*Note.* Means with different subscripts differ significantly at $p < .05$

A one-way, within-subjects ANOVA was conducted with confidence rating as the dependent variable and strength category of evidence as the independent
variable. This analysis of variance is summarised in Table 10. Mauchly’s test of sphericity revealed that this assumption had been violated by these data ($\chi^2 (5) = 24.88, p < .05$) therefore the degrees of freedom in the analysis were corrected using the Greenhouse-Geisser estimate of sphericity ($\varepsilon = .89$) (Field, 2009).

Table 10

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength category (sphericity assumed)</td>
<td>94.54</td>
<td>3</td>
<td>31.52</td>
<td>34.48</td>
<td>.000</td>
</tr>
<tr>
<td>Greenhouse-Geisser correction</td>
<td>94.54</td>
<td>2.70</td>
<td>35.04</td>
<td>34.48</td>
<td>.000</td>
</tr>
<tr>
<td>Error (sphericity assumed)</td>
<td>444.21</td>
<td>486</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser correction</td>
<td>444.21</td>
<td>437.07</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ANOVA indicated that overall there was a significant difference in the mean confidence ratings between the various strength categories of evidence, $F (2.70, 437.07) = 34.48, p < .05, \omega^2 = .21$. Post hoc comparisons using the Bonferroni correction found that there was a significant difference between all categories of evidence, with one exception; there was no significant difference in mean confidence between the ‘high mobility/high relevance’ and ‘low mobility/low relevance’ categories.
These analyses partially confirm the hypotheses as the strongest evidence category was indeed associated with the highest confidence scores as predicted, and the middle two categories of strength received significantly lower confidence scores which did not significantly differ from one another. What was not predicted was that the weakest category of evidence would be associated with significantly lowest confidence ratings overall.

3.4.7 Source of Forensic Knowledge and Evidence Evaluation

In order to determine whether there was any difference in the mean strength ratings for each category of evidence as a function of the source of participants’ forensic knowledge a multivariate analysis of variance (MANOVA) was conducted. This analysis allows a comparison of multiple means (strength ratings for four categories of evidence) for multiple levels of an independent variable (source of forensic knowledge) and is appropriate in situations where the dependent variables are moderately positively correlated with one another (Tabachnick & Fidell, 2007).

The results of the MANOVA indicated that there were no significant differences in the mean strength ratings for the various sources of forensic knowledge, (for all $F$ values, $p > .05$). As illustrated in Figure 4, the evaluations of strength for each evidence type follow a similar pattern regardless of the source of forensic knowledge reported by participants.
Figure 4. Strength ratings of evidence by category (mobility/relevance) and source of forensic knowledge.

A second MANOVA was used to investigate differences in mean confidence ratings (dependent variables) for the four evidence categories as a function of the reported source of forensic knowledge (independent variable). The results of this MANOVA revealed no significant differences in confidence ratings across the four evidence categories and sources of knowledge, (for all $F$ values, $p > .05$). Figure 5 illustrates the common patterns in confidence ratings across the evidence categories for each level of the independent variable. Although the graphical representation of these confidence ratings by source of forensic knowledge group appears to show some substantial differences between groups, the dependent variables are quite highly correlated with one another which results in a MANOVA with reduced power (Ramsey, 1982 as cited in Field, 2009).
Figure 5. Confidence ratings by evidence category (mobility/relevance) and source of forensic knowledge.

3.5 DISCUSSION

3.5.1 OVERALL FORENSIC AWARENESS

Almost half of respondents indicated that their knowledge of forensic science originates mainly from television and movie representations of criminal investigations. This is perhaps not surprising as forensic science is widely represented in fictional media and features as a central storyline in many popular, current television dramas (Mann, 2006). A number of participants indicated that the news media is their main source of information about forensic science, and a smaller proportion of participants reported that they have learned about forensic techniques in a more formal environment (e.g. school or through their career/profession). The only category of
knowledge source which was not selected by any participants was ‘personal involvement/experience in a court case’. In terms of where prior knowledge about forensic science comes from, and of particular relevance to juries, this means that none of the jury-eligible members in this current sample have had any personal experience with forensic science techniques in a legal setting. It is reasonable therefore to assume that much of the knowledge held by this sample of potential jurors originated in fictional representations of forensic science.

Exposure to mainly fictional sources of information about forensic science does seem to have some impact on people’s overall awareness of some of common forensic techniques, albeit not as large an effect as was hypothesised. Whereas educational level was not found to be related to overall forensic awareness scores in the current study, it was found that those who indicated that their knowledge comes from mainly television and movies scored lowest in terms of overall accuracy of their beliefs about forensic techniques (although only significantly lower than the highest category – crime books). There were, however, some beliefs which the vast majority of participants accurately held about forensic science, regardless of their source of knowledge. It was widely believed that fingerprints and DNA could be used to positively identify someone, which is a reasonable finding as this refers to the basis of these two very common and well established forensic techniques. What was less unanimously accurately believed were items referring to the specific conditions under which a fingerprint or DNA sample could be obtained and used by police.

Instances where a low proportion of participants held accurate beliefs were often the result of an overestimation in the abilities of the forensic science technique
in question. For example 49% of participants believed that fingerprints could be recovered from any type of surface, and 42% believed that trace amounts of glass will remain on a suspect’s clothing until they are washed. In both of these cases respondents’ beliefs are incorrect due to the fact that they are overly optimistic about the ability of forensic science to recover evidence using these techniques. It is possible that this tendency to believe that forensic science is more effective than it actually is in practice, may be explained by the content of media representation of these techniques. As described by Schweitzer and Saks (2007) such exaggerated beliefs about the capabilities of forensic science (as often seen in media representations) are thought to be one potential result of the so-called CSI effect, which may ultimately burden the prosecution of a case as these high expectations of forensic science cannot be met in reality.

There was also variation in the accuracy of beliefs held about different types of forensic evidence, which is to be expected if people are exposed to varying amounts of information about different evidence types, for example in the media. The participants in this study were most knowledgeable about DNA and ballistics evidence which may be due to the fact that regardless of your source of forensic knowledge these are likely to appear often in both fictional (e.g. tv and movies) and factual (e.g. news programmes, documentary-style programmes) sources of information. It should be noted however that participants were not accurate in their beliefs about all aspects of these evidence types, with very high accuracy on items concerning the basic use of these techniques, but lower accuracy on items referring to the successful recovery and practical value of evidence in these categories. This trend in these results seems to
echo the opinion of a judge in the United States who commented that “TV has taught jurors about DNA tests, but not enough about when to use them” (Mann, 2006, p. 166).

3.5.2 Evidence evaluation

The findings in the current study regarding the participants’ ability to evaluate the strength of various pieces of evidence were promising, as they do suggest that mobility and investigative relevance are important factors in the evaluation process. The fact that participants were not given any contextual information about the evidence in the evaluation task suggests that the manipulation of mobility and relevance were indeed responsible for the strength ratings obtained in this study. The fundamental importance of relevance above that of mobility was also demonstrated by participants’ ratings of the evidential strengths and the fact that there was no interaction between these two main effects. This finding indicates that these participants did understand the need for evidence to be indicative of guilt in order for it to be useful, regardless of its level of mobility.

By demonstrating that participants were sensitive to two important factors of evidence strength (mobility and relevance) in the absence of any contextual information and that they were not making strength judgments based solely on the type of evidence (e.g. DNA, fingerprints, footwear), the findings suggest that potential jurors are capable of understanding some of the intricacies which determine the probative value of forensic evidence. This seems contrary to some previous research which has suggested that mock jurors do not reason effectively about forensic evidence in courtroom situations (Nance & Morris, 2005; Faigman & Baglioni, 1988;
Goodman, 1992; Smith et al., 1996), however these previous studies have presented mock jurors with more peripheral information as part of the evidence evaluation task. The resulting biases which have been found to interfere with mock jurors’ decision making may therefore be attributed to this additional contextual information. Indeed even experts have been shown to evaluate forensic evidence differently when it is presented with contextual information (Dror, Charlton, & Peron, 2006). The current study has taken a step back from these previous studies and specifically shown that participants displayed the ability to recognise and understand correctly some of the features of physical evidence which contribute to the value of the evidence, namely mobility and relevance.

3.5.3 Confidence and evidence evaluation

In the current study, potential jurors reported being most confident when asked to evaluate the strength of high quality evidence which is to be expected as these items of evidence would quite obviously be useful during an investigation. The middle two categories of evidence strength resulted in significantly lower confidence ratings, which was presumed to be due to the increasing ambiguity concerning the usefulness of these items. These two findings are consistent with the manipulations made in the evidence examples and the participants’ abilities to correctly rank the strengths of each category of evidence.

The unexpected finding was that the evidence category which was associated with the lowest confidence levels was the weakest strength category of evidence. This is surprising as these items of evidence were presumed to be of obviously poor quality and therefore should have been associated with high levels of confidence in the
assignment of low strength scores. Participants in this study seemed to be able to recognise the fact that these pieces of evidence were of low quality and strength, yet they were not confident when making this assessment.

It is not clear from this analysis why participants were not confident in their rating of the weakest evidence, however one possible explanation may be related to previous findings which suggest that jurors have elevated expectations of forensic evidence (both quantity and quality) due to increasing interest in scientific developments in forensic science (Shelton et al., 2006). It could be the case that participants in this study were able to recognise comparatively weak evidence but they were not confident about rating forensic evidence of any kind as ‘weak’ due to elevated expectations about how useful forensic evidence should be in general in criminal cases. It does seem to be the case that participants in the current study did not rate even the weakest evidence as particularly low in strength, which may suggest that even though they correctly identified that some evidence was relatively weak there was still a belief that it was somewhat useful.

3.5.4 LIMITATIONS AND FUTURE RESEARCH

The study described in this chapter had a very specific objective in terms of investigating participants’ comprehension of forensic evidence and therefore much information which is normally presented to jurors was omitted in order to isolate certain aspects of the decision making process. The direct applicability of these findings to the actual behaviour of jurors in a courtroom setting is therefore limited. However, some interesting future research questions can be generated to further
address jurors’ understanding of forensic evidence and how this impacts on their decision making strategies.

Now that it has been established that mobility and relevance are likely to be important aspects of physical evidence when jurors assess the strength of evidence, the next chapter in this thesis aimed to identify the impact of introducing contextual information into the decision making process. For example, different types of evidence (alibi, eye-witness, expert testimony) were incorporated into the following studies to determine if this additional information interferes with the appropriate evaluation of forensic evidence by mock jurors. It could be the case that in isolation jurors can understand the importance of mobility and relevance but when faced with a multitude of other pieces of information, the integration of these into a coherent narrative (as described by the story model) and correct evaluation of forensic evidence may become more difficult.
CHAPTER FOUR: THE IMPACT OF CASE CONTEXT ON EVIDENCE EVALUATION

4.1 ABSTRACT

This study follows on from the previous chapter by investigating the impact of contextual information on the ability to correctly evaluate the probative value of forensic evidence. A vignette methodology was used in order to present participants with mock trial information which included forensic evidence of varying strengths using the same mobility and relevance manipulations as in the previous chapter. The results for the control condition replicated the findings in the previous study, and support the argument that participants were able to distinguish varying probative values when no case context is provided. The results of the remaining conditions suggested that contextual information does interfere with the evaluation of evidential strength, and in all cases strength estimates were inflated by the presence of case context. This effect was also observed in the probability of guilt estimates, although it did not seem to have an impact on verdict decisions. These results are discussed in relation to narrative models of juror decision making, and provide the empirical framework for the remainder of the thesis chapters.

4.2 INTRODUCTION

In the study described in the previous chapter jury-eligible participants evaluated the strength of various pieces of forensic evidence which had been manipulated on two levels; (i) the mobility of the evidence was varied resulting in evidence which could have been transported into the crime scene without requiring attendance at the scene (high mobility evidence) or evidence which could not be deposited at the scene without the person being present (low mobility evidence), (ii)
the investigative relevance of the evidence was either of high relevance (highly probable indication of guilt) or low relevance (not directly indicative of guilt).

The manipulation of these two features of evidence resulted in a 2x2 factorial, within-subjects design and the results indicated that participants were able to recognise the importance of these features and as a result they tended to rank the evidence items in the appropriate order of strength. The level of confidence for each strength rating was also recorded by participants and these results showed that the strongest evidence was associated with the highest confidence levels and the weakest evidence with the lowest confidence ratings. This suggests that participants were very confident that they could identify strong forensic evidence but significantly less confident about evaluating forensic evidence that was weak.

The findings of the previous study seem promising regarding juror decision making about forensic evidence, as they demonstrated that participants were able to distinguish between evidential items which differed in terms of mobility and relevance. However the evidence was presented to the participants in the absence of any case information or other evidence, which would normally be an important part of a juror’s decision making process. This chapter aims to follow up on these previous findings to determine whether the introduction of case context and other evidence affects the ability of participants to reason effectively about the strength of forensic evidence when mobility and relevance are varied.
4.3 Method

4.3.1 Participants

The same target population was required for this study as was used in the previous chapter; members of the public who are eligible for selection as jurors based on the basic criteria described by the Criminal Justice System for England and Wales (www.cjsonline.gov.uk). Consistent with the previous chapter, basic jury eligibility was determined by asking participants to indicate their age and whether they had ever been convicted of a criminal offence.

In total 160 people completed the online task, who were randomly allocated across the four conditions ($N = 40$ per condition). The final sample consisted of 113 female and 47 male respondents, aged between 18 and 67 years ($M_{age} = 34.77, SD = 12.10$). Similar to the sample in the previous chapter, the highest educational level reported by participants represented a good range, with 19% having completed high school, 66% with an undergraduate degree, and 15% having postgraduate qualifications.

4.3.2 Materials

The online materials presented to participants in this study were developed using the University of Leicester online content management system ‘Plone’. This system allows the creation of various presentation formats for interactive online tasks, which can then be published as open-access websites for distribution to participants. Plone also collects and stores the results which can then be exported into Microsoft Excel or SPSS for analysis.
In order to allow random allocation of participants to one of the four conditions in this study, a single main webpage was designed for the study which contained a welcome message and brief overview of the research objectives. Participants were given this web address for participation in the study, and by visiting this site they were then randomly forwarded to one of the four Plone webpages, each containing one of the experimental conditions. Each of these four Plone sites included a further introductory message and more detailed description of the research and instructions for completing the task, as well as the informed consent details.

4.3.3 Procedure

Control (no context) condition. Participants allocated to the control (or no case context) condition responded to an evidence evaluation task containing different material and format than the other three conditions. This material consisted of statements about various evidence types (which were similar to the evidence presented in the trial stimulus used in the other conditions) and participants were asked to estimate the relative strengths of each. This task was identical to the evidence evaluation task described in the previous chapter, and a list of the items is provided in Appendix B. Participants used the same five-point Likert scale to indicate how useful they perceived the evidence items to be in determining the guilt of a suspect, which was coded from irrelevant/useless evidence (1) to conclusive/definite evidence (5).

Case context conditions. In order to provide participants with contextual information about a criminal case, and the associated evidence, a case summary and excerpts from court testimony were developed with reference to a fictional murder
trial. In order to control for a variety of case variables, such as source of forensic evidence and seriousness of offence, one case scenario was used with only the strength of the forensic DNA evidence being varied across the three conditions. The DNA evidence testimony in the transcript followed the presentation format used in previous research in jury decision making (Koehler, 2001). The case information included testimony from an eyewitness (the victim’s neighbour), the arresting police officer, a forensic pathologist, a forensic laboratory analyst as well as closing summaries by the prosecution and defence lawyers.

Initially, participants were presented with a summary of the case which did not include any detail about evidence recovered from the scene or witnesses who testified in the trial. After reading this summary participants indicated how probable they thought it was that the defendant was guilty of the murder as described in the summary (expressed as a percentage between 0 and 100%). This measure provided a baseline measure of perception of guilt for each participant against which comparisons could be made with later measures in the study.

Following this summary, participants read through the testimony of a number of key witnesses in the murder trial. Appendix C contains a full record of the witness statements as presented. The material presented in the witness statements was identical in each of the three conditions, with the exception of the scenes of crime examination information which varied in terms of the mobility and relevance of the DNA evidence source recovered from the crime scene (colour coded in Appendix C). Although in the previous chapter there were four strength categories of evidence, the ‘low mobility / low relevance’ category was not included as a condition in this study.
due to the implausibility of such DNA evidence in the murder scenario used. Following each witness statement participants were asked to indicate how useful they felt the evidence was for determining the guilt of the defendant, using the same five-point Likert scale described in the control condition.

After all of the witness statements were read participants were once again asked to estimate the perceived probability that the suspect was guilty of the murder. This measure allows for a quantitative estimation of how probative the participants perceived all of the prosecution evidence presented at trial to be, as it can be compared to their initial estimate of guilt. The final piece of trial information presented at this stage was a summary of the case presented by the prosecution and then by the defence lawyers. These summaries were presented as brief narratives which accounted for the evidence presented, but either supported the guilt (prosecution summary) or innocence (defence summary) of the accused. Following these summaries participants were asked to estimate the probability of guilt one final time, as well as make a verdict decision (guilty or not-guilty of murder).

4.4 RESULTS

4.4.1 SOURCE OF FORENSIC KNOWLEDGE

Participants were asked to indicate where they believed most of their knowledge of forensic evidence comes from, and Figure 6 illustrates the frequency of responses to this question. The distribution of these responses is very similar to the responses from participants in the previous study.
To investigate whether the reported source of forensic knowledge had an impact on the ratings of the evidence presented in the case scenario, a MANOVA was conducted with evidence strength ratings as the dependent variables (eye-witness, arresting officer, pathologist, and DNA) and source of forensic knowledge as the independent variable. The results confirmed that there was no difference in the mean strength ratings of the various types of evidence between participants who reported various sources of forensic knowledge (for all $F$ values, $p > .05$). This finding is consistent with the results from the previous study.

4.4.2 Tests of evidence manipulation across conditions

A series of analyses were conducted to determine whether the manipulation of the independent variable (strength of the DNA evidence) was successful in the three case context scenarios. Firstly, the measures of prior probability of guilt should not differ significantly between the three conditions due to the fact that this estimate is
made before any evidence is presented to participants. A one-way, between-subjects ANOVA was used to compare the mean prior probability of guilt between the three conditions, and the results confirmed that there was no significant difference in this measure between conditions ($F(2, 177) = .05, p = .95$). The mean prior probability estimates were 49.40 (strongest evidence), 50.42 (moderate evidence), and 50.75 (weakest evidence) and one-sample t-tests confirmed that these values did not differ significantly from a test score of 50 which indicates that prior to hearing any evidence participants perceived the case summary to favour neither the prosecution nor the defence.

It was also necessary to confirm that there was no difference in the mean strength ratings for the non-DNA evidence in the scenario between the conditions as this information was the same for all participants. A MANOVA was used to compare the mean strength ratings for the eye-witness, arresting officer, and forensic pathologist testimony between the three conditions. The results indicated that there was no significant difference in the mean strength ratings for any of these categories of evidence (for all $F$ values, $p > .05$).

### 4.4.3 Evaluation of the DNA Evidence

**Hypothesis:** Participants in the control condition will correctly identify the relative strengths of the forensic evidence, similar to the findings presented in Chapter three.

In the no-context control condition participants used a five point Likert scale to indicate the perceived probative value of various statements describing forensic
evidence of varying strengths. Table 11 summarises the mean strengths assigned to each category of evidence in this condition. A one-way, within-subjects ANOVA was found to be significant \( (p < .05) \) and post hoc tests confirmed that the mean strength ratings were significantly different between each category of evidence strength in the control condition, \( F(2, 78) = 64.37, p < .001 \). This result replicates the findings from the previous study and confirms that in the absence of case context participants assessed the weight of evidence in the correct relative direction as defined by the mobility/relevance manipulation.

Table 11

Mean strength ratings for DNA evidence in all conditions

<table>
<thead>
<tr>
<th>Category of evidence</th>
<th>Mean strength rating (no-context)</th>
<th>SD</th>
<th>Mean strength rating (with case context)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak evidence</td>
<td>2.52(_a)</td>
<td>0.67</td>
<td>3.10(_d)</td>
<td>1.00</td>
</tr>
<tr>
<td>Moderate evidence</td>
<td>3.28(_b)</td>
<td>0.69</td>
<td>3.85(_e)</td>
<td>0.74</td>
</tr>
<tr>
<td>Strong evidence</td>
<td>3.73(_c)</td>
<td>0.62</td>
<td>3.88(_e)</td>
<td>0.69</td>
</tr>
</tbody>
</table>

*Note. Column means with different subscripts differ significantly at \( p < .05 \) \((N = 160)\)*
**Hypothesis**: Participants in the case context conditions would find it more difficult to distinguish between the evidence strengths due to the influence of the contextual information presented at the trial.

In the remaining three conditions, all of which included identical case context, participants rated the strength of DNA evidence (which was manipulated across conditions) and therefore a one-way, between-subjects ANOVA was used to compare the strength ratings of the DNA evidence between conditions. These results are also summarised in Table 11.

The results of the ANOVA indicated that there was a significant difference in the strength ratings, \( F(2, 119) = 11.20, p < .001 \) and post hoc tests using the Bonferroni correction confirmed that the DNA evidence in the weakest evidence condition was rated significantly lowest, but there was no significant difference in strength ratings between moderate and strong DNA evidence conditions. These findings partially confirmed the hypothesis, as case context seems to have made it more difficult for participants to distinguish between moderate and strong evidence.

The impact of presenting case context on the mean strength ratings of the various strengths of DNA evidence is illustrated in Figure 7.
Figure 7. Effect of case context on mean strength ratings for all DNA evidence strength categories ($N = 160$).

All categories of evidence were rated as stronger on average when accompanied by case context (compared to the control condition) and this difference is most prominent when the evidence is of a weak or moderate standard. It is also worth noting that the only condition in which evidence was rated as ‘weak’ (e.g. below a mid-scale score of 3, indicated with a dotted line in Figure 7) is in the weakest evidence category, but only when no case context is presented.

4.4.4 Probability of guilt estimates

Participants in all of the case context conditions were required to estimate the probability that the defendant was guilty of the murder described in the scenario.

These estimates were expressed as percentages, and were measured at three stages in the scenario resulting in a within-subjects measure: The prior probability (after a brief summary, but before any evidence was presented), post-evidence probability
(immediately after all of the evidence/testimony was heard), and the final probability (after the prosecution and defence summaries).

**Hypothesis:** Participants who were in the weakest evidence condition will estimate the lowest probability that the defendant committed the crime, and those in the strongest evidence condition will perceive the probability of guilt as significantly highest.

As previously described, the mean prior probability of guilt estimates were compared between conditions and no significant differences were found. Therefore, in order to compare the remaining two probability of guilt estimates between the three case context conditions, a MANOVA was used. Table 12 summarises the mean probability of guilt estimates for each condition and Figure 8 represents these relationships visually.
Table 12
Mean probability of guilt estimations for each DNA evidence strength condition

<table>
<thead>
<tr>
<th>Category of DNA evidence</th>
<th>Prior probability of guilt</th>
<th>Post-evidence probability of guilt</th>
<th>Final probability of guilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak evidence</td>
<td>49.40&lt;sub&gt;a&lt;/sub&gt;</td>
<td>55.28&lt;sub&gt;b&lt;/sub&gt;</td>
<td>52.55&lt;sub&gt;d&lt;/sub&gt;</td>
</tr>
<tr>
<td>Moderate evidence</td>
<td>50.42&lt;sub&gt;a&lt;/sub&gt;</td>
<td>67.95&lt;sub&gt;c&lt;/sub&gt;</td>
<td>64.48&lt;sub&gt;e&lt;/sub&gt;</td>
</tr>
<tr>
<td>Strong evidence</td>
<td>50.75&lt;sub&gt;a&lt;/sub&gt;</td>
<td>68.48&lt;sub&gt;c&lt;/sub&gt;</td>
<td>68.50&lt;sub&gt;e&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. Column means with different subscripts differ significantly at $p < .05$ ($N = 120$)

Figure 8. Mean probability of guilt estimates by DNA evidence case context condition ($N = 120$)
Using Pillai’s Trace, it was confirmed that there was an overall effect of strength of evidence condition on probability of guilt estimates, $F(4, 234) = 3.43, p < .05$. However separate univariate ANOVAs for the measures of post-evidence ($F(2, 117) = 4.93, p < .05, r = .28$), and final probability of guilt ($F(2, 117) = 6.16, p < .05, r = .31$) revealed that the only significant difference was between the weakest evidence and the other two evidence strength conditions. This partially confirmed the hypothesis, however it was not predicted that the moderate evidence condition would result in similar probability of guilt estimates as the strong evidence condition.

4.4.5 Verdict decisions

At the end of the online task, participants in the case context conditions were asked to indicate whether they would vote guilty or not-guilty given the evidence they had received in their condition. Table 13 summarises the frequencies of each verdict for each of the three conditions.

**Hypothesis:** Only participants in the strong evidence condition would vote guilty, as the other two evidence conditions did not contain enough evidence to support a guilty verdict.
Table 13

*Frequency of guilty and not-guilty verdicts by evidence condition*

<table>
<thead>
<tr>
<th>Category of DNA evidence</th>
<th>Guilty verdict</th>
<th>Not-guilty verdict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak evidence</td>
<td>6 (15%)</td>
<td>34 (85%)</td>
</tr>
<tr>
<td>Moderate evidence</td>
<td>9 (23%)</td>
<td>31 (77%)</td>
</tr>
<tr>
<td>Strong evidence</td>
<td>16 (40%)</td>
<td>24 (60%)</td>
</tr>
<tr>
<td>Total</td>
<td>31 (26%)</td>
<td>89 (74%)</td>
</tr>
</tbody>
</table>

*Note. N = 120*

There was a significant overall association between DNA evidence strength condition and verdict decision, $\chi^2 (2) = 6.87, \ p < .05$, however further analyses confirmed that only the strongest evidence category differed significantly in verdict decision from the two other evidence conditions (which did not differ significantly from one another). This partially confirms the hypothesis, however a fair number of participants in the weak and moderate evidence conditions decided on a guilty verdict (which was not predicted).

4.5 **Discussion**

Similar to the findings in the previous chapter, these results suggest that participants were able to correctly evaluate the relative strengths of DNA evidence when information was presented in the absence of any contextual information, as was the case in the control condition in the current study. The replication of these findings lends support for the use of mobility and relevance as salient factors in the
determination of probative value, as suggested by previous research (Bond, 2007; White, 2004).

The inclusion of case information in the other three conditions seems to have impacted on participants’ evaluations of the probative value of the forensic evidence presented in the trial. In particular, participants rated moderately strong DNA evidence as having the same probative value as the strongest DNA evidence. This is despite the fact that in the absence of context the differentiation between these evidence strengths was obvious to participants. Another interesting finding in the context conditions is the fact that all strengths of DNA evidence were rated as stronger compared to similar evidence presented in the control condition. Of particular concern is the fact that on average participants in the case context conditions did not rate any of the DNA evidence as ‘weak’ (e.g. below a mid-point score of 3) regardless of the mobility/relevance manipulation. In other words, even the weakest category of DNA evidence was rated as stronger than ‘moderate’ when accompanied by case context.

These findings shed some light on the likely impact of information presented in a courtroom format on a juror’s ability to objectively evaluate the probative value of evidence. It appears that the incorporation of evidence into a narrative, as described by the story model (Pennington & Hastie, 1986), resulted in evidence being perceived as having more probative value than would otherwise be attributed to it. This effect of context was particularly prevalent when evidence was ambiguous, as in the case of the moderate and weak DNA evidence category. This could be explained by the Liberation Hypothesis which proposes that when trial evidence is ambiguous (or not
clearly favouring either guilt or innocence), jurors may rely more heavily on their beliefs, assumptions or attitudes about various aspects of the judicial process (Devine, Buddenbaum, Houp, Studebaker, & Stolle, 2009).

The observed difficulty in distinguishing between moderate and strong DNA evidence in the case context conditions was also evident in the participants’ probability of guilt estimates. In all conditions the prior probability of guilt estimates were indistinguishable at approximately 50%. However, probability of guilt estimates made after the evidence and the prosecution/defence summaries were presented showed that moderate and strong DNA evidence resulted in similar high probabilities while the weak DNA evidence resulted in probabilities which remained around 50%. This is interpreted as support for the argument that the case context and the associated narrative interfered with participants’ ability to distinguish the difference between moderate and strong DNA evidence.

It does appear, however, that despite high strength ratings and estimates of the defendant’s guilt in the moderate evidence strength category, the decision to vote guilty was significantly more prominent in only the strongest DNA evidence category. This finding is important as it raises the question of whether the observed difficulties in assessing the relative strength of evidence in a case scenario impacts on the jurors’ ultimate verdict decision. It is not entirely clear from this research why the verdict decision results did not mirror the probative value evaluations and the probability of guilt estimates. Although there were more guilty verdicts in the moderate evidence condition (compared to the weak evidence condition), this difference was not statistically significant.
The probability of guilt estimates for each condition do suggest that in the moderate strength DNA evidence condition the presentation of prosecution and defence summaries resulted in a slight decrease in probability of guilt (although this decrease was not statistically significant). This presentation of competing narratives at the end of the trial to explain the DNA evidence may have reduced the probability of guilt by introducing alternative explanations for the ambiguous forensic evidence which participants may not have considered previously. This may partly explain the slight decrease in final probability of guilt estimates, and the increase in not-guilty verdicts when compared with responses in the strong DNA evidence condition. The consideration of alternative explanations for the presence of ambiguous forensic evidence is an important issue which has been raised previously in the literature in relation to the impact of technological advances on courtroom decision making (Duncan, 2008; Gallop & Stockdale, 2004).

Having demonstrated an effect of trial context on estimates of probative value of forensic evidence, the following chapters in this thesis apply an attitudes-behaviour paradigm to investigate the existence of pre-trial beliefs and attitudes which may account for the observed variability in participants’ perceptions of forensic evidence strength.
5.1 ABSTRACT

In an attempt to explain the high levels of disagreement between participants concerning the probative value of weak DNA evidence when accompanied by case context, a measure of attitudes about forensic science evidence was developed. In keeping with narrative models of juror decision making, the rationale for this study was the important role that prior beliefs play in decision making about ambiguous information. An initial scale of 31 items was developed with reference to the CSI Effect literature, which provided a conceptual framework for the content of the items. After item analyses on 21 items (excluding filler items), the final version of the scale consisted of 10 items representing pre-trial beliefs about forensic science evidence. These 10 items were then analysed using Principal Components Analysis (PCA), and the results indicated that the items loaded on two components. These two clusters of attitudes corresponded with the pro-prosecution and pro-defence beliefs which are hypothesised to be problematic by the CSI Effect literature.

5.2 INTRODUCTION

The previous chapters described two studies which have demonstrated that jury eligible members of the public are able to distinguish between strong, moderate and weak forensic evidence in the absence of contextual information. However, this ability to recognise the objective strength of forensic evidence may be diminished when the evidence is framed in the context of a criminal trial. Although the responses from participants showed some significant effects between evidence strength conditions, the perception of varying strengths of DNA evidence when accompanied
by case context was not uniform across participants within each condition. This is summarised in Table 14 below (with the percentages from the no-case-context condition in brackets).

Table 14

*Percentage of participants’ strength ratings for DNA evidence in the case context conditions*

<table>
<thead>
<tr>
<th></th>
<th>Rated ‘strong’</th>
<th>Rated ‘moderate’</th>
<th>Rated ‘weak’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongest evidence</td>
<td>80% (82.5%)</td>
<td>15% (15%)</td>
<td>5% (2.5%)</td>
</tr>
<tr>
<td>Moderate evidence</td>
<td>75% (32.5%)</td>
<td>20% (40%)</td>
<td>5% (27.5%)</td>
</tr>
<tr>
<td>Weakest evidence</td>
<td>37.5% (2.5%)</td>
<td>32.5% (37.5%)</td>
<td>30% (60%)</td>
</tr>
</tbody>
</table>

*Note.* The percentages in brackets are from the control (no case context) condition (*N* = 160)

From Table 14 it is clear that, when accompanied by case context, strong and moderate DNA evidence were rated by participants in a similar pattern. This is contrary to the response patterns when evidence is presented without any case context (see figures in brackets). Furthermore, case context affected the perceived probative value of the weakest DNA evidence such that without context the data were as expected (based on the previous results), but with context there was substantial disagreement among participants regarding the ratings of probative value.
There are two particularly interesting findings here, both of which can be discussed within the parameters of the story model of jury decision making (Pennington & Hastie, 1986; 1988). The story model would predict that the organisation of trial information into a narrative representation of the events leading up to the crime would have some impact on how the evidence is interpreted. This is partially due to the fact that each individual item of evidence is evaluated with reference to the rest of the evidence and the case as a whole, and is therefore not considered in isolation (Pennington & Hastie, 1988). This may be an explanation for the difference between the rated strength of moderately probative evidence in the absence of case context (Chapter 3) and when presented as a case with accompanying evidence (Chapter 4). However, this explanation does not account for the lack of consensus across participants regarding the weakest DNA evidence.

Another important aspect of juror decision making, as posited by the story model, is the impact of individual juror beliefs and biases which every person brings to the courtroom (Pennington & Hastie, 1988). This background knowledge helps jurors make sense of large amounts of information and determine the most likely, reasonable story to explain the evidence presented during a trial. The interactionist view of juror decision making predicts that individual differences will have the most influence when the situational cues are ambiguous or weak, and this appears to be the case with the greatest lack of consensus in the weak evidence category in Chapter 4.

If the prior knowledge or attitudes operating in cases of weak or ambiguous forensic evidence could be isolated and measured, it may be possible to predict how a participant would perceive and evaluate ambiguous evidence, and possibly their
resulting verdict preference. It does not seem appropriate however to use existing scales which measure general juror bias, for example the JBS (Kassin & Wrightsman, 1983) or LAQ (Boehm, 1968), for two reasons. Firstly, these scales do not make any reference to beliefs about forensic science evidence as these bias measures were developed prior to the availability of these types of scientific evidence. Secondly, if the goal is to specifically measure juror pre-trial attitudes about forensic science evidence, then generalised measures of attitudes toward the criminal justice system will likely not be specific enough to predict decisions about certain types of evidence. Therefore the aim of the study described in this chapter is to develop a measure of pre-trial juror attitudes about forensic evidence, and test whether such a measure predicts the decisions made about weak forensic evidence in mock trial situations.

5.3 Scale Development

As this is the first measure designed to identify juror pre-trial attitudes about forensic evidence specifically, the initial pool of items had to be developed with consideration of the theoretical basis for the potential existence of this particular bias. This theory comes largely from the CSI effect literature (see Chapter one for a review), and anecdotal support for the existence of a juror bias concerning forensic science. Although the existence of this bias has yet to be empirically confirmed in the literature, authors who support the possibility of this effect describe the impact of such a bias in two ways. Firstly, there is thought to be one manifestation of the CSI effect which burdens the prosecution in a criminal case, as jurors may have unrealistically high expectations of forensic science evidence in all types of trial (Podlas, 2006). It is thought that these unrealistic expectations result in the
prosecution not being able to satisfy jurors, and it follows that jurors may be unwilling to convict in cases where no forensic evidence is presented (Cole & Dioso-Villa, 2007). The second form that the CSI effect might take would have an opposite effect and place an increased burden on the defence. This occurs when jurors place too much faith in forensic science testimony and evidence, and therefore award forensic evidence more weight than it deserves (Kim, Barak, & Shelton, 2009). This may result in very tenuous or ambiguous forensic evidence leading to convictions simply because jurors are ‘blinded by the science’.

An initial pool of 31 items was generated for the first version of the scale. The items reflected beliefs that jurors have been anecdotally reported to hold (within the CSI effect literature), which are thought to be potentially problematic in cases where forensic evidence is presented. Also included were four items which were not directly related to the potentially problematic beliefs, in that they referred to general attitudes about jury service. An example of this type of item is “I look forward to being a juror one day, as it would be interesting and exciting”.

In addition, six items were included which were ‘filler’ items taken from the original JBS (Kassin & Wrightsman, 1983) and LAQ (Boehm, 1968) scales, and were included to make the purpose of this current measure less obvious to participants. It was thought that if participants were aware that the scale was measuring something similar to the CSI effect (which has appeared in recent media reports) there may be a social desirability effect in the responses. The inclusion of filler items made the scale appear to be a more general measure of beliefs about the criminal justice system, rather than focusing only on forensic science specifically. Participants responded to
the items by indicating their level of agreement with each using a 5-point Likert scale ranging from ‘strongly disagree’ (coded as 1) to ‘strongly agree’ (coded as 5).

Appendix D contains the full list of initial scale items used.

5.4 Reliability and Item Selection for Initial Scale

5.4.1 Pilot Study

The initial 31 item scale was piloted on psychology postgraduate students at the University of Leicester (N = 25). These participants completed the scale online and were asked to give feedback on the clarity of the items and the overall features of the scale. Comments relating to the wording of some of the items were taken into account, and in particular some of the items were amended to reduce perceived ambiguity. It was also observed in this pilot study that a high proportion of participants selected the neutral (“neither agree or disagree”) response category. As a result it was difficult to determine what people’s opinions were on some of the items.

The use of a neutral response category is important for scale items when there is the possibility that some participants may not understand the topic of the item (Loewenthal, 2001). However for the purposes of the current subject matter it seemed unlikely that respondents would not be able to express an opinion for each item. Therefore the response categories for the scale were adjusted, resulting in a response scale with no entirely neutral option. It was hoped that by using this new response scale participants would not be able to avoid making a decision concerning their opinion of the items (Howitt & Cramer, 2005a). Although some authors argue that removing the neutral response choice results in less accurate data by forcing participants to make a choice at random when they do not have an opinion (Field,
2009), it seemed reasonable to assume that the vast majority of respondents will have some opinion of the subject matter investigated by this scale. This revised version of the scale was then used for the further analyses described in the following sections.

5.4.2 Participants

Participants were asked to confirm that they met the basic jury eligibility criteria described by the Criminal Justice System for England and Wales (www.cjonline.gov.uk). Similar to the first two studies, basic jury eligibility was determined by asking participants to indicate their age and whether they had ever been convicted of a criminal offence.

In total 219 eligible participants completed the online version of the scale. The final sample consisted of 155 female and 64 male respondents, aged between 18 and 69 years ($M_{age} = 30.66, SD = 11.99$). The distribution of highest educational level of participants was similar to the previous studies with 21% having completed high school, 64% with undergraduate degrees, and 14% with postgraduate qualifications.

5.4.3 Materials

The online version of the scale was developed using the University of Leicester content management system, ‘Plone’. This allowed the scale to be presented and completed online, and the results tabulated in Excel which could then be exported to SPSS for analysis. The website consisted of an initial page which described the purpose of the research as a questionnaire measuring general beliefs and attitudes about the criminal justice system. After completing the informed consent page, participants were able to complete the 31 items which were presented in a fixed,
randomised order (full list available in Appendix D). At the end of the online task participants were thanked for their participation and encouraged to forward the website link on to other people for completion.

5.5 Results

The following analyses for scale reliability and item characteristics were carried out on 21 scale items, representing problematic beliefs and attitudes as proposed in the literature, and excluding the filler items and neutral items.

5.5.1 Item characteristics

There were two initial criteria against which all of the scale items were evaluated; items which did not result in a full range of responses were considered for exclusion, as were any items with poor item-total correlations (Field, 2009; Howitt & Cramer, 2005a; Loewenthal, 2001). All of the scale items elicited the full range of possible responses on the scale used by participants, indicating a wide range of opinions about the items, and therefore no items were excluded from the scale based on this criterion.

The scale responses were coded such that higher scores indicated stronger agreement with each statement, and the total scale score was calculated by summing a participant’s responses to all 21 items. The item-total correlations (correlation between the item score and the total scale score minus that item) were calculated for each of the 21 items, and items with correlations less than .3 were excluded from the final version of the scale (Field, 2009; Loewenthal, 2001). The inter-item correlation matrix was also examined for any items which did not correlate sufficiently with the
other items. The final version of the scale consisted of ten items with sufficiently high inter-item and item-total correlation scores. These items and the corresponding item-total correlations are summarised in Table 15.

Cronbach’s alpha was used as a measure of reliability for this reduced ten item scale, and was found to be in the acceptable range, $\alpha = .78$ (Field, 2009).
Table 15

*Final scale items with descriptive statistics and item-total correlations*

<table>
<thead>
<tr>
<th>Item description</th>
<th>Mean score</th>
<th>SD</th>
<th>Item-total correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every crime can be solved with forensic science</td>
<td>2.22</td>
<td>.75</td>
<td>.54</td>
</tr>
<tr>
<td>Every criminal leaves trace evidence at every scene</td>
<td>2.66</td>
<td>.77</td>
<td>.46</td>
</tr>
<tr>
<td>Forensic evidence is enough to convict</td>
<td>2.20</td>
<td>.66</td>
<td>.38</td>
</tr>
<tr>
<td>Forensics always IDs the guilty person</td>
<td>2.39</td>
<td>.72</td>
<td>.50</td>
</tr>
<tr>
<td>Forensic always provides a conclusive answer</td>
<td>2.41</td>
<td>.76</td>
<td>.53</td>
</tr>
<tr>
<td>Science is the most reliable way to ID perpetrators</td>
<td>2.77</td>
<td>.56</td>
<td>.34</td>
</tr>
<tr>
<td>No forensics means investigators did not look hard enough</td>
<td>2.21</td>
<td>.63</td>
<td>.43</td>
</tr>
<tr>
<td>If no forensics, the jury should not convict</td>
<td>2.08</td>
<td>.60</td>
<td>.40</td>
</tr>
<tr>
<td>Police should not charge someone without forensics</td>
<td>2.23</td>
<td>.69</td>
<td>.55</td>
</tr>
<tr>
<td>If no forensics recovered the defendant is probably innocent</td>
<td>1.89</td>
<td>.57</td>
<td>.37</td>
</tr>
</tbody>
</table>

*Note.* All correlations significant at $p < .05$; This table presents abbreviated versions of the scale items. For the exact wording as presented to participants, see Appendix D. $N = 219$. 
5.5.2 Principal components analysis

In order to investigate the underlying constructs which are potentially measured by this scale, principal components analysis (PCA) was carried out. The goal of PCA is to analyse patterns in the correlation matrix of the scale items and reduce the data by defining the underlying components which represent linear combinations of observed variables (Tabachnick & Fidell, 2001). This approach was exploratory in this research due to the fact that this is a novel scale and there is no robust empirical research to suggest that it was testing any particular hypothesis. However the items were developed using reference to the juror bias proposed by the CSI Effect literature and therefore the use of PCA results in an empirical confirmation of the scale’s construct validity (Lecci & Myers, 2002).

The PCA was run on the ten item scale with oblique rotation (direct Oblimin method in SPSS). Oblique rotation was selected as it is appropriate when it is reasonable to assume that the underlying factors might correlate with one another (Field, 2009), however the PCA was also conducted with orthogonal rotation (varimax method) which produced the same component structure. The results of the oblique rotation are reported here, as the theoretical basis on which the scale was developed would suggest that the components are not likely to be completely independent of one another. The adequacy of the sampling size was confirmed by the Kaiser-Meyer-Olkin measure which was in the ‘acceptable’ range, KMO = .81 (Field, 2009). The appropriateness of PCA for these data, based on the correlations between the items, was confirmed with Bartlett’s test of sphericity, \( \chi^2 (45) = 463.58, p < .001 \).
Using the criteria of retaining components with eigenvalues greater than 1 (Kaiser, cited in Field, 2009) two components emerged which explained 47.99% of the variance in responses. Although a third component obtained an eigenvalue approaching 1 (.94), inspection of the scree plot confirmed that the extraction of two components appeared to be appropriate (Howitt & Cramer, 2005b), and therefore only the first two components were retained from the analysis. In addition to using Kaiser’s criteria and the scree plot as guides for retaining factors, a parallel analysis was also conducted in order to determine how many factors were contributing significantly to the model. Parallel analysis compares the eigenvalues obtained to those which would occur for randomly generated data with the same parameters as the study data (Loehlin, 2004; Hayton, Allen, & Scarpello, 2004). Only eigenvalues greater than those resulting from the random data generation are retained and the results of the current parallel analysis supported the retention of the first two factors. Table 16 displays the factor loadings after rotation using the pattern and structure matrices with oblique rotation (Tabachnick & Fidell, 2001; Pallant, 2007).

When interpreting the factor loadings in Table 16, Tabachnick and Fidell (2001) suggest only interpreting loadings which are greater than .32, which is the case for all of the item loadings. All factor loadings are in the ‘good to excellent’ range as outlined by Tabachnick and Fidell. The variables that loaded onto component one all seem to correspond with beliefs about forensic evidence which might burden the defence in a trial, as these statements reflect an unrealistic amount of faith in the capability of forensic analyses. Items which loaded onto component two appear to be beliefs
associated with unrealistic expectations for the presence of forensic science evidence, which could potentially burden the prosecution in a trial.

The item ‘every criminal leaves some trace evidence behind at every crime scene’, was originally thought by the author to be a belief that might contribute to a burden on the defence, however it was a somewhat ambiguous item compared with some of the other statements. This item loaded onto component two, and given the factor structure revealed in this analysis more consideration was given to this item, and it was concluded that this belief could be conceived as one that might burden the prosecution, as it would contribute to increased expectations for the presence of forensic evidence in every criminal case.
Table 16

*Summary of principal components analysis for 10-item bias scale*

<table>
<thead>
<tr>
<th>Item</th>
<th>Pattern coefficients</th>
<th>Structure coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Component 1</td>
<td>Component 2</td>
</tr>
<tr>
<td>Forensics always IDs the guilty person</td>
<td>.81</td>
<td>.79</td>
</tr>
<tr>
<td>Forensic evidence is enough to convict</td>
<td>.66</td>
<td>.64</td>
</tr>
<tr>
<td>Science is the most reliable way to ID perpetrators</td>
<td>.64</td>
<td>.61</td>
</tr>
<tr>
<td>Forensic always provides a conclusive answer</td>
<td>.63</td>
<td>.17</td>
</tr>
<tr>
<td>Every crime can be solved with forensic science</td>
<td>.58</td>
<td>.23</td>
</tr>
<tr>
<td>If no forensics, the jury should not convict</td>
<td>-.13</td>
<td>.79</td>
</tr>
<tr>
<td>No forensics means investigators did not look hard enough</td>
<td>.71</td>
<td>.23</td>
</tr>
<tr>
<td>If no forensics recovered the defendant is probably innocent</td>
<td>.67</td>
<td>.18</td>
</tr>
<tr>
<td>Police should not charge someone without forensics</td>
<td>.20</td>
<td>.65</td>
</tr>
<tr>
<td>Every criminal leaves trace evidence at every scene</td>
<td>.27</td>
<td>.46</td>
</tr>
<tr>
<td>Eigenvalues (before rotation)</td>
<td>3.41</td>
<td>1.39</td>
</tr>
<tr>
<td>Cronbach’s α</td>
<td>.72</td>
<td>.70</td>
</tr>
</tbody>
</table>

*Note: N = 219*
5.6 DISCUSSION

This preliminary analysis suggested that the ten items selected based on reliability and item correlation criteria may be useful for measuring pre-trial bias about forensic evidence. In particular the PCA analysis described in this chapter lends support to the existence of two distinct clusters of attitudes about forensic science. These two components identified by the PCA seem to be conceptually similar to the two main types of bias described in the *CSI effect* literature. The items which comprised the first component are beliefs which exaggerate the capabilities of forensic science, which are described in the literature as potentially burdening the defence in criminal cases (Cole & Dioso-Villa, 2007). The second component consisted of items which reflect unrealistic expectations for the presentation of forensic evidence in a criminal case. These expectations are not always able to be met by the prosecution, and therefore these beliefs are thought to burden the prosecution (Thomas, 2006).

Based on the results of the PCA described in this chapter, there appears to be some support for the existence of the attitudes hypothesised to be problematic in the *CSI Effect* literature, however this chapter has not investigated the impact that these beliefs may have on the decision making process. In order to further validate these findings, to compare scores on this 10-item scale (henceforth referred to as the Forensic Evidence Evaluation Bias Scale – FEEBS) with existing measures of juror bias, and to determine whether these attitudes predict differences in juror decision making, a follow up study was conducted and is described in the next chapter.
CHAPTER SIX: VALIDATION OF THE FEEBS WITH A MURDER TRIAL SCENARIO

6.1 ABSTRACT

In the previous chapter the Forensic Evidence Evaluation Bias Scale (FEEBS) was developed and tested and the results of a Principal Components Analysis (PCA) suggested that two distinct constructs were being measured, corresponding to pro-prosecution and pro-defence attitudes about forensic evidence. In the current study, scores on these two subscales were compared with other existing juror bias measures (Juror Bias Scale and Belief in a Just World) and in a mock juror decision making task only the pro-prosecution subscale of the FEEBS predicted the perceived strength of weak forensic DNA evidence. A partial mediation model is presented which explains the relationship between these pro-prosecution attitudes and verdict preferences. The implications of this potential juror bias are discussed in the context of real juries, the CSI Effect and peremptory challenges.

6.2 INTRODUCTION

The aim of this study was to further test the usefulness of the 10-item scale developed in the previous chapter, which is referred to as the Forensic Evidence Evaluation Bias Scale (FEEBS). In the previous study a principal components analysis (PCA) revealed that the responses to this new scale demonstrated patterns in correlations between items, such that two distinct components were identified. Each of these components contained five scale items and interpretation of these components revealed them to be conceptually similar to a pro-prosecution and pro-defence bias towards forensic evidence, as hypothesised in the CSI effect literature (Cole & Dioso-Villa, 2007).
In this chapter the 10-item FEEBS was administered along with the Juror Bias Scale (JBS; Kassin & Wrightsman, 1983) and the General Belief in a Just World Scale (GBJW; Dalbert, Montada, & Schmitt, 1987) in order to investigate how these measures relate to one another. The JBS and GBJW scales were selected for comparisons in this study due to the fact that these attitude measures have previously been used widely in juror decision making research, and in the case of the JBS it is still widely cited in juror bias literature (see Chapter one). In addition to these relationships, the effectiveness of these pre-trial attitude measures for predicting mock juror decision making in the context of a fictional murder trial was also investigated.

6.3 METHOD

6.3.1 PARTICIPANTS

Undergraduate psychology students at the University of Leicester participated in this study in exchange for course credit, and were recruited using the psychology department’s Experimental Participation Requirement (EPR) system. As in the previous studies, participants met the basic eligibility requirement for jury service in England and Wales (www.cjsonline.gov.uk). In total 159 jury eligible students participated in the current study. The final sample consisted of 140 female and 19 male participants, aged between 18 and 46 years ($M_{age} = 20.16, SD = 3.90$).

6.3.2 MATERIALS

As in the previous studies, the materials were developed in an online format, which allowed participants to complete the study without being required to attend sessions in person. The study described in the previous chapter was developed online
using the University of Leicester Plone content management system, which was no longer available when materials were being developed for this current study. Therefore another software package was used, ‘Zoomerang’ (Market Tools, 2010) which is available by online subscription. This software produced very similar online stimuli to those used in the previous chapters, and allowed for data to be exported in Microsoft Excel format which could then be transferred into SPSS for analysis.

The materials adapted into the online format were the 10-item FEEBS, JBS, and the GBJW scale. In addition to responding to these scales, participants also read through the same fictional murder trial scenario used in the study described in Chapter four (available in Appendix C). The version of the murder trial used in this study included only weak DNA evidence for two reasons; Firstly, in the previous study participants overwhelmingly agreed on the high probative value of the strong DNA evidence and therefore it would appear that any possible differences in juror attitudes which might be identified using the FEEBS measure would not influence juror decision making when the evidence was obviously of high value. This previous finding and the subsequent decision to exclude strong DNA evidence from this study is supported by an interactionist model of juror decision making which would predict biases to influence decision making only when evidence is weak or ambiguous (Kassin & Wrightsman, 1988; Devine, Buddenbaum, Houp, Studebaker, & Stolle, 2009). Therefore for the purpose of testing the predictive validity of the FEEBS and in an attempt to explain the disagreement over weak DNA evidence seen in Chapter four, only weak DNA evidence was included in the current study.
6.3.3 Procedure

After completing the informed consent and demographic information, participants completed three scales; the FEEBS, JBS, and GBJW. After completing these measures, participants were presented with the fictional trial stimuli, which were the exact same materials used in the study described in Chapter four and available in Appendix C. Throughout the trial information, participants were required to indicate their opinion of the strength of each piece of trial evidence (on a 5-point Likert scale), as well as provide estimates of the probability that the defendant was guilty of the crime (at three points in the trial stimulus) and a final verdict decision (guilty/not-guilty of murder).

6.4 Results

6.4.1 Principal components analysis of the FEEBS

In order to investigate the underlying structures of the FEEBS measure, and to determine whether responses in this study are similar to those described in the previous study, a principal components analysis (PCA) was conducted on the 10-item FEEBS. As in the previous study, the PCA was carried out using oblique rotation (direct Oblimin in SPSS) and the adequacy of the sampling size was confirmed using the Kaiser-Meyer-Olkin measure which was in a good range, $KMO = .88$ (Field, 2009). The decision to conduct another exploratory PCA in this study was made, as the sample of participants differed significantly from the previous study (community sample versus student sample), and therefore moving on to confirmatory factor analysis would not have been justified at this stage.
The results of the current PCA were very similar to the previous PCA and are summarised in Table 17. Similar to the previous study, Kaiser’s criteria and visual inspection of the scree plot (Field, 2009) were both considered and supported the retention of two factors. In addition, parallel analysis also confirmed this decision to retain two factors (Hayton, et al. (2004). The same two components emerged in this PCA, with the only difference being that component one in the previous study represented the pro-prosecution biased items and in the current PCA these were identified as component two. This means that in the previous study the pro-prosecution component accounted for more of the variance in responses (34.06%) than the pro-defence component (13.92%) prior to rotation, however in the current study the pro-defence component accounted for more of the variance (44.05%) than the pro-prosecution component (14.13%). Overall these two components explained 58.17% of the total variance in the current analysis.

In order to determine the extent of the similarity between the component structure in this and the previous study, Tucker’s Coefficient of Congruence was calculated. This coefficient is calculated using the following formula (where \( a \) and \( b \) are loadings of a given variable on factors which are being compared), and is a measure of similarity between two factor structures obtained from subsequent samples.

\[
\text{Tucker's Coefficient of Congruence} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} b_{ij}}{N}
\]


For the current two PCA analyses, Tucker’s coefficient was found to be $r_{\text{congruence}} = .997$. This value indicates a very high level of similarity between these two studies, as typically values of greater than .95 are considered to reflect structures that are essentially equivalent (Krzanowski, 1979; Lorenzo-Seva & Berge, 2006).

This replication in the component structure of the FEEBS supports the separation of the measure into two subscales, one measuring pro-prosecution attitudes (FEEBpp) and the other representing pro-defence attitudes (FEEBpd). These subscales were negatively correlated with one another ($r = -.53, p < .001$) which is to be expected as they represent conceptually opposite biases about forensic science evidence. This correlation also supports the use of oblique rotation in the PCA, although similar to the previous chapter an orthogonal rotation (varimax method) was also conducted and resulted in the same component structure. These two subscales each had high reliability, as for both Cronbach $\alpha > .80$ (see Table 17).
Table 17

*Summary of principal components analysis for 10-item bias scale*

<table>
<thead>
<tr>
<th>Item</th>
<th>Pattern coefficients</th>
<th>Structure coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Component 1</td>
<td>Component 2</td>
</tr>
<tr>
<td>Police should not charge someone without forensics</td>
<td>.85</td>
<td>.81</td>
</tr>
<tr>
<td>If no forensics, the jury should not convict</td>
<td>.82</td>
<td>.81</td>
</tr>
<tr>
<td>If no forensics recovered the defendant is probably innocent</td>
<td>.78</td>
<td>.76</td>
</tr>
<tr>
<td>No forensics means investigators did not look hard enough</td>
<td>.70</td>
<td>.73</td>
</tr>
<tr>
<td>Every criminal leaves trace evidence at every scene</td>
<td>.56</td>
<td>-.19</td>
</tr>
<tr>
<td>Forensics always IDs the guilty person</td>
<td>.12</td>
<td>.90</td>
</tr>
<tr>
<td>Science is the most reliable way to ID perpetrators</td>
<td></td>
<td>.74</td>
</tr>
<tr>
<td>Forensic evidence is enough to convict</td>
<td>-.12</td>
<td>.72</td>
</tr>
<tr>
<td>Forensic always provides a conclusive answer</td>
<td>.69</td>
<td>-.40</td>
</tr>
<tr>
<td>Every crime can be solved with forensic science</td>
<td>-.18</td>
<td>.65</td>
</tr>
<tr>
<td>Eigenvalues (before rotation)</td>
<td>4.41</td>
<td>1.41</td>
</tr>
<tr>
<td>Cronbach’s α</td>
<td>.81</td>
<td>.82</td>
</tr>
</tbody>
</table>

*Note. N = 159.*
6.4.2 Relationship with other juror attitude measures

The descriptive statistics for the scale measures (FEEBpp, FEEBpd, JBS, and GBJW) are presented in Table 18. The JBS was found to be significantly correlated with the GBJW measure ($r = .21, p < .05$) which is consistent with the original findings by Kassin and Wrightsman (1983). Although the measure of Belief in a Just World used in this study was different from the measure used in the Kassin and Wrightsman study, this should not have changed the relationship between these two concepts, which indeed it has not. The FEEBpp was positively correlated with the JBS ($r = .24, p < .05$) which may be expected as the JBS is scored in the direction of a general pro-prosecution bias. The FEEBpp was also found to be correlated somewhat with the GBJW scale ($r = .17, p < .05$). The FEEBpd was not significantly correlated with either the JBS or the GBJW scale.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
<th>Range (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEBpp</td>
<td>13.50</td>
<td>2.97</td>
<td>7-19</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>11.41</td>
<td>2.83</td>
<td>7-19</td>
</tr>
<tr>
<td>JBS</td>
<td>45.72</td>
<td>5.28</td>
<td>32-66</td>
</tr>
<tr>
<td>GBJW</td>
<td>21.68</td>
<td>4.81</td>
<td>7-34</td>
</tr>
</tbody>
</table>

*Note. N = 159*
6.4.3 Evaluation of trial evidence

Participants read through witness statements from a fictional murder trial and were presented with four pieces of prosecution evidence; eye-witness testimony (the victim’s neighbour), testimony from the arresting police officer, pathologist’s testimony, and testimony from the DNA laboratory analyst (see Appendix C). After reading about each piece of evidence participants rated the usefulness of the evidence for determining the guilt of the defendant on a 5-point Likert scale (1=irrelevant/useless to 5=conclusive/definite). Descriptive statistics for the strength ratings are provided in Table 19.

It is clear from Table 19 that this weak DNA evidence is perceived as ambiguous by this sample of participants, as there is the full range of responses regarding the strength of this evidence, and similar to the previous study the mean strength rating for this evidence is in the ‘strong’ range (e.g. above the moderate value of three on the Likert scale).
Table 19

*Descriptive statistics for ratings of evidence strength*

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Mean</th>
<th>SD</th>
<th>Range (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye-witness evidence</td>
<td>2.81₂</td>
<td>0.78</td>
<td>1-5</td>
</tr>
<tr>
<td>Police officer</td>
<td>2.91₂</td>
<td>0.75</td>
<td>1-5</td>
</tr>
<tr>
<td>Pathologist</td>
<td>2.46₃</td>
<td>1.00</td>
<td>1-5</td>
</tr>
<tr>
<td>DNA analyst</td>
<td>3.75₄</td>
<td>0.78</td>
<td>1-5</td>
</tr>
</tbody>
</table>

*Note.* N = 159; subscripts indicate means which differ significantly at p < .05

### 6.4.4 Probability of guilt estimates

Participants were asked to estimate the probability that the defendant was guilty of the murder charge at three points in time within the online task. First, they estimated probability of guilt after reading a brief summary of the case which did not contain any details of evidence (referred to as the ‘prior probability of guilt’). Second, they made a subsequent estimate after hearing all of the testimony and evidence (‘post evidence probability of guilt’) and finally again after reading the summary of the case made by the defence and prosecution lawyers (‘final probability of guilt’). These estimates were expressed by participants as percentages, and the descriptive statistics are summarised in Table 20.
Table 20

Descriptive statistics for probability of guilt estimates

<table>
<thead>
<tr>
<th>DNA evidence</th>
<th>Mean</th>
<th>SD</th>
<th>Range (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior probability</td>
<td>66.16</td>
<td>17.78</td>
<td>0-100</td>
</tr>
<tr>
<td>Post-evidence</td>
<td>73.55</td>
<td>18.26</td>
<td>0-100</td>
</tr>
<tr>
<td>Final probability</td>
<td>67.73</td>
<td>20.58</td>
<td>0-100</td>
</tr>
</tbody>
</table>

*Note. N = 159*

The mean prior probability of guilt estimates did differ significantly from a test value of 50% ($t (158) = 11.46, p < .001$) which indicates that this particular sample of participants found the summary of the case to indicate a moderate probability that the defendant was guilty of murder, before hearing any of the evidence. There is a similar trend in the probability of guilt estimates to the earlier study (see Chapter four), as the mean probability increases after the trial evidence is presented, and then decreases slightly after the prosecution and defence summaries are presented.

6.4.5 Verdict decisions

The final part of the online task required participants to decide on a verdict of guilty or not-guilty of murder. Overall 73 participants (46%) voted guilty and 86 participants (54%) voted not-guilty in this scenario. This further demonstrates the ambiguity of this case, and the fact that there was wide disagreement among participants regarding the strength of the evidence presented.
A t-test comparing the mean final probability of guilt estimates between the two verdict categories confirmed that participants who voted guilty had significantly higher mean final probability of guilt estimates (80.15) than those who voted not-guilty (57.19), \( t = -8.42, p < .05 \).

### 6.4.6 Relationship between juror attitudes and decision making

Having demonstrated that participants who voted guilty had significantly higher final probability of guilt estimates than those who voted not-guilty, a multiple regression analysis was conducted to determine which pieces of evidence predicted the perceived final probability of guilt. Participants read details of four pieces of evidence relevant to the trial; eye-witness testimony, police officer testimony, pathologist’s statement and the DNA analyst testimony. For each of these pieces of evidence participants indicated their perception of the strength of the evidence in terms of its usefulness for determining the guilt of the defendant. Therefore in the multiple regression analysis described below the strength ratings of the four pieces of evidence were entered as predictor variables, and the final probability of guilt estimate was the outcome variable. The results of the regression are summarised in Table 21.

For these data the assumptions of independent errors (D-W = 2.08), normally distributed errors, linearity, and homoscedasticity were all met and multicollinearity was not a concern (VIF = 1.3). The results of the multiple regression indicated that the perceived strength of the DNA evidence was the only significant predictor of final probability of guilt estimates.
Table 21

*Predictors of final probability of guilt*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model $B$</th>
<th>SE $B$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.75</td>
<td>8.10</td>
<td></td>
</tr>
<tr>
<td>Eye-witness</td>
<td>3.56</td>
<td>2.11</td>
<td>.14</td>
</tr>
<tr>
<td>Police</td>
<td>1.83</td>
<td>2.17</td>
<td>.07</td>
</tr>
<tr>
<td>Pathologist</td>
<td>-.77</td>
<td>1.50</td>
<td>-.04</td>
</tr>
<tr>
<td>DNA</td>
<td>12.22*</td>
<td>1.99</td>
<td>.46*</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>.32</td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td></td>
<td>14.38*</td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 159, *$p < .001*

Having determined that the perceived strength of the weak DNA evidence was the only significant predictor of final probability of guilt, a second multiple regression analysis was run to determine if any of the pre-trial attitude measures predicted how strong participants perceived the DNA evidence to be in the trial scenario. In this regression model the predictor variables were the FEEBpp, FEEBpd, JBS, and GBJW scores. The results of the regression are summarised in Table 22.

The only significant predictor of the strength rating for the weak DNA evidence was the participant’s score on the FEEBpp. The FEEBpd scale approached significance as a predictor of DNA evidence strength ($t = -1.87, p = .06$) however the JBS and GBJW scores did not emerge as useful predictors in the regression model. The sign of $\beta$-
values in the regression output indicate the direction of the relationship between the predictors and the strength rating of DNA evidence. For the FEEBpp scores this value is positive which confirms that as the FEEBpp score increases (indicating stronger agreement with pro-prosecution attitudes) so does the strength rating of the weak DNA evidence. The FEEBpd predictor has a negative β value in the regression output, which indicates a negative relationship between the score on the FEEBpd subscale and the strength rating of the DNA evidence (although this relationship did not reach statistical significance).

Table 22

*Predictors of DNA strength rating*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.32</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>FEEBpd</td>
<td>-.04</td>
<td>.02</td>
<td>-.15</td>
</tr>
<tr>
<td>FEEBpp</td>
<td>.13*</td>
<td>.02</td>
<td>.48*</td>
</tr>
<tr>
<td>JBS</td>
<td>.00</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>GBJW</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>16.81*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 159, *p < .001*
In order to further understand the underlying process connecting participants’ evidence evaluation bias (FEEBpp) and their probability of guilt estimates a simple, three variable mediation model was hypothesised and tested. The model is depicted in Figure 9. In order to test the model, three regression equations were calculated following the procedures outlined in Baron and Kenny (1986) and MacKinnon (2008). In order to confirm that a relationship is mediated by a third variable, the independent variable must significantly predict both the mediator variable and the dependent variable (Baron & Kenny, 1986). In addition to these criteria, the direct relationship between the independent and dependent variables must decrease when controlling for the effect of the mediator variable in a third regression model (MacKinnon, 2008). The standardised beta coefficients for these relationships are given in Figure 9, and it can be seen that the criteria for a partial mediation have been met.

The methods outlined above for confirming the effect of a mediator variable in a relationship do not provide a test of statistical significance of the mediation effect, and therefore the Sobel test was used to evaluate the significance of the mediation in the current model (MacKinnon, 2008). An SPSS macro for running a Sobel test of a mediation model is provided by Preacher and Hayes (2004) and the results of the test confirmed a significant indirect effect of the mediator variable (perceived strength of DNA evidence) on probability of guilt ($p < .001$).
**Figure 9.** Partial mediation model for relationship between evidence bias (FEEBpp) and final probability of guilt estimates including standardised regression coefficients. The standardised regression coefficient between FEEBpp and probability of guilt controlling for perceived evidence strength is in parentheses.

*coefficients significant at \( p < .05 \).
6.5 Discussion

In an effort to explain differences in mock jurors’ perceived strength ratings of forensic evidence in cases with weak DNA evidence, a measure of pre-trial attitudes about forensic evidence was developed. The Forensic Evidence Evaluation Bias Scale (FEEBS) was developed with reference to the anecdotal claims in the CSI Effect literature concerning problematic attitudes held by some jurors, which have been reported to contribute to the jury decision making process.

Initially, potential items for the FEEBS were developed and tested, and eventually ten items were selected for the final scale based on their high inter-item and item-total correlations (see Chapter five). An exploratory principal components analysis (PCA) revealed that the ten items represented two distinct underlying components, each consisting of five items, which were interpreted as corresponding to pro-prosecution and pro-defence related attitudes about forensic science evidence.

In the current study, the component structure of the FEEBS was tested again, with the same two components emerging from the overall scale.

The only difference between the component structure in the initial study compared with the structure in this second study was that the order of the components was reversed. In the first PCA (Chapter five), component one represented the pro-prosecution items but in the current validation study component one corresponded to the pro-defence items. This reversal of the components is not considered to be problematic for two reasons. First, the components in both analyses are identical to one another (in that they contain the same items) and therefore the configuration of the components remains constant between studies. This is
considered to be more important than the magnitude of the patterns, especially when comparing two studies with significantly different samples of participants (Rummel, 1970). Second, after oblique rotation of the components, the relative importance of the two components is equalised (Jolliffe, 2002). Therefore the rotated component structure for both studies is essentially the same. The similarity between the two PCA results were also quantitatively confirmed using Tucker’s Coefficient of Congruence.

The replication of the PCA conducted on the FEEBS supports the use of two subscales, as defined by the components, representing pro-prosecution attitudes (FEEBpp) and pro-defence attitudes (FEEBpd). The relationship between these subscales and the other attitude measures used in this study revealed that the FEEBS approach to measuring juror bias does not seem to be measuring the same concepts as previous bias measures. The Juror Bias Scale (JBS) was significantly, but not highly, correlated with the FEEBpp which indicates that these scales may be conceptually similar, but not likely to be measuring the same construct. This is certainly a reasonable result, as the JBS is scored in the direction of a general pro-prosecution bias and the FEEBpp measures a more specific evidence-related bias which would act in favour of the prosecution. The FEEBpp was also found to be very slightly correlated with General Belief in a Just World (GBJW) which is also to be expected as previous research found the JBS to be moderately correlated with Belief in a Just World attitudes (Kassin & Wrightsman, 1983).

The relationships between the FEEBpp and other juror bias measures is encouraging as it suggests that the evidence bias being measured by the FEEBpp is conceptually related to other pre-trial, pro-prosecution attitudes previously measured
in jurors. Also of importance is the fact that these correlations are not very strong, which lends support to the argument that the FEEBS is measuring a fundamentally different bias, and is therefore a unique contribution to the jury bias literature. The fact that the FEEBpd subscale did not correlate significantly with the JBS or the GBJW is perhaps somewhat surprising, as it could be argued that a small negative correlation would be expected. However, it is possible that the FEEBpd subscale is measuring an evidence-specific construct which is not addressed by either the JBS or the GBJW scales.

The development of the FEEBS and the theoretically relevant results of the PCA analyses are encouraging, however it was also important to investigate whether the juror bias being measured was related to the decisions made in the context of a criminal trial. As one might expect, participants who voted guilty in the murder trial scenario perceived the defendant as more likely to have committed the offence, as measured by the final probability of guilt estimates. The perceived strength of the weak DNA evidence was the only significant predictor of final probability of guilt, which is a reasonable finding, as the DNA evidence was the only testimony which had the potential to physically link the defendant to the crime (albeit this link was tenuous).

The relationship between verdict, probability of guilt, and the perceived strength of the weak DNA evidence is perhaps common sense, however of interest in terms of this particular juror bias research is what predicts how strong participants will perceive the weak DNA evidence to be. The only pre-trial measure which significantly predicted the perceived strength of the DNA evidence was the FEEBpp subscale score.
This suggests that participants who responded with higher levels of agreement to pro-prosecution evidence attitudes perceived weak DNA evidence to be of higher probative value in this murder trial scenario.

In order to further understand the relationship between perceived strength of evidence and probability of guilt estimates, and the way in which pre-trial bias may impact on these decisions, a simple mediation model was proposed. This model suggested that in cases where forensic evidence is of a weak standard a pre-trial pro-prosecution bias exists which partially predicts how strong a juror will perceive forensic evidence to be, and in turn how probable they conclude it is that the defendant is guilty. This relationship has been hypothesised in the literature previously but no studies to date have confirmed the model. This type of approach is endorsed in the following quote by Lecci and Myers (2008); “It is expected that indirect effects, particularly as mediated by the evaluation of evidence, would carry the greatest influence of pretrial bias” (p. 2029).

From a practical perspective awareness of this evidence bias, and the potential impact on criminal trials, is crucial as it demonstrates the importance of jurors’ prior beliefs and attitudes in cases where the evidence is not overwhelmingly strong. Indeed it could be argued that many cases decided by juries are inherently of this nature, given that if a case was supported by undeniably strong physical evidence it would probably not require a trial at all. It seems reasonable then to assume that there is likely to be ambiguity in the forensic evidence at many jury trials. If the FEEBS is found to be a robust indicator of a pre-trial evidence evaluation bias which can predict juror perceptions of ambiguous evidence it would be possible to use such a
measure in peremptory challenges. For example, in a criminal case where the prosecution’s main physical evidence is of a weak probative standard, the defence could argue that jurors who hold strong pro-prosecution beliefs about forensic science (as measured by the FEEBpp subscale) should be excluded from the jury as they are likely to attribute too much weight to the weak forensic evidence.

There are limitations to the conclusions that can be drawn at this stage about the FEEBS. Firstly, only the FEEBpp subscale was a significant predictor of perceived evidence strength in the current study, while the FEEBpd subscale was not. Referring back to the theoretical concepts represented by each of the subscales, the FEEBpd items refer to pro-defence beliefs about forensic evidence, and many of these beliefs are most relevant to cases in which no forensic evidence is presented (e.g. ‘if there is no forensic evidence then investigators didn’t look hard enough’). In the present study there was forensic evidence presented by the prosecution and therefore the pro-defence attitudes measured by the FEEBpd may not be influential in this type of scenario. This concept will be tested in the next chapter when conditions with no forensic evidence are included in the experimental manipulation.

A second limitation to the current study is the fact that this sample consisted entirely of undergraduate students. Although all of the participants were jury eligible, the use of students results in a very homogenous sample which is not representative of people who typically serve on juries in England and Wales. The Crown Court Jurors Survey (HMCS, 2009) reported that only 6% of jurors were between the ages of 18-24, while 94% of the students in this study were in this age range. Similarly there are significant differences in gender distribution between the student sample and actual
jurors, with 88% of the student sample being female in the current study, while 55% of the jurors in 2009 were female (HMCS, 2009). These differences impact on the ability to generalise from the findings of the validation study to real juries, and the studies in the following chapters aimed to recruit more representative samples for further testing of the usefulness of the FEEBS.

The same murder trial scenario was used in this study as previously in this thesis, which allows a comparison of the current student sample with the community-based sample used in Chapter four. One notable difference between the two studies is in the prior probability of guilt estimates given before any evidence is provided. In Chapter four the community sample’s mean prior probability of guilt was not significantly different from 50%, however the current student sample’s mean prior probability of guilt was 66%, which did differ significantly from 50%. The reason for this difference is not clear, as these estimates were made by participants before they heard any evidence in the case. One possible explanation for the higher estimates in the student sample is that by responding to the FEEBS, JBS and GBJW scales at the beginning of the online task, this may have primed participants to be more judgmental causing them to perceive higher probabilities of suspect guilt before evidence is presented. However previous research has shown that administering a measure of pre-trial bias had no significant priming effect on how participants evaluated evidence or their overall verdict preference (Morris & Lecci, 2005). Another potential reason for the different prior probabilities is offered by Martin, De La Fuente, De La Fuente, and Garcia (2007), as their research findings revealed a guilty bias in student samples for
cases with ambiguous evidence, compared with jury eligible members of the general public.

In order to address some of the limitations of the current study, the study described in the following chapter used jury-eligible participants recruited from the general population which resulted in a more representative sample. In addition, the trial scenario used in the following chapter is not a murder case, so the impact of the FEEBS on decision making about evidence in a robbery and sexual assault scenario can be tested. In order to investigate whether the FEEBpd subscale predicts decision making in cases where no forensic evidence is presented, the next chapter also included versions of the trial scenarios with no forensic evidence presented by the prosecution.
CHAPTER SEVEN: TESTING THE PREDICTIVE VALIDITY OF THE FEEBS WITH ROBBERY AND SEXUAL ASSAULT SCENARIOS

7.1 ABSTRACT

In this study the predictive validity of the FEEBS was further tested with two additional crime types; robbery and sexual assault. In addition, the presence of DNA evidence was manipulated which resulted in additional versions of the trial scenarios with no forensic evidence presented as part of the prosecution’s case. The results replicated the previous findings, as for both robbery and sexual assault the FEEBpp subscale was the only significant predictor of the perceived strength of the weak DNA evidence. In addition, scores on the FEEBpd subscale significantly predicted the perceived strength of circumstantial evidence in cases which included no DNA evidence. The results are discussed in the context of partial mediation models and the potential impact of these attitudes on probability of guilt estimates and subsequent verdict preferences. This chapter also presents the results of a Confirmatory Factor Analysis of the FEEBS, which supported the two-factor model previously identified using exploratory PCA.

7.2 INTRODUCTION

The studies presented in the previous chapter suggested that the Forensic Evidence Evaluation Bias Scale (FEEBS) may be an effective measure for identifying juror pre-trial bias about forensic evidence. More specifically it was demonstrated that scores on the pro-prosecution subscale of the FEEBS significantly predicted participants’ perceptions of the strength of weak DNA evidence presented in the context of a murder trial, which contributed to their verdict decision. However, thus
far this relationship has only been tested for one crime type (murder) and only using
jury eligible student participants.

In the previous study (Chapter six) only the FEEBpp subscale was found to be a
significant predictor of perceived DNA evidence strength, while the FEEBpd subscale
did not predict any of the measured mock juror decisions. This may be due to the fact
that the FEEBpp subscale measures attitudes people have concerning the importance
and infallibility of forensic evidence, which was particularly relevant in the murder
scenario used. However, the FEEBpd subscale contains items which relate to people’s
beliefs about cases in which forensic evidence is expected by a juror but not presented
at trial, and it is therefore conceivable that this subscale would be more effective at
predicting juror decisions for cases in which no forensic evidence is presented as part
of the prosecution’s case.

This chapter aimed to extend the previous findings to two new crime types,
robbery and sexual assault, as well as manipulating the presence of forensic DNA
evidence in the scenarios. It is important to test the FEEBS with multiple crime types
as some previous research measuring juror bias has found that sexual assault trial
scenarios have produced unexpected and inconsistent findings (Kassin & Wrightsman,
1983) which may be the result of other underlying biases or preconceptions
concerning certain types of offenders and victims involved in these types of crime
scenarios. It will also be possible in this chapter to investigate the possibility that the
FEEBpd subscale will be a significant predictor of juror decisions in cases where no
forensic evidence is presented through the inclusion of conditions containing no
forensic DNA evidence.
7.3 Method

7.3.1 Participants

The same target population was required for this study as was used in Chapters three through five; members of the public who are eligible for selection as jurors based on criteria described by the Criminal Justice System for England and Wales (www.cjonline.gov.uk). As in the previous chapters, basic jury eligibility was determined by asking participants to indicate their age and whether they had ever been convicted of a criminal offence.

The final sample consisted of 200 participants, of which 71 (35.5%) were male and 129 (64.5%) were female. The age of participants ranged from 18-69 years ($M_{age} = 33.72$, $SD = 13.20$). Overall, 25% of participants had completed high school, 58% had completed undergraduate degrees, and 17% held postgraduate qualifications.

7.3.2 Materials

The online presentation of the study materials was developed using Zoomerang, an online survey tool available by subscription. In total, four versions of the online material were developed separately which represent the experimental conditions in this study. The Zoomerang software then published each condition as a separate survey and assigned each a unique webpage address (URL). These four URLs were then used in the design of a separate webpage which was programmed to randomly allocate and forward participants to one of the four conditions/webpages.

The trial scenarios consisted of two versions of a robbery trial (with and without DNA evidence) and two versions of a sexual assault trial (with and without DNA evidence). The scenarios contained nearly identical testimony from an
eyewitness who saw the perpetrator briefly (but did not witness the crime occur), the responding police officer, and the arresting police officer. The only evidence manipulated was the DNA evidence, which was either present (but of weak probative value) or absent. In conditions where DNA evidence was presented, the testimony of the laboratory analyst was included in the trial materials and the evidence presented was of a weak probative standard, similar to the evidence in the previous murder trial study reported in Chapter four. The trial scenario text is available in Appendix E.

Participants initially completed some demographic information to confirm basic jury eligibility followed by the FEEBS, the JBS (Kassin & Wrightsman, 1983), and GBJW (Dalbert, Montada, & Schmidt, 1987) scale before moving on to complete the mock juror task. Once their responses were submitted, the Zoomerang software compiled the data which could then be exported into Microsoft Excel and then SPSS for statistical analyses.

7.3.3 PROCEDURE

Recruitment of participants was carried out in a similar way to the previous studies in this thesis, using the School of Psychology’s participant panel, online advertising websites and a snowballing technique. When a participant visited the main study webpage they were presented with a brief explanation of the study aims (described as investigating criminal justice attitudes and juror decisions) and asked to click on a link within the page to continue. This link activated the random allocation program which forwarded the participant to one of the four Zoomerang study web pages (each containing one of the experimental conditions).
Each trial scenario began with a brief summary of the case, after which participants indicated how probable they thought it was that the defendant was guilty of the crime (expressed as a percentage between 0 and 100). Testimony from each of the key witnesses followed on separate pages, with participants indicating the probative value of each piece of evidence on a 5-point Likert scale (ranging from irrelevant/useless to conclusive/definite). After hearing all of the testimony participants read summaries of the case presented by the prosecution and the defence attorneys, after which they again estimated the probability that the defendant committed the crime as well as their verdict decision (guilty/not-guilty).

After completing the online task, participants were directed to a final page which thanked them for their participation and provided a debriefing statement and the researchers’ contact details.

7.4 Results

7.4.1 Confirmatory factor analysis of the FEEBS

In the previous chapters the results of exploratory Principal Components Analyses (PCA) suggested that the FEEBS is best conceptualised as having two distinct components. Each of the components contained five items with high factor loadings, and this structure was previously replicated with both a jury-eligible student sample and a sample drawn from the general public. The current study once again collected responses on the FEEBS from participants, and in order to further test the hypothesised component structure of this scale Confirmatory Factor Analysis (CFA) was used. The consistency of the PCA results in the previous chapters provides empirical and theoretical support for an a priori model hypothesis, which can be
tested using confirmatory factor analytic techniques (Byrne, 2010). The use of CFA is appropriate in this chapter as it can be used to verify the hypothesised dimensional structure of the FEEBS with a new sample of jury-eligible participants (DiStefano & Hess, 2005).

The CFA was conducted using AMOS version 16 and a graphical representation of the hypothesised two-factor model of the FEEBS is provided in Figure 10. An assumption for the use of AMOS for CFA is that the data are confirmed to be multivariate normal. This assumption was confirmed to be met for these data and therefore the analysis was run using the Generalised Least Squares (GLS) method of estimation which is appropriate for this sample size of 200 with no missing data (Loehlin, 2004; Kline, 1994). AMOS provides a number of goodness-of-fit statistics which indicate how well the hypothesised model fits the observed data, and a selection of these are summarised in Table 23.
Figure 10. Confirmatory factor analysis of the two-factor FEEBS with standardised regression weights (factor loadings)
Table 23

*Summary of Goodness-of-Fit statistics for 2-factor model of the FEEBS*

<table>
<thead>
<tr>
<th></th>
<th>$X^2$</th>
<th>df</th>
<th>GFI</th>
<th>TLI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-factor model</td>
<td>39.06</td>
<td>34</td>
<td>.96</td>
<td>.95</td>
<td>.03</td>
</tr>
<tr>
<td>Independence model</td>
<td>170.10</td>
<td>45</td>
<td>.83</td>
<td>0.00</td>
<td>.12</td>
</tr>
</tbody>
</table>

The $X^2$ value for the model (39.06) was non-significant ($p = .25$), which indicates that the hypothesised model is a good fit with these data (DiStefano & Hess, 2005), however it has been noted that the $X^2$ statistic can be misleading with large sample sizes and therefore it is important to also interpret the other available goodness-of-fit statistics (Oakman, Van Ameringen, Mancini, & Farvolden, 2003; Bentler, 1990). The most common statistic to report in CFA is the Goodness of Fit Index (GFI), which should be greater than .90 if the model fits the data adequately (Oakman, et al.), although some authors suggest values of .95 or greater should be achieved (DiStefano & Hess). Using either of these suggested threshold values, the GFI for the current CFA indicates a model that fits the data quite well. The final two statistics which are presented in Table 23 are the Tucker-Lewis Index (TLI) and the Root Mean Squared Error of Approximation (RMSEA). The TFI compares the hypothesised model to an alternative model with no structure and this value should be close to .95 if the proposed model is a good fit to the data (Byrne, 2010), which indeed the current value
is. The RMSEA takes into account the complexity of the model and should be less than .05 if the model is a close fit (Oakman, et al.), which is true of the current model.

In addition to comparing the two-factor model with the independence model, which assumes a zero correlation between items and latent factors (reported above), it is also important to compare the hypothesised model with a single factor model (Mulaik, 2010). Therefore the factor analysis was run again forcing a solution with one single factor using the generalised least squares method (Kline, 1994), and the resulting chi-square goodness-of-fit test confirmed that this solution was not an adequate fit to these data ($\chi^2 = 86.07, p < .001$).

7.4.2 COMPARISON OF JUROR ATTITUDE MEASURES

Descriptive statistics for the attitude measures (FEEBpp, FEEBpd, JBS, GBJW) for the overall sample are summarised in Table 24. The correlations between the pre-trial attitude measures are summarised in Table 25. These correlations are similar to the previous studies, however in Chapter six there was also a significant correlation between JBS and GBJW, which was not the case with this current sample.
Table 24

*Descriptive statistics for attitude scale measures*

<table>
<thead>
<tr>
<th></th>
<th>Mean score</th>
<th>SD</th>
<th>Min-max score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEBpp</td>
<td>13.68</td>
<td>3.28</td>
<td>6-20</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>12.76</td>
<td>3.40</td>
<td>5-20</td>
</tr>
<tr>
<td>JBS</td>
<td>46.56</td>
<td>5.84</td>
<td>22-66</td>
</tr>
<tr>
<td>GBJW</td>
<td>20.72</td>
<td>5.36</td>
<td>6-34</td>
</tr>
</tbody>
</table>

*Note. N = 200*

Table 25

*Correlations between attitude scale measures*

<table>
<thead>
<tr>
<th></th>
<th>FEEBpp</th>
<th>FEEBpd</th>
<th>JBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEBpd</td>
<td>-.49**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBS</td>
<td></td>
<td>-.18*</td>
<td>-18*</td>
</tr>
<tr>
<td>GBJW</td>
<td>.18*</td>
<td>-.23**</td>
<td>.09</td>
</tr>
</tbody>
</table>

*Note. *p < .05, **p < .001; N = 200.*
7.4.3 Relationship between source of forensic knowledge and measures of pre-trial attitudes/bias

Participants were given a list of five potential sources of their prior knowledge about forensic science (e.g. a range of media types and personal experience), and were asked to indicate how influential they felt each source was for their own personal knowledge about forensic science (using a 5-point Likert scale ranging from ‘not at all important’ to ‘extremely important’). Table 26 summarises the descriptive statistics for these ratings, and Table 27 provides the results of a correlation analysis for participants’ ratings of each source of forensic knowledge.

It appears from the patterns in the correlations between sources of knowledge, that participants who rated TV/movies as being an important source of knowledge about forensic science also rated crime fiction novels as highly important ($r = .46, p < .001$) as well as news media ($r = .22, p < .05$). It is also worth noting that the only significant correlation with ratings for the importance of formal training was the importance of the internet as a source of knowledge about forensic science ($r = .23, p < .001$).
Table 26

Descriptive statistics for influence ratings of sources of forensic science knowledge

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Range (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV/movies</td>
<td>3.20</td>
<td>1.04</td>
<td>1-5</td>
</tr>
<tr>
<td>Crime fiction novels</td>
<td>2.59</td>
<td>1.13</td>
<td>1-5</td>
</tr>
<tr>
<td>News media</td>
<td>3.62</td>
<td>.99</td>
<td>1-5</td>
</tr>
<tr>
<td>Internet</td>
<td>3.23</td>
<td>1.09</td>
<td>1-5</td>
</tr>
<tr>
<td>Formal training</td>
<td>3.35</td>
<td>1.60</td>
<td>1-5</td>
</tr>
</tbody>
</table>

Note. N = 200

Table 27

Correlations between ratings of importance for sources of forensic knowledge

<table>
<thead>
<tr>
<th></th>
<th>TV/Movies</th>
<th>Crime fiction novels</th>
<th>News media</th>
<th>Internet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime fiction novels</td>
<td>.46**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News media</td>
<td>.22*</td>
<td>.17*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>.13</td>
<td>.06</td>
<td>.23**</td>
<td></td>
</tr>
<tr>
<td>Formal training</td>
<td>.00</td>
<td>.07</td>
<td>-.05</td>
<td>.23**</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .001; N = 200.
As the *CSI effect* literature claims that attitudes and beliefs about forensic science originate from the source of jurors’ information (e.g. particularly crime fiction television programmes), multiple regression was used to determine whether any of the influence ratings for the sources of forensic science knowledge predicted participants’ scores on the FEEBS subscales. The results of these multiple regression analyses are provided in Table 28. The only significant predictors of participant scores on the FEEBpp subscale were the ratings for TV/movies and crime fiction novels (negative $\beta$-value), while the only predictor of scores on the FEEBpd was the influence rating of news media (negative $\beta$-value) as a source of knowledge about forensic science.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>$\beta$</th>
<th>B</th>
<th>SE B</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>12.41</td>
<td>1.20</td>
<td></td>
<td>15.52</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>TV/Movies</td>
<td>.55</td>
<td>.25</td>
<td>.18*</td>
<td>-.18</td>
<td>.26</td>
<td>-.06</td>
</tr>
<tr>
<td>Crime novels</td>
<td>-.59</td>
<td>.23</td>
<td>-.20*</td>
<td>.10</td>
<td>.24</td>
<td>.03</td>
</tr>
<tr>
<td>News media</td>
<td>.10</td>
<td>.24</td>
<td>.03</td>
<td>-.52</td>
<td>.26</td>
<td>-.15*</td>
</tr>
<tr>
<td>Internet</td>
<td>.36</td>
<td>.22</td>
<td>.12</td>
<td>-.17</td>
<td>.23</td>
<td>-.05</td>
</tr>
<tr>
<td>Formal training</td>
<td>-.14</td>
<td>.15</td>
<td>-.07</td>
<td>.00</td>
<td>.16</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05; *N* = 200.
7.4.4 Evaluation of Trial Evidence

Across the four conditions the manipulated variables were crime type (robbery or sexual assault) and DNA evidence (present, not present). In each condition participants read testimony from an eyewitness (who saw the perpetrator, but did not witness the crime event), the responding police officer, and the arresting police officer. In the DNA evidence conditions, there was additional testimony from the DNA laboratory analyst, although the probative value of the DNA evidence was very low as defined by the mobility/relevance concepts presented in Chapter three (see Figure 1). After reading about each piece of evidence participants were asked to rate the usefulness of that evidence for determining the guilt of the defendant, using a 5-point Likert scale ranging from ‘irrelevant/useless’ to ‘conclusive/definite’. The mean strength ratings for each piece of evidence by condition are summarised in Table 29.
Table 29

*Mean strength ratings of evidence types by condition*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Eye-witness evidence</th>
<th>Responding officer</th>
<th>Arresting officer</th>
<th>DNA laboratory analyst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery no DNA</td>
<td>2.36&lt;sub&gt;a&lt;/sub&gt;</td>
<td>2.46&lt;sub&gt;b&lt;/sub&gt;</td>
<td>3.60&lt;sub&gt;d&lt;/sub&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td>Robbery with DNA</td>
<td>2.56&lt;sub&gt;a&lt;/sub&gt;</td>
<td>3.44&lt;sub&gt;c&lt;/sub&gt;</td>
<td>2.48&lt;sub&gt;e&lt;/sub&gt;</td>
<td>3.80</td>
</tr>
<tr>
<td>Sexual assault no DNA</td>
<td>2.52&lt;sub&gt;a&lt;/sub&gt;</td>
<td>3.14&lt;sub&gt;c&lt;/sub&gt;</td>
<td>3.62&lt;sub&gt;d&lt;/sub&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td>Sexual assault with DNA</td>
<td>2.72&lt;sub&gt;a&lt;/sub&gt;</td>
<td>3.10&lt;sub&gt;c&lt;/sub&gt;</td>
<td>3.40&lt;sub&gt;d&lt;/sub&gt;</td>
<td>3.80</td>
</tr>
</tbody>
</table>

*Note.* $N = 50$ in each condition; column means with different subscripts differ significantly at $p < .05$

Due to the fact that the mean strength ratings for the DNA evidence were identical for the robbery and sexual assault scenarios, a MANOVA was used to investigate whether there were any significant differences between the conditions on the strength ratings of the other three evidence types (eyewitness, responding police officer, and arresting police officer). The assumption of multivariate homogeneity was violated by these data, and therefore the Pillai’s trace statistic was used as this value is robust to this violation when sample sizes are equal (Field, 2009). Overall there was an effect of condition on the perceived strength of the responding officer’s and arresting officer’s testimony, $V = 0.43, F (9, 588) = 10.79, p < .001$. Table 29 provides details of the separate univariate ANOVAs and indicates which mean strength ratings differed significantly across conditions (denoted by subscripts).
Overall the ratings of the evidence types were quite similar across conditions with a couple of exceptions. The testimony from the responding officer was rated significantly lower on average in the ‘robbery with no DNA’ condition, and the arresting officer’s testimony was rated significantly lower on average in the ‘robbery with DNA’ condition.

7.4.5 Probability of guilt estimates

As described in the studies presented in the previous chapters, participants were asked to estimate the probability that the defendant was guilty of the crime (robbery or sexual assault) at three points in time during the online task. The prior probability of guilt was estimated after participants read a summary of the case which did not contain any evidence, the post-evidence probability was estimated immediately after all of the evidence had been presented, and the final probability was reported after the prosecution and defence summaries were presented. Participants were asked to express these guilt estimates as percentages (from 0-100) and the mean probabilities by condition are summarised in Table 30.

As indicated in Table 30, the probability of guilt estimates for the robbery and sexual assault scenarios show the same pattern as those in the murder scenario presented in Chapters four and six. In all cases the probability of guilt estimates increase after the evidence is presented, and decrease slightly after hearing the prosecution and defence lawyers’ summaries at the end of the trial.
Table 30

Mean probability of guilt estimates by condition

<table>
<thead>
<tr>
<th></th>
<th>Prior probability of guilt</th>
<th>Post evidence probability of guilt</th>
<th>Final probability of guilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery no DNA</td>
<td>50.80</td>
<td>67.46</td>
<td>65.60</td>
</tr>
<tr>
<td>Sexual assault no DNA</td>
<td>49.64</td>
<td>74.42</td>
<td>67.88</td>
</tr>
<tr>
<td>Robbery with DNA</td>
<td>52.30</td>
<td>80.08</td>
<td>71.14</td>
</tr>
<tr>
<td>Sexual assault with DNA</td>
<td>51.60</td>
<td>82.66</td>
<td>78.20</td>
</tr>
<tr>
<td>Overall</td>
<td>51.08</td>
<td>76.16</td>
<td>70.70</td>
</tr>
</tbody>
</table>

Note. N = 50 in each condition

The mean prior probability of guilt values did not differ significantly from a test value of 50, which indicates that this sample of participants did not perceive the trial summaries to favour either the prosecution or the defence before reading any of the trial evidence.

To investigate the impact of crime type and presence of DNA evidence on the post-evidence and final probabilities of guilt, a factorial MANOVA was conducted with the independent variables crime type (robbery or sexual assault) and evidence (DNA or no DNA) and the dependent variables prior, post-evidence and final probability of guilt. Box’s test confirmed that the assumption of equality of covariance matrices was
violated with these data, so the Pillai-Bartlett statistic was interpreted as this statistic is robust when the MANOVA assumptions are violated, when sample sizes in each cell are equal (Field, 2009).

Using Pillai’s Trace statistic ($V = .09$, $F (3, 194) = 6.73, p < .001$), there was a significant main effect of evidence on both the post-evidence probability of guilt ($F (1, 196) = 19.07, p < .001, r = .29$) and the final probability of guilt ($F (1, 196) = 8.33, p < .05, r = .20$). There was no significant main effect of crime type, and no interaction between the independent variables. This confirms that when DNA is presented the probability of guilt estimates were significantly higher, regardless of the crime type.

7.4.6 VERDICT DECISIONS

The final question asked participants to indicate whether they would vote guilty or not guilty, of robbery or sexual assault, depending on their condition. Table 31 summarises the frequencies of each verdict decision, by condition.

It is clear that these cases were considered by participants to be quite ambiguous, which was the intention of these scenarios in order to increase the potential for impact of pre-trial attitudes and biases on participants’ decisions.
Table 31

*Frequency of verdict decisions by condition*

<table>
<thead>
<tr>
<th></th>
<th>Guilty</th>
<th>Not guilty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery no DNA</td>
<td>20 (40%)</td>
<td>30 (60%)</td>
</tr>
<tr>
<td>Sexual assault no DNA</td>
<td>16 (32%)</td>
<td>34 (68%)</td>
</tr>
<tr>
<td>Robbery with DNA</td>
<td>21 (42%)</td>
<td>29 (58%)</td>
</tr>
<tr>
<td>Sexual assault with DNA</td>
<td>26 (52%)</td>
<td>24 (48%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>83 (41.5%)</td>
<td>117 (58.5%)</td>
</tr>
</tbody>
</table>

*Note. N = 200.*

As in the previous studies, participants who voted guilty had significantly higher final probability of guilt estimates ($M_{\text{prob}} = 89.88$) than those who voted not guilty ($M_{\text{prob}} = 57.11$), $t = -14.47$, $p < .001$. There was no significant effect of condition on the distribution of verdict decisions ($\chi^2 = 4.18$, $p = .24$).

**7.4.7 Juror bias measures and decision making in ‘DNA present’ conditions**

Similar to the study described in Chapter six, participants who voted guilty had significantly higher mean final probability of guilt estimates ($M_{\text{robbery}} = 90.24$, $M_{\text{sexual assault}} = 89.62$) than those participants who voted not guilty ($M_{\text{robbery}} = 57.31$, $M_{\text{sexual assault}} = 65.83$), $t = 13.97$, $p < .001$. This makes intuitive sense and replicates the previous murder trial findings, so similar to Chapter six a multiple regression analysis was utilised to determine what items of evidence predicted the final probability of
guilt for the robbery and sexual assault scenarios. The previous chapter indicated that the weak DNA evidence was the only significant predictor of final probability of guilt, so to test this relationship with these data multiple regression was conducted. For these data the assumptions of independent errors ($D-W_{\text{robbery}} = 1.27; D-W_{\text{sexual assault}} = 1.56$), normally distributed errors, linearity, and homoscedasticity were all met and multicollinearity was not a concern ($VIF_{\text{robbery}} = 1.21; VIF_{\text{sexual assault}} = 1.03$).

The results of these regression models are summarised in Table 32 by crime type. These results indicated that in the case of robbery, only the perceived strength of the DNA evidence significantly predicted probability of guilt. In the case of the sexual assault scenario, the strongest predictor of probability of guilt was the perceived strength of the DNA evidence, however the perceived strength of the eyewitness testimony was also a significant predictor in the regression model.
Table 32

*Multiple regression analyses predicting final probability of guilt in the robbery and sexual assault with DNA evidence conditions*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Robbery condition (N=50)</th>
<th>Sexual assault condition (N=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
</tr>
<tr>
<td>Constant</td>
<td>31.43</td>
<td>14.24</td>
</tr>
<tr>
<td>DNA</td>
<td>15.72</td>
<td>2.67</td>
</tr>
<tr>
<td>Eye witness</td>
<td>-2.13</td>
<td>2.82</td>
</tr>
<tr>
<td>Responding police</td>
<td>-1.36</td>
<td>2.38</td>
</tr>
<tr>
<td>Arresting police</td>
<td>-4.00</td>
<td>2.13</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.56</td>
<td></td>
</tr>
</tbody>
</table>

*Note.  * $p < .05$,  **$p < .001$  

In order to determine whether any of the juror attitude measures predicted the perceived strength of the weak DNA evidence, another multiple regression was conducted. The results of the regression analysis are summarised in Table 33. For both crime types, the only significant predictor of perceived strength of the DNA evidence was the participant’s score on the FEEBpp scale, which replicated the findings from the previous murder scenario study.
Table 33

Multiple regression analyses predicting perceived strength of DNA evidence in the robbery and sexual assault with DNA evidence conditions.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Robbery condition (N=50)</th>
<th>Sexual assault condition (N=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
</tr>
<tr>
<td>Constant</td>
<td>3.19</td>
<td>1.35</td>
</tr>
<tr>
<td>FEEBpp</td>
<td>.10</td>
<td>.04</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>-.03</td>
<td>.04</td>
</tr>
<tr>
<td>JBS</td>
<td>.001</td>
<td>.02</td>
</tr>
<tr>
<td>GBJW</td>
<td>-.02</td>
<td>.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.31</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .001

In order to test whether these data fit the mediation model proposed in the previous chapter, three regression equations were calculated following the procedures outlined in Baron and Kenny (1986) and MacKinnon (2008). The standardised beta coefficients for these relationships are given in Figure 11 (for robbery condition) and Figure 12 (for the sexual assault condition), and it can be seen that for both crime types the criteria for a partial mediation have been met (MacKinnon, 2008). Unlike the mediation model tested in Chapter six, the coefficients for the direct path between the FEEBpp score and the final probability of guilt estimate was not reduced to non-significance after partialling out the effect of the mediator variable (perceived strength...
of DNA evidence). The direct effect was however reduced from .80 to .58 for robbery and .68 to .43 for sexual assault, and therefore these models meet the criteria for partial mediation as outlined by Baron and Kenny (1986). Indeed Baron and Kenny argue that requiring the direct relationship between the independent and dependent variables to be reduced to non-significance is not appropriate for most research in the social sciences. In addition, the indirect effect sizes in these models are very similar to that of the model in Chapter four (.22 for the robbery model, .25 for the sexual assault model, compared with .23 for the previous murder scenario model), which provides some evidence for similarity between these models overall.

The Sobel test was used to evaluate the statistical significance of the mediation models for the robbery and sexual assault scenarios (MacKinnon, 2008). An SPSS macro for running a Sobel test of a mediation model is provided by Preacher and Hayes (2004) and the results of the test confirmed a significant indirect effect of the mediator variable on probability of guilt ($Z = 3.17, p < .05$ for robbery; $Z = 2.81, p < .05$ for sexual assault). Both mediation models produced similar effect sizes using Fairchild, MacKinnon, Taborga, and Taylor’s (2009) $R^2$ effect size measure which expresses the amount of variance in final probability of guilt estimates which is accounted for by the mediator variable (perceived strength of DNA evidence); for the robbery scenario $R^2 = .40$ and for sexual assault $R^2 = .34$ both of which represent medium sized effects (Field, 2009).
Figure 11. Partial mediation model for relationship between FEEBpp and final probability of guilt in the robbery condition, including standardised regression coefficients. The standardised regression coefficient between FEEBpp and probability of guilt controlling for perceived strength of DNA evidence is in parentheses.

* coefficient significant at $p < .001$
Figure 12. Partial mediation model for relationship between FEEBpp and final probability of guilt in the sexual assault condition, including standardised regression coefficients. The standardised regression coefficient between FEEBpp and probability of guilt controlling for perceived strength of DNA evidence is in parentheses.

* coefficient significant at $p < .001$
7.4.8 Juror bias measures and decision making in the ‘DNA absent’ conditions

In the current study, conditions with no forensic DNA evidence were included for each crime type in order to determine whether the pro-defence attitudes subscale of the FEEBS contributes to the perception of these versions of the cases. The attitudes and beliefs which are measured in the FEEBpd subscale relate to cases where a potential juror expects there to be forensic evidence, and if this expectation is not met it may result in a decreased likelihood of the juror voting guilty.

Similar to the conditions with DNA evidence, participants who voted guilty had significantly higher final probability of guilt estimates ($M_{\text{robbery}} = 90.00$; $M_{\text{sexual assault}} = 89.62$) compared with those who voted not-guilty ($M_{\text{robbery}} = 49.33$; $M_{\text{sexual assault}} = 57.65$), $t = 10.63, p < .001$ and $t = 10.37, p < .001$ respectively. To investigate what evidence contributed to the perceived probability of guilt, a multiple regression was conducted with perceived strength of the eye-witness, responding police, and arresting police testimony entered as predictor variables, and the final probability of guilt as the outcome variable. The results of the regression analyses for robbery and sexual assault are summarised in Table 34. For these data the assumptions of independent errors ($D-W_{\text{robbery}} = 1.23$; $D-W_{\text{sexual assault}} = 1.68$), normally distributed errors, linearity, and homoscedasticity were all met and multicollinearity was not a concern ($VIF_{\text{robbery}} = 1.53$; $VIF_{\text{sexual assault}} = 1.37$).
Table 34

*Multiple regression analyses to predict probability of guilt estimates in no DNA evidence conditions*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Robbery condition (N=50)</th>
<th>Sexual assault condition (N=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
</tr>
<tr>
<td>Constant</td>
<td>8.95</td>
<td>15.95</td>
</tr>
<tr>
<td>Eyewitness</td>
<td>-6.38</td>
<td>5.17</td>
</tr>
<tr>
<td>Responding police</td>
<td>1.41</td>
<td>4.09</td>
</tr>
<tr>
<td>Arresting police</td>
<td>18.96</td>
<td>4.31</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.34</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *$p < .05$; **$p < .001$*

For both the crime types the only significant predictor of final probability of guilt was the perceived strength of the arresting officer’s testimony. This seems appropriate in these scenarios as this testimony provides details of the only circumstantial evidence for the guilt of the defendant, and therefore in the absence of any physical evidence this would be the most convincing prosecution testimony in these scenarios. In order to determine whether any of the attitude measures predicted the perceived strength of this testimony, a further multiple regression was conducted with the FEEBpp, FEEBpd, JBS, and GBJW scores as predictor variables, and the perceived strength of the arresting officer’s testimony as the outcome variable. The results are summarised in Table 35.
Multiple regression analyses to predict perceived strength of arresting officer’s testimony in the no DNA evidence conditions

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Robbery condition (N=50)</th>
<th>Sexual assault condition (N=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>SE $B$</td>
</tr>
<tr>
<td>Constant</td>
<td>3.79</td>
<td>1.28</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>-.09</td>
<td>.03</td>
</tr>
<tr>
<td>FEEBpp</td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>JBS</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>GBJW</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.37</td>
<td></td>
</tr>
</tbody>
</table>

Note. *$p < .05$

The negative, significant $\beta$ values for the FEEBpd predictor variable indicated that higher scores on the FEEBpd scale were associated with lower strength ratings for the arresting officer’s testimony. So, it would appear that participants who held pro-defence attitudes about forensic science (e.g. that suspects should not be charged/convicted without physical evidence, etc.) were more likely to rate the circumstantial evidence as weak, and thus were more likely to vote not-guilty in both the robbery and the sexual assault versions of the trial stimulus.
In order to investigate whether the relationship between the FEEBpd score and probability of guilt estimates operate in a similar way to the FEEBpp and forensic evidence conditions, a mediation analysis was conducted for both robbery and sexual assault. Only the robbery condition demonstrated a significant partial mediation relationship between the FEEBpd score, perceived strength of the circumstantial evidence, and the probability of guilt ($R^2 = .25$). This relationship is illustrated in Figure 13. The lack of a significant mediation model in the sexual assault (with no DNA) condition may be explained by the relatively small $R^2$ values in the multiple regression tables, compared with the $R^2$ values for the robbery conditions. The smaller amount of variability in perceived strength of evidence and probability of guilt explained by the predictor variables in both regression equations for the sexual assault condition (see tables 34 and 35) explains why a mediation model was not found to be significant for this crime type.
Figure 13. Partial mediation model for relationship between FEEBpd and final probability of guilt in the robbery with no DNA condition, including standardised regression coefficients. The standardised regression coefficient between FEEBpd and probability of guilt controlling for perceived strength of circumstantial evidence is in parentheses.

* coefficient significant at $p < .05$

**coefficient significant at $p < .001$
7.5 Discussion

In this chapter a confirmatory approach to testing the underlying structure of the FEEBS was taken, drawing on the results of the Principal Components Analyses described in the previous studies for empirical and theoretical support. The hypothesised two-factor model of the FEEBS was found to be a good fit to the data collected in this study, which lends support to the use of the FEEBS as a measure of attitudes about forensic science evidence held by members of the public. By confirming this factor structure there is also evidence for the existence of reliable subscales, which represent pro-prosecution (FEEBpp) and pro-defence (FEEBpd) clusters of attitudes about forensic science.

Similar to the findings in the previous chapters, the FEEBpp subscale was significantly, albeit not highly, correlated with participants' scores on the JBS which is reasonable as both scales are scored in the direction of a pro-prosecution bias. However the fact that the FEEBpp and JBS were not highly correlated further supported the argument that they are in fact measuring different attitudes held by participants, which is likely due to the fact that the JBS items refer to general beliefs about the criminal justice system (Kassin & Wrightsman, 1983) while the FEEBpp refers to beliefs about forensic science evidence specifically. The FEEBpp scores were also found to be slightly correlated with GBJW scores, which replicated the studies described in the previous chapters, although the JBS and GBJW were not found to be significantly correlated as they were found to be previously.

In the study described in Chapter six there was no significant correlation found between the FEEBpd subscale and the JBS or GBJW measures, which seemed
surprising as it could be argued that a small negative correlation might be expected. However, such negative correlations were found in this study. It is not clear why these correlations differ between these two studies, although it could be due to the fact that the previous study sample consisted of jury-eligible student participants while the current study sampled jury-eligible members of the public. It is well documented that student samples differ in many respects from community-based samples, and have been shown to display lower correlations between legal attitude measures overall compared with general members of the public or jury samples (Narby, Cutler, & Moran, 1993).

Previously the usefulness of the FEEB subscales for predicting participants’ perceptions of weak DNA evidence in a criminal trial scenario was only tested using a murder scenario, and therefore it was important in the current study to adapt the trial materials for different crime types (robbery and sexual assault). The inclusion of multiple crime types is crucial as some previous juror bias measures have not been shown to have consistent findings across crime types, particularly in the case of sexual assault (Kassin & Wrightsman, 1983, 1988; Lecci & Myers, 2008). In addition trial versions in which no forensic DNA evidence was presented during the trial scenario were included in order to determine whether the FEEBpd subscale scores are associated with perceptions of cases where no forensic evidence is included.

In the robbery scenario which included weak DNA evidence the perceived strength of the DNA evidence was the only significant predictor of probability of guilt estimates made by participants, which is consistent with the previous study and also with the existing literature (Skolnick & Shaw, 2001). For the sexual assault scenario
the DNA evidence was also the most significant predictor of probability of guilt, however the perceived strength of the eyewitness testimony was also found to be a significant predictor. Given that the eyewitness in the scenario was not able to identify the perpetrator, it may seem strange that participants found this testimony to be important in the sexual assault trial. However some research has found that people’s perceptions of the credibility of rape allegations were dependent upon whether a witness could provide evidence that the victim had suffered an attack, for instance by coming across the victim after the attack had occurred (Sleath & Woodhams, 2010). Based on such research it could be the case that participants in the current study associated the eyewitness testimony with increased victim credibility, which in turn may have impacted on their estimates of the probability of the defendant’s guilt.

For both the robbery and sexual assault scenarios the only pre-trial attitude measure which significantly predicted the perceived strength of the weak DNA evidence was the FEEBpp subscale. The relationship was such that the stronger the pro-prosecution beliefs about forensic evidence, as measured by the FEEBpp scale, the higher the strength rating for the weak DNA evidence in the scenario. This finding is consistent with the results in the previous chapter involving the murder trial scenario, and this provides some evidence for the usefulness of the FEEBpp subscale for predicting how mock jurors will perceive weak DNA evidence in the context of a criminal trial. This relationship was shown to operate as a similar partial mediation model to that which was demonstrated in the previous chapter.
In the robbery and sexual assault scenarios in which no DNA evidence was included as part of the prosecution’s case, the testimony from the arresting police officer was the only significant evidence for predicting participants’ probability of guilt estimates. This finding seems reasonable as in the absence of any physical evidence it seems that participants were relying on the circumstantial evidence for their decisions about how likely it was that the defendant committed the crime. The only pre-trial attitude measure which was shown to be associated with the perceived strength of this circumstantial evidence was the FEEBpd subscale. This relationship indicated that the stronger a participant’s pro-defence beliefs about forensic science (e.g. unrealistically high expectations for the presence of forensic evidence in criminal trials) the less likely they were to perceive the defendant as guilty, which seems to correspond to the so-called ‘anti-prosecution’ version of the CSI effect (Podlas, 2006).

The findings for both the DNA present and DNA absent versions of the trial scenarios seem to be consistent with what would be predicted by the CSI effect hypothesis, in that the beliefs held about forensic science predicted how evidence (both forensic and circumstantial) was perceived and in turn the verdict decision made by participants. However, the CSI effect hypothesis would go on to predict that these beliefs come from exposure to crime fiction dramas on television. The failure to find this direct link between television viewing and decision making in the previous literature has led some authors to conclude that no such effect exists (Podlas, 2006; Schweitzer & Saks, 2007; Shelton, Kim, & Barak, 2006).

However, it has been demonstrated here that these beliefs about forensic science did contribute significantly to participants’ verdict decisions, and when
participants were asked to rate the influence of a variety of sources of information on their knowledge about forensic science some interesting findings emerged. In particular, there was a significant positive relationship between the reported influence of TV and movies and scores on the FEEBpp subscale. The results indicated that participants who rated TV and movies as influential for their knowledge about forensic science held stronger pro-prosecution attitudes about forensic science.

Although the relationships which emerged between the reported influence of various types of media and pre-trial attitudes were not of particularly high magnitude, they do make some sense in the context of the CSI effect literature. Forensic science is most often portrayed in TV and movies as being infallible and always conclusive (Tyler, 2006), and indeed these are the beliefs captured by the FEEBpp subscale. However the argument often made in the literature that viewers of crime fiction are more likely to require forensic evidence in order to convict a defendant (e.g. the beliefs measured by the FEEBpd subscale) (Marquis, 2007; Podlas, 2006) was not supported by the current findings. The weak support for a direct link between viewing habits and mock juror decisions in the present study and the wider literature may be reflective of a flaw in methodology requiring participants to correctly identify the source of their attitudes (Cole & Dioso-Villa, 2007).

The rationale for the development of the FEEBS was based on the argument that pre-trial attitudes about forensic science evidence held by potential jurors may have an impact on how ambiguous forensic evidence is perceived in the context of criminal trials. The source of these pre-trial attitudes in the popular media has been the pursuit of the previous CSI effect literature, but is of less concern in the studies
described within this thesis. Although some evidence has been presented which suggests a possible link between the influence of TV and movies and pre-trial attitudes, it seems unlikely that there is a direct relationship to be found. However what has been demonstrated is that the pre-trial beliefs hypothesised by the CSI effect literature can be measured and used to predict mock juror perceptions of forensic evidence in cases where this evidence is of a low probative value.

Although some of the limitations identified in the previous chapter were improved upon in this study (e.g. using a sample from the general public, including a no DNA condition), there are still some methodological issues which limit the applicability of these findings to juries in actual courtrooms. Firstly, the presentation format of the trial materials has been limited in an attempt to control for extraneous variables which might impact on the decision making process, however the result is trial scenarios which are lacking much of the case detail that would be presented to real juries. The use of more realistic material presentation format has been found to impact on participants’ responses (Martin, De La Fuente, De La Fuente, & Garcia, 2007) and so the simplistic format of the materials in the current studies is likely to limit the generalisability of the results.

Secondly, a real jury verdict decision depends on the outcome of group deliberations and in the studies described within this thesis the participants were not participating in group decision making. It is unclear how the process of deliberation might impact on decisions about forensic evidence and verdicts, however for the purposes of this research it is the individual juror pre-trial attitudes which are of interest and for this reason no deliberation has been included in the methodology.
Some previous research has demonstrated that the deliberation process exacerbates the impact of some biases, such as pre-trial publicity (Shaw & Skolnick, 2004), so this would be an interesting area to investigate in relation to the attitudes identified by the FEEBS. This may be of particular relevance as juries may be increasingly likely to be homogenous with respect to attitudes and pre-trial beliefs as a result of peremptory challenges exercised by the prosecution and defence attorneys (Hastie, 1991).
CHAPTER EIGHT: INVESTIGATING THE FEEBS AND NEGATIVE EVIDENCE TESTIMONY

WITH A SAMPLE OF POTENTIAL JURORS

8.1 ABSTRACT

The use of forensic experts called to trial to explain the absence of forensic evidence to the jury (referred to as negative evidence witnesses) is a relatively new trial phenomenon used to counter the hypothesised ‘anti-prosecution’ version of the CSI Effect. In the previous chapter, the predictive validity of the FEEBpd subscale in trial scenarios containing no forensic evidence presented by the prosecution has been tested for robbery and sexual assault crime scenarios. The study presented in this chapter extends this methodology to a murder trial scenario, and also includes a condition which tests the effectiveness of a negative evidence witness. In an attempt to improve the ecological validity of the previous findings, participants in this study were recruited from a jury venire at a County Courthouse in Chicago, Illinois (N = 48). The results indicate that the responses to the FEEBS obtained from this sample were similar to those described in previous chapters, and some similar relationships between the FEEBpd subscale and verdict preferences emerged. There does appear to be some indication that the use of a negative evidence witness may increase the proportion of guilty verdicts (when no physical evidence is presented), although this was not a statistically significant effect. The results are discussed with reference to those presented in the previous chapters, and the implications of the limited sample size are considered.
8.2 INTRODUCTION

In the previous chapters, the factor structure of the FEEBS has been confirmed, and the predictive ability of the FEEB subscales have been tested using three different crime types; murder, robbery, and sexual assault. The results described in the preceding chapters indicated that the pro-prosecution subscale of the FEEBS significantly predicted how weak DNA evidence was perceived by participants, which was related to their verdict preference for all three crime types. The pro-defence subscale of the FEEBS was demonstrated to predict the perceived strength of circumstantial evidence when the prosecution’s case did not include any forensic evidence, and this relationship was true for both robbery and sexual assault scenarios. The predictive ability of the FEEBpd subscale has not yet been tested for a murder case, and this is the main reason for a murder trial being selected for the study described in this final study.

It has been documented that prosecutors are beginning to include so-called ‘negative evidence’ witnesses in cases where there is a lack of forensic evidence to support the prosecution’s case (Trask, 2007). Such negative evidence witnesses are usually forensic science experts who are called by the prosecution to testify in order to explain to the jury why the lack of forensic evidence is not problematic for the case against the defendant, and to manage any high expectations held by jurors about forensic science. No empirical research exists which tests the effectiveness of this type of testimony for reducing unrealistically high expectations for forensic evidence, and so this study aims to evaluate the use of negative evidence testimony for countering the pro-defence bias as measured by the FEEBpd subscale.
The current study also aims to compare the predictive ability of commonly used *voir dire* questions with the FEEB subscales in order to contribute to the current jury selection literature. Similar to the recent use of negative evidence witnesses, prosecutors and defence lawyers are reportedly using specific questions during *voir dire* hearings, which are designed in an attempt to identify forensic science biases in potential jurors (Crocker & Bull-Kovera, 2010). In order to compare the effectiveness of some of these common juror selection questions, this study compares their predictive validity with that of the FEEB subscales.

Finally, this study also seeks to improve the ecological validity of the previous studies in this thesis by recruiting participants from an actual jury venire. Mock juror research is often criticised for relying mainly on student participants or volunteers from the general public (Diamond, 1997), due to possible demographic differences between these samples and actual jurors. By recruiting participants from actual jury venires, this chapter provides an opportunity to compare findings across the other samples used in this thesis and discuss the implications of these findings for future mock juror research.

### 8.3 Method

#### 8.3.1 Participants

Volunteers for this study were recruited from the DuPage County Criminal Courts (Chicago, Illinois, USA) with cooperation from the Jury Administrator and the Chief Judge. In order to ensure that participants’ pre-trial attitudes were being measured, participants were required to be part of the jury venire, but not randomly selected to participate in a trial on the day that they reported for jury service. In an
average week, the DuPage County Criminal Courts summon approximately 400 potential jurors, of which between 200-250 report as directed, and the jury venire consists of all those who report to the courthouse in response to their summons. From this venire, people are randomly selected to participate in *voir dire* questioning, before being selected to form part of a jury on a criminal or civil trial. Participants for this study were volunteers from the pool of potential jurors not randomly selected to participate in *voir dire* proceedings, and were therefore being sent home without actually serving as jurors. Although these participants did not actually go on to be jurors in a criminal trial, they had the same likelihood of being randomly selected for pre-trial questioning as other members of the venire.

In total, 48 volunteers (see section 8.3.3 below for recruitment procedure) completed the online task, which represented an 86% response rate (56 email addresses were collected from potential participants). The sample consisted of 27 males (56.3%) and 21 females (43.7%), and overall 16.7% reported completing high school, 56.2% held undergraduate qualifications, and 27.1% reported postgraduate degrees. The age of the participants ranged from 18-69 years (*M*<sub>age</sub> = 43.35, *SD* = 14.83).

**8.3.2 Materials**

The online presentation of the study materials was developed using Zoomerang, as in the studies described in the previous chapters. Two versions of the murder trial stimulus were developed separately which represent the experimental conditions in this study. This resulted in two URLs which could then be emailed to
participants, and by following the link provided in the email they could get access to the online task.

The trial scenarios consisted of two versions of a murder trial; both trial versions contained no physical forensic evidence, and one condition contained additional negative evidence testimony which was presented by a forensic expert who explained possible reasons for the lack of physical evidence to the jury. The murder scenario used was the same as in the earlier study (see Chapter six), and this trial contains testimony from an eyewitness, arresting police officer, and a pathologist (plus a negative expert witness in one condition). The only evidence manipulated across the conditions was the presence of negative evidence expert testimony (present or absent) provided by the forensic expert. The content of the murder trial stimulus for the two conditions is provided in Appendix F.

Participants initially completed some demographic information followed by the FEEBS, and then some commonly used voir dire questions before moving on to complete the mock juror task (for a list of voir dire questions included, see Appendix F). Once their responses were submitted, the Zoomerang software compiled the data which could then be exported into Microsoft Excel and then SPSS 18 for statistical analyses.

8.3.3 Procedure

Information about the research, and how to volunteer to participate was read aloud by the Jury Administrator to all unselected potential jurors before they were sent home at the end of the jury selection process each day (from 15 Dec, 2010 until 1 April, 2011; excluding the Christmas closure period and weeks with no scheduled jury
trials). Those interested in volunteering for the study were asked to provide an email address, which was then sent by the Jury Administrator to the researcher who then forwarded the link to the online task to volunteers for completion. When volunteers signed up for the study, they were also provided with written information about the research which they could take with them when they left the courthouse.

After the online administration of the FEEBS and voir dire questions, each trial scenario began with a brief summary of the case, after which participants indicated how probably they thought it was that the defendant was guilty of the crime (expressed as a percentage between 0 and 100). Testimony from each of the key witnesses followed on separate pages, with participants indicating the probative value of each piece of evidence on a 5-point Likert scale (ranging from irrelevant/useless to conclusive/definite). After hearing all of the testimony participants read summaries of the case presented by the prosecution and the defence attorneys, after which they again estimated the probability that the defendant committed the crime as well as their verdict decision (guilty/not-guilty).

After completing the online task, participants were directed to a final page which thanked them for their participation and provided a debriefing statement and the researcher’s contact details.

8.4 Results

8.4.1 Comparison of pre-trial questions

In both conditions, participants responded to the FEEBS measure and also to five pre-trial questions which resemble commonly used voir dire questions reported to
be used in the State of Illinois (Judge Schenkier, 2010). The five *voir dire* questions presented to participants were:

1. Do you watch CSI (or similar programmes such as Law and Order, NCIS, Bones, etc.)? (yes or no)
   a. If yes, how many times per week?

2. Forensic television programmes (such as CSI) provide a realistic view of the Criminal Justice System.
   a. Level of agreement on five-point Likert scale

3. How realistic do you think the forensic science techniques are in shows like CSI?
   a. Five-point Likert scale

4. How much of your knowledge about the Criminal Justice System comes from television programmes?
   a. Five-point Likert scale

5. In what percentage of cases would you expect forensic evidence to be part of the prosecution’s case?
   a. Expressed as a percentage from 0-100

The significant Pearson’s correlations between these items are presented in Table 36.
Table 36

*Significant Pearson correlations between responses to pre-trial questions and FEEBS subscales*

<table>
<thead>
<tr>
<th></th>
<th>FEEBpp</th>
<th>Realistic portrayal of CJS?</th>
<th>Realistic forensic science?</th>
<th>Knowledge from TV</th>
<th>Expectation for forensics</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEBpd</td>
<td>-.55**</td>
<td></td>
<td></td>
<td></td>
<td>.37**</td>
</tr>
<tr>
<td>Forensic TV Viewing/week</td>
<td></td>
<td>.71**</td>
<td>.55**</td>
<td></td>
<td>.31*</td>
</tr>
<tr>
<td>Realistic portrayal of CJS?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.30*</td>
</tr>
<tr>
<td>Realistic forensic science on tv?</td>
<td></td>
<td></td>
<td></td>
<td>.32*</td>
<td>.36*</td>
</tr>
<tr>
<td>Knowledge from TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.56**</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .001; N = 48. The items in this table are abbreviated versions of the items presented to participants (see above, or Appendix F for the exact wording).

The majority of participants (72.9%) reported watching forensic-related fiction on television, with the mean number of times per week being 3.11 (SD = 1.80). On average, participants reported that they would expect forensic science to be presented by the prosecution in 61% of criminal cases as indicated by the mean response to this pre-trial question.
8.4.2 Scores on the FEEBS subscales

The mean scores on the FEEBpp and FEEBpd subscales, for the entire sample \((N = 48)\), were \(M = 12.63\) \((SD = 3.27)\) and \(M = 12.75\) \((SD = 3.94)\) respectively. The mean scores on these measures did not differ significantly between the two experimental conditions, as for the FEEBpp \(t = -1.47, p = .15\) and for the FEEBpd \(t = -0.09, p = .93\).

The mean scores on the FEEBS subscales are similar to the results in previous chapters, which are summarised in Table 37.

Table 37

Summary of mean FEEBS scores between studies

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Current study</th>
<th>Chapter 6</th>
<th>Chapter 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jury venire</td>
<td>Students</td>
<td>General public</td>
</tr>
<tr>
<td>(N)</td>
<td>48</td>
<td>159</td>
<td>200</td>
</tr>
<tr>
<td>FEEBpp mean score</td>
<td>12.63</td>
<td>13.50</td>
<td>13.68</td>
</tr>
<tr>
<td>FEEBpd mean score</td>
<td>12.75</td>
<td>11.41</td>
<td>12.76</td>
</tr>
</tbody>
</table>

8.4.2 Evaluation of evidence

In both conditions participants rated the strength of testimony from an eyewitness (the victim’s neighbour), the arresting police officer, and the pathologist. In condition two, participants also heard ‘negative evidence’ testimony from a forensic expert, which explained the absence of forensic evidence in the prosecution’s case. The mean strength ratings for these items of evidence are summarised in Table 38.
Table 38

*Mean evidence strength ratings by condition*

<table>
<thead>
<tr>
<th></th>
<th>EWT</th>
<th>Arresting Officer</th>
<th>Pathologist</th>
<th>Negative evidence witness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>2.43</td>
<td>2.96</td>
<td>2.96*</td>
<td>NA</td>
</tr>
<tr>
<td>Condition 2</td>
<td>2.40</td>
<td>2.48</td>
<td>1.68*</td>
<td>2.96</td>
</tr>
</tbody>
</table>

Note. *indicates means differ significantly at $p < .05$; $N = 48$.

As participants in both conditions read the same testimony (with the exception of the negative evidence witness), there should be no difference in strength ratings between conditions. As indicated in Table 38, there was no significant difference in strength ratings between conditions for the eyewitness and arresting officer testimony, however the strength rating for the pathologist’s testimony did differ significantly between conditions ($t = 4.84$, $p < .05$). It is not clear why participants in condition one rated the pathologist’s testimony as stronger on average than those in condition two.

8.4.3 Probability of guilt estimates and verdict preferences

Hypotheses:

i. In the absence of forensic evidence, participants will rely mainly on the circumstantial evidence presented in the arresting officer’s testimony, as found in Chapter seven.
ii. The perceived strength of the arresting officer’s testimony will be predicted by the participants’ scores on the FEEBpd (negative relationship), similar to the findings in Chapter seven.

iii. The presentation of negative evidence expert testimony will increase the number of guilty verdicts rendered in the absence of physical evidence.

Participants in both conditions were asked to estimate the probability that the defendant committed the murder, before hearing any of the trial testimony or evidence. This prior probability of guilt did not differ on average between conditions ($t = .43, p = .67$), and overall did not differ from a test value of 50% ($M = 47.44, t = -.81, p = .42$) which indicates that participants did not favour the prosecution or defence before hearing the trial evidence. This is consistent with the results presented in previous chapters where members of the general public were recruited as participants.

The distribution of verdicts by condition is displayed in Figure 14, and the results of a $\chi^2$ analysis confirmed that there was no significant association between verdict and condition ($\chi^2 = 2.97, p = .08$). Although there appears to be an association displayed in Figure 14, whereby presenting negative evidence testimony increases the number of guilty verdicts, this association was not statistically significant. It is possible that this non-significant result was due to the relatively small sample size in this study, which resulted in a power of .44 for detecting a medium effect size. This increased potential for a type II error requires these results to be interpreted with caution. The hypothesis was therefore not confirmed.
As in the previous chapters, participants who voted guilty estimated significantly higher final probability of guilt values than those who voted not-guilty (89.68 and 40.22, respectively; $t = -10.78$, $p < .001$). The sample size in this study is not ideal for conducting multiple regression analyses (which was used in the previous chapters), as the minimum required sample described in Tabachnick and Fidell (2007) is $N \geq 50+8m$, where $m$ is the number of independent variables entered into the regression model. A small sample size has a detrimental effect on the power achieved by a regression model, and therefore increases the risk of a Type II error (e.g. not detecting an effect when one exists, due to loss in statistical power). However, multiple regression analyses were run, similar to the previous chapters, in order to
illustrate any potential medium to large effects which might be detected in these data despite the small sample size.

The first multiple regression analysis was used to determine which pieces of trial evidence predicted the final probability of guilt estimates in each condition. The results are summarised in Table 39. Further multiple regression analyses were conducted to determine which pre-trial questions predicted the perceived strength of the significant evidence predictors in each of the conditions, and the results are summarised in Tables 40 and 41. The initial hypotheses were partially confirmed as the arresting officer’s testimony was the only significant predictor of probability of guilt in condition one, however it was the eyewitness testimony and negative evidence testimony which were significant in condition two. The hypothesis that the FEEBpd scores would predict the perceived strength of the arresting officer’s testimony (negative relationship) was confirmed in condition one, and is therefore consistent with the findings reported in Chapter seven.
Table 39

*Multiple regression analysis predicting final probability of guilt*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>No DNA condition (N=23)</th>
<th>No DNA + negative evidence condition (N=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>SE $B$</td>
</tr>
<tr>
<td>Constant</td>
<td>-16.10</td>
<td>18.69</td>
</tr>
<tr>
<td>Eye witness</td>
<td>12.35</td>
<td>7.31</td>
</tr>
<tr>
<td>Arresting officer</td>
<td>8.27</td>
<td>3.46</td>
</tr>
<tr>
<td>Pathologist</td>
<td>7.37</td>
<td>4.57</td>
</tr>
<tr>
<td>Negative evidence witness</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Power (medium effect)</td>
<td>.30</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, ** $p$ < .001
Table 40

Multiple regression analysis predicting perceived strength of arresting officer testimony in no DNA condition

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.43</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>FEEBpp</td>
<td>-.14</td>
<td>.12</td>
<td>-.27</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>-.22</td>
<td>.09</td>
<td>-.63*</td>
</tr>
<tr>
<td>Perceived realism of TV forensics</td>
<td>-.12</td>
<td>.18</td>
<td>-.19</td>
</tr>
<tr>
<td>Expectations for forensic evidence</td>
<td>.32</td>
<td>.22</td>
<td>.39</td>
</tr>
<tr>
<td>Amount of forensic knowledge from TV</td>
<td>.18</td>
<td>.26</td>
<td>.25</td>
</tr>
<tr>
<td>Frequency of viewing forensic TV</td>
<td>-.21</td>
<td>.23</td>
<td>-.29</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power (medium effect)</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * $p < .05$, $N = 23$
Table 41

**Multiple regression analysis predicting perceived strength of negative evidence expert testimony in no DNA plus negative evidence condition**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.93</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>FEEBpp</td>
<td>.25</td>
<td>.10</td>
<td>.75*</td>
</tr>
<tr>
<td>FEEBpd</td>
<td>.23</td>
<td>.10</td>
<td>.73*</td>
</tr>
<tr>
<td>Perceived realism of TV forensics</td>
<td>-.14</td>
<td>.15</td>
<td>-.24</td>
</tr>
<tr>
<td>Expectations for forensic evidence</td>
<td>.02</td>
<td>.16</td>
<td>.03</td>
</tr>
<tr>
<td>Amount of forensic knowledge from TV</td>
<td>.02</td>
<td>.13</td>
<td>.04</td>
</tr>
<tr>
<td>Frequency of viewing forensic TV</td>
<td>.10</td>
<td>.15</td>
<td>.18</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power (medium effect)</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. * $p < .05$, $N = 25$
8.5 DISCUSSION

The first aim of the study described in this chapter was to extend the findings in Chapter seven (which included robbery and sexual assault scenarios, with no prosecution forensic evidence) to a murder trial scenario in which the prosecution did not present any forensic evidence as part of their case. The results for the murder scenario in condition one replicated the previous findings and confirmed the hypothesis, as in the absence of physical evidence participants appeared to be relying on the circumstantial evidence presented in the arresting officer’s testimony when deciding on the probability of the defendant’s guilt. As in the previous chapter, in the current study the perceived strength of the arresting officer’s testimony was significantly predicted only by the participants’ scores on the FEEBpd subscale, which confirmed the hypothesis. This relationship was negative, indicating that participants who demonstrate unrealistically high expectations for forensic evidence (as measured by the FEEBpd) are less likely to rate the circumstantial evidence as compelling, and therefore are more likely to vote not-guilty.

One novel aspect of the design in the current study was to include pre-trial questions similar to those used in voir dire hearings in the State of Illinois (where the jurors were recruited) in order to compare these questions with the attitudes measured by the FEEBS. The five pre-trial questions included asking participants whether they watch forensic-related television programmes (and if so, how often), how realistic they believe the portrayal of the criminal justice system and forensic science to be in such programmes, and in what proportion of cases they believed that forensic science would be included in the prosecution’s case. As these types of
questions are routinely used in pre-trial hearings, based on the belief that they predict a bias as hypothesised in the *CSI Effect* literature, it was considered important to compare the predictive validity of such questions with the attitudes measured by the FEEBS.

The results suggested that high expectations for the presence of forensic evidence in criminal trials, was associated with increased viewing of forensic-related television. In addition, high levels of forensic television viewing was associated with increased ratings of the accuracy of the forensic science portrayed in such shows, and a belief that the criminal justice system is depicted realistically in television fiction. This finding replicates the results in Shelton et al. (2006), who found that participants who reported watching CSI regularly, rated it as more realistic than those who did not watch such shows. It is interesting to note that the FEEBpp subscale did not correlate with any of the other pre-trial questions (apart from a negative correlation with the FEEBpd), which suggests that the attitudes measured by this subscale are sufficiently different from those measured by commonly used *voir dire* questions. The FEEBpd subscale was only significantly associated with the expectations for forensic evidence question, which seems reasonable as this is the underlying concept being measured by the FEEBpd items.

A second original contribution made in this study is the inclusion of negative evidence expert testimony by the prosecution, in the second condition. This type of testimony aims to explain to the jury the absence of forensic evidence in the prosecution’s case, and is thought to reduce the impact of unrealistic expectations held by jurors. As it has already been demonstrated that high expectations for
forensic evidence are measured by the FEEBpd (see previous chapters for results of factor analysis, and correlations reported in this chapter), this provides an opportunity to determine whether this type of testimony had an impact on the perception of a case and the subsequent verdict preference. Although there was no statistically significant association between verdicts and condition, in the condition with no negative evidence testimony the majority of participants voted not-guilty (61%), while in the condition which included negative evidence testimony participants who voted not-guilty were in the minority (36%). There is a risk that the lack of statistical significance may have been due to the reduced statistical power resulting from a limited sample size, however the hypothesis was not confirmed by this analysis.

Of particular interest is the finding that the most significant predictor of probability of guilt in the second condition was the perceived strength of the negative evidence testimony. Participants who perceived this testimony as strong were more likely to report a high probability of guilt and therefore vote guilty. Both the FEEBpp and FEEBpd scores predicted the perceived strength of the negative evidence testimony, and both relationships were positive. First, if we consider the relationship between the FEEBpd and perceived strength of the negative witness testimony, this seems to suggest that this type of testimony might be advantageous to the prosecution, as participants with higher FEEBpd scores find this evidence to be particularly compelling which leads them to be more likely to convict. This is contrary to the previous findings presented in this thesis in which high FEEBpd scores were associated with higher frequency of acquittals.
The relationship between the FEEBpp subscale and the perceived strength of the negative evidence testimony may also be explained by the attitudes being conceptualised by this subscale. The FEEBpp measures attitudes about the infallibility and conclusiveness of forensic science, which in the previous chapters has been demonstrated to be associated with higher perceived strength of weak forensic evidence. It could therefore be the case that when no physical forensic evidence is presented by the prosecution, then these attitudes have a similar impact on the perceived strength of the negative evidence witness (who is usually a forensic science expert) and therefore this witness is considered compelling enough to contribute to a guilty verdict.

There were some unexpected results in the current study, which are important to mention and attempt to investigate with further research. Firstly, it is not clear why the participants in the first condition considered the pathologist’s testimony to be significantly more probative on average than participants in the second condition. As the only evidence manipulation was the presence of negative evidence testimony, differences in the perceived strengths of the other evidence types were not anticipated and could not be explained by the analysis of these data. There was a similar anomaly reported in Chapter seven, in which some of the evidence types were perceived differently between conditions, however this was thought to be potentially explained by the fact that the conditions differed in crime type (robbery or sexual assault). This explanation would not be relevant in this study, as both conditions contained exactly the same offence of murder.
A second unexplained difference in the perceived probative value of evidence is seen in the regression analysis in which the significant predictors of final probability of guilt differ between conditions. In the first condition only the arresting officer’s testimony predicted final probability of guilt, which is consistent with the findings in Chapter seven. However, in the second condition the eyewitness testimony was also a significant predictor, along with the negative evidence testimony, while the arresting officer’s testimony was not significant in this condition. This appears similar to results in the sexual assault condition in Chapter seven in which the eyewitness was an unexpected significant predictor (which was not the case in the robbery condition). However, in Chapter seven this was thought to be related to previous research which suggests that corroborating evidence (e.g. an eyewitness) is considered by jurors to be an important factor in determining victim credibility in sexual assault cases (Sleath & Woodhams, 2010). It therefore remains unclear why this effect was found in the current study. Further research should aim to test whether this was an isolated finding, or whether there are other factors which might explain the results.

Although this study aimed to improve the ecological validity of the research in this thesis by recruiting participants from a real jury venire, this created a limitation in the current study which may have had an impact on the results. Due to the significant difficulties negotiating access to these participants, we were only granted a limited amount of time for the recruitment of volunteers. Coupled with the fact that we had to rely on the courthouse’s jury administrator for the advertising of the study and collection of participants’ email addresses, the result was a sample of 48 participants who completed the online task. As a result, the statistical analyses presented within
this chapter are suffering from low statistical power, and therefore the conclusions need to be considered with caution. Further research should be conducted with the aim of securing a larger juror-based sample in order to attempt to replicate the current findings. It is hoped that this will be achieved in part due to a research travel grant recently awarded to the researcher, which will facilitate the dissemination of these findings to the Chief Judge at DuPage County Courts with the aim of planning future, larger scale research with jury venire members.
CHAPTER NINE: OVERALL DISCUSSION

This final chapter considers the findings of this thesis in relation to the previous literature and theoretical perspectives on juror decision making, as well as comparing the results between chapters which will allow some comments to be made about the use of different participant samples in mock juror decision making research. The strengths and weaknesses of the research methodology will be discussed, and the theoretical and practical implications of the FEEBS will be considered. Finally, the chapter will conclude with recommendations for the directions of future research.

9.1 CONSIDERING THE FINDINGS IN THE CONTEXT OF PREVIOUS LITERATURE AND THEORY

The most widely accepted cognitive models of juror decision making emphasise the importance of jurors’ prior knowledge and beliefs and the influence these prior beliefs may have on the evaluation of the information presented during criminal trials (Pennington & Hastie, 1986, 1988; Wagenaar, et al., 1993). The aim of Chapter three was to assess participants’ existing knowledge about various forensic techniques (referred to as forensic awareness), and the results demonstrated wide variability in participants’ levels of forensic awareness. In particular, many of the erroneous beliefs held by participants were the result of over-estimations concerning the ability of forensic science techniques, which is hypothesised as being problematic in the CSI effect literature (Cole & Dioso-Villa, 2007; Marquis, 2007; Tyler, 2006). Although the results of study one did not find a relationship between self reported sources of forensic knowledge (e.g. tv/movies, internet, etc.) and forensic awareness, as would
be predicted by the CSI effect literature, the accuracy of participants’ knowledge about forensic science was also not related to their educational level, age, or gender.

In addition to assessing forensic awareness, Chapters three and four also tested whether participants could differentiate between forensic evidence of varying probative value. Previous literature focusing on whether mock jurors understand the relative contribution of different forms of forensic evidence has overwhelmingly concluded that this is a difficult task which people are generally ill-equipped to carry out (Faigman & Baglioni, 1988; Goodman, 1992; Schklar & Diamond, 1999). Somewhat contrary to these previous findings, the results presented in Chapter three suggested that participants were able to distinguish between the four strengths of forensic evidence presented to them.

There are two explanations for the discrepancy between the previous literature and the results presented in Chapter three; firstly, previous research has focused on the ability of participants to correctly assess probabilistic testimony (e.g. Faigman & Baglioni, 1988; Goodman, 1992; Nance & Morris, 2005; Schklar & Diamond, 1999, etc.) and in Chapter three there was no manipulation of probability or frequency information given to participants. Secondly, the stimuli used in Chapter three did not involve case details (which was usually included in previous research), and therefore participants were assessing the evidence in the absence of trial context. Although this is obviously not a realistic way to present evidence in a mock juror paradigm, it allowed for some conclusions to be drawn about the impact of context on the decision making process.
As predicted by the narrative models of juror decision making, Chapter four demonstrated an effect of case context on participants’ evaluations of the probative value of forensic evidence. The introduction of case context resulted in inflated estimates of evidence strength, particularly in cases where the forensic evidence was of a weak or moderate probative value. In the absence of context participants correctly identified weak forensic evidence as such, however in the context of a criminal case the same evidence was rated as significantly stronger on average. This finding is supported by existing theory as cognitive models predict that the effect of a narrative process results in evidence being evaluated not in isolation, but with reference to the contextual information provided by the prevailing story of events (Pennington & Hastie, 1988; Huntley & Costanzo, 2003). Indeed, contextual effects have been previously documented in relation to experts’ evaluation of forensic evidence (Dror, et al., 2006) so it seems reasonable to assume that laypersons (e.g. jurors) would exhibit similar effects.

The fact that the effect of context was most pronounced in the weak evidence condition, and this condition produced the most disagreement between participants, is potentially explained by the influence of prior attitudes and beliefs. Interactionist theories of juror decision making would predict the results found in Chapter four, as the experimental condition with the most variability in perceived evidence strength and verdict was the condition in which weak forensic evidence was presented by the prosecution. As outlined in the previous literature, heuristics and biases are most likely to influence decisions when the situational cues (in this case the trial evidence) are not particularly salient (Kaplan & Miller, 1978; Kassin & Wrightsman, 1988; Bull
Kovera, et al., 1999). This finding in Chapter four, and its congruence with the previous literature and decision making theory, formed the basis for the construction of the FEEBS in an effort to identify and measure pre-trial attitudes which might predict decision making in cases with weak forensic evidence.

This attempt to predict juror decisions using attitude measures has been the pursuit of previous literature for decades, and one of the most robust findings is that general attitudes are not reliable predictors of juror decision making (Sealy, 1981; Visher, 1987; Moran, et al., 1994), while more legally relevant attitudes demonstrate greater predictive validity (Narby, et al., 1993; Moran & Comfort, 1982; Kassin & Wrightsman, 1983). This finding was confirmed by the results presented in Chapters five through eight of this thesis, as the FEEBS measure was a significant indirect predictor of verdict preference, while more general attitude measures (the JBS and GBJW) and commonly used *voir dire* questions did not predict participants’ decisions. These findings suggest that when the prosecution’s case relies on forensic evidence which is of weak probative value, then jurors’ pre-trial attitudes about forensic science may predict how strong they perceive the prosecution’s case to be, which in turn predicts their verdict decision.

Although the studies described in this thesis have demonstrated that certain attitudes about forensic science significantly predict juror perceptions of evidence strength and verdict preferences, there are likely to be many other factors contributing to the decision making process. Consideration of these other factors should be pursued in future research in order to further improve the model of juror decision making about cases, which will be discussed further in section 9.5.
9.2 Theoretical and Practical Implications of the FEEBS

This section will consider how the FEEBS measure was developed with reference to the previous literature, and how the predictive validity of this scale might be explained by existing theory. In addition, the potential practical applications of the FEEBS will be discussed with reference to the use of peremptory challenges in some jurisdictions, and other tactics such as negative evidence witnesses.

9.2.1 Theoretical Basis for the FEEBS

In an effort to determine whether prior beliefs and attitudes about forensic science could partially explain the wide variation in participants’ perceptions of weak forensic evidence found in Chapter four, the remainder of the thesis focused on developing a measure of such attitudes and testing its predictive validity. The theoretical support for the development of the FEEBS items was derived mainly from the anecdotal claims made in the CSI effect literature concerning potentially problematic attitudes held by jurors about forensic science (Podlas, 2006; Tyler, 2006; Cole & Dioso-Villa, 2007). The factor analytic results reported in Chapters five through seven suggest that there are two reliable, distinct clusters of attitudes about forensic science being measured by the FEEBS, which correspond conceptually to the two main effects hypothesised by the CSI Effect literature.

Across all of the crime types used in this thesis (murder, robbery, and sexual assault) the subscales of the FEEBS significantly predicted verdict preferences through the mediating variable ‘perceived strength of evidence’. The two subscales of the FEEBS were demonstrated to play different predictive roles depending on the presence or absence of forensic evidence as part of the prosecution’s case.
Specifically, unrealistically high expectations for the presence of forensic evidence (as defined by high scores on the FEEBpd subscale) predicted not-guilty verdicts when there was no forensic evidence presented by the prosecution. Conversely, attitudes associated with exaggerated faith in the capabilities of forensic science (defined by high scores on the FEEBpp subscale) predicted guilty verdicts when the prosecution presented weak forensic evidence. These findings confirm what the CSI Effect literature suggests anecdotally about juror attitudes and verdict preferences, and what the previous empirical research has failed to demonstrate. The systematic manipulation of evidence presence and probative value, and the focus on attitude measures as the independent variable in this thesis seem like the most likely explanations for these novel findings.

Despite the fact that previous literature, which has attempted to empirically confirm the existence of the CSI effect, has failed to find a direct link between media consumption and juror behaviour, the results presented in this thesis provide some support for the existence of the hypothesised attitudes and their indirect impact on verdict preference. There are a number of differences between the approach taken in this thesis and the previous literature, which may explain these findings. Firstly, the focus on attitude measurement as the predictor variable rather than viewing habits (as often used in the existing literature) approaches the research question from a more psychologically sound framework (Ajzen & Fishbein, 1977). Secondly, previous research in the attitude-behaviour literature suggests that trial stimuli should contain ambiguous or weak evidence in order for pre-trial attitudes to have a measurable impact on behaviour (Kaplan & Miller, 1978; Ruva et al., 2007; Devine et al. 2009).
Some of the previous literature has not considered this in the design of the trial material, and this may partially explain the lack of an effect found in such studies.

Finally, the approach taken in this thesis to analyse indirect effects of attitudes on verdict preference has been suggested by some authors to be a more theoretically-grounded approach (Kim et al., 2009; Lieberman et al., 2008), particularly as the so-called CSI effect is thought by some to be more likely the result of a general popular culture phenomenon, rather than the result of particular television viewing habits (Tyler, 2006). Social psychological research has historically recognised the importance of mediating variables in explaining the mechanism by which an independent variable affects a dependent variable, and this has been a particularly favoured approach in research investigating the relationship between attitudes and behaviour (Baron & Kenny, 1986). Indeed throughout this thesis a significant partial mediation model has been used to explain the relationship between the FEEBS subscales, perceived strength of evidence, and probability of guilt estimates. The only exception was in Chapter seven in the condition containing a sexual assault scenario with no forensic DNA evidence presented by the prosecution. In this condition the mediation model was not found to be significant, and the reason for this anomaly is not entirely clear. However this is not the first time that sexual assault scenarios have resulted in inconsistent results (e.g. see Kassin & Wrightsman, 1983), and this could be due to other salient attitudes about this specific crime type which are operating alongside those being measured in these studies.
9.2.2 Practical implications of the FEEBS

Despite the fact that the majority of previous empirical research has concluded that the *CSI Effect* was not operating as hypothesised, reports of courtrooms responding to the perceived threat of the *CSI Effect* have been documented. In a survey of attorneys conducted by Robbers (2008), 85% of respondents stated that the way they conduct their job had changed in response to the perceived *CSI Effect*, including an increased use of negative evidence expert witnesses (23%), additional time spent on conducting *voir dire* hearings (36%), and an increasing proportion of their communication to jurors spent explaining the absence of forensic evidence (67%). The usefulness of these reported techniques for overcoming juror bias about forensic evidence is yet to be tested in the literature, although the FEEBS may be a preliminary step toward understanding relevant juror attitudes and the impact that various legal tactics may have on reducing any existing biases.

As discussed in Chapter one, decades of juror decision making research and theory has demonstrated that jurors do not arrive at court without prior beliefs and attitudes which are very likely to contribute to their evaluation of the trial evidence (Pennington & Hastie, 1986). Given that this is the case, some jurisdictions attempt to identify potential biases and problematic pre-trial attitudes through the legal mechanism of peremptory challenges (Brooks, 2004). In particular, in the United States, the process of *voir dire* hearings can be very lengthy and resource intensive with some lawyers believing that cases are won or lost at this pre-trial stage (Fulero & Penrod, 1990). Although it is beyond the remit of this thesis to argue the philosophical debate surrounding the legal ethics of juror exclusions, given that this tactic is
commonly used in some countries this thesis does argue that the basis on which these exclusions are made can be informed and improved by social psychological research.

Relying on “jury selection folklore” (Fulero & Penrod, 1990, p. 229) as the basis for peremptory challenges, rather than empirical evidence, is likely to result in unjustified juror exclusions and undermines the safeguards which are theoretically in place to avoid discriminating against jurors on the basis of race, gender, and other socio-economic characteristics (Norton, et al., 2007). Of particular relevance to this thesis, the empirical literature would suggest that asking jury venire-persons about their television viewing habits is unlikely to predict any bias or verdict preference, and should therefore not form the basis of excluding a juror. However this thesis has demonstrated that there may be some attitudes which do predict juror decisions in cases with certain types of evidence, and therefore the FEEBS might be a more empirically justified set of questions to use in pre-trial hearings for relevant trials.

Some jurisdictions (such as England and Wales) have much more stringent restrictions in place regarding pre-trial juror questioning and exclusion, and so the ability to overcome potential juror bias or pre-trial attitudes relies on other methods in the courtroom. One such method for overcoming the high expectations for forensic evidence (as measured by the FEEBpd) is to introduce negative evidence witnesses to explain to the jury why no forensic evidence was recovered from a particular crime scene. The effectiveness of this type of testimony was tested in Chapter eight, and found to have a potential effect on the proportion of guilty verdicts. In particular, participants who heard negative evidence testimony as part of a murder trial convicted more often than those who heard the same murder trial with no such
testimony. However, the association between verdict and condition was not found to be statistically significant, although this may have been the result of the small sample size of real jurors.

If we cautiously interpret the results in Chapter eight as supporting the effectiveness of using negative evidence experts to counter the unrealistic expectations for forensic evidence held by some jurors, further research is needed to investigate the practical implications of these results. Prior to this research, the empirical evidence suggested that the CSI Effect claim that jurors hold these high expectations was not entirely supported, however the FEEBpd appears to be a more reliable way to measure these types of attitudes. Having demonstrated within this thesis that such attitudes exist and can be used to predict verdicts in cases with no prosecution forensic evidence, perhaps the use of negative evidence witnesses could be an effective method for managing these expectations held by jurors.

9.3 COMPARISONS OF DIFFERENT SAMPLES ACROSS STUDIES

Mock juror research is often criticised in the literature for recruiting participants who are not sufficiently similar in demographic characteristics to members of actual juries (Diamond, 1997). Substantial differences between mock juror samples and actual jury members may limit the ecological validity of research findings, however the difficulties inherent in accessing actual jury members results in the continued use of somewhat dissimilar mock jurors as research participants. The studies reported within this thesis have intentionally recruited different groups of mock juror participants across the chapters, in order to enable some conclusions to be
made about the extent to which these samples differ from one another in terms of the key variables measured in the studies.

The most widely used sample population in psychological research generally, and mock juror research specifically, is the undergraduate student population. This type of sample is certainly a result of convenience for academic researchers, however it is also argued to be the most significantly different from real jurors, which often attracts criticism in the literature (Bray & Kerr, 1982). In Chapter six of this thesis, 159 undergraduate students acted as mock jurors and, similar to previous research, the average age of the sample was substantially lower than the average age in the general population or real juror samples reported in the rest of the thesis, despite the fact that a similar overall range of jury-eligible ages was represented in all samples. The student sample was also slightly more biased toward having a majority of female participants compared with the general population samples, although the real juror sample in Chapter eight contained a much more balanced gender split. Finally, the student sample clearly did not have the range in educational backgrounds exhibited in the general population and jury samples, due to the fact that the students were all at the same stage in their academic careers.

Despite the demographic characteristics in the samples described above, what is of particular interest is whether the samples differed substantially in their responses to the variables measured during the present research. Previous research has suggested that student and non-student mock juror samples may only differ significantly in as few as 19% of relevant published research (Bornstein, 1999) in terms of verdict decisions. In the current thesis, one of the measures that did demonstrate a
difference by sample was the prior probability of guilt estimates. For all of the non-student mock juror samples reported in the current research, the prior probability of guilt estimates did not differ significantly from a test value of 50%, however the student sample did differ significantly with a mean of 66%. There is previous research to suggest that student populations are more prone to guilty judgements in cases with ambiguous evidence (Martin et al., 2007), which may explain this finding within this thesis.

The final probability of guilt estimates given by participants who voted guilty appear to show some difference between participant samples, with the student sample averaging 80% while the non-student samples averaged 90% probability of guilt for guilty verdicts. What was also inconsistent between the samples was the average probability of guilt estimates for the participants who voted not-guilty. The average student probability was 57% for not-guilty verdicts, while general members of the public estimated a mean of 62% and real jurors reported a mean probability of 40% for not-guilty verdicts. Regardless of these apparent differences in the mean probability of guilt estimates, similar proportions of participants in each sample voted guilty, on average 44%, which indicates that regardless of the sample the cases were indeed perceived as ambiguous.

Another comparison that can be made between samples is the scores on the attitude measures used in these studies. The FEEBS subscales were used with all three sample types, and resulted in very consistent mean scores across sample types. The student sample scored slightly lower on the FEEBpd subscale compared to the non-student samples ($M = 11.41$ versus $M = 12.76$, respectively), however this does not
appear to constitute a problematic difference in mean score. More important to the methodological debate is the fact that the real juror sample did not differ substantially on their attitudes about forensic science as measured by the FEEBS. Similarly the JBS and GBJW scores can be compared between the student and general population samples, and these are also very similar indicating that these personality characteristics were not substantially different between students and general members of the public. This finding could not be extended to the real juror sample, as these participants were not required to complete these measures.

In conclusion, although there were significant differences in average age and educational level between the student and non-student samples, and a clear gender bias in every sample except the real juror sample, the verdict results across the studies were remarkably similar. Of particular importance for the studies in this thesis, the FEEBS factor structure was identical for students and non-students, and the relationship between pre-trial attitudes, perceived strength of evidence, and probability of guilt was consistent across the studies regardless of the sample. It would therefore appear that the sample recruited from the jury venire did not display different attitudes about forensic science, despite the demographic differences in this sample, which provides some support for research conducted using mock juror participants from the general public. Although real jury participants should perhaps always be the goal of this type of research, it appears that the significant access issues may be safely overcome by using less-realistic samples for mock jury research.
9.4 STRENGTHS AND WEAKNESSES OF THE RESEARCH

This section will consider some of the strengths and weaknesses of the research contained within this thesis, and consider how these methodological issues may have affected the results and the ecological validity of the findings.

9.4.1 STRENGTHS OF THE RESEARCH DESIGN

As described in Chapter two, there is extensive debate in the literature about the validity of mock jury research and in particular the use of participant samples which do not adequately resemble people who sit on real juries (Diamond, 1997; Bray & Kerr, 1982). Despite this debate, there are few studies which have incorporated a range of sample types in order to directly compare the performance of various types of mock jury sample. It is therefore considered a strength of the research within this thesis, that volunteers were recruited from a range of populations including students, jury eligible members of the public, and a real jury venire. As discussed in the previous section, this allowed a comparison of these samples on some of the relevant measures, which led to the conclusion that despite some demographic differences between samples the verdict decision process appeared to be largely consistent.

Perhaps the most significant strength of the present research is the integration of various existing mock juror methodologies and theoretical perspectives in order to contribute a more robust psychological framework to the existing literature investigating the impact of pre-trial attitudes about forensic evidence on juror decisions. The vast majority of the previous literature investigating the CSI Effect has not been grounded in psychological theory and methodology, resulting in many of these studies being designed far too simplistically, which has not allowed for sufficient
understanding of the cognitive mechanisms underlying the results. The use of multiple dependent variables, such as verdict and probability of guilt estimates, in this thesis is the result of combining the existing CSI Effect literature with methodology commonly used in previous mock juror research concerning decisions about forensic evidence (e.g. probability theory approaches). The outcome of this more comprehensive approach to the collection of data is the ability to conduct more meaningful statistical analyses, which allows for a better understanding of the decision making processes at work in these studies.

Another strength in the research approach taken within this thesis is the focus on indirect effects of attitudes on juror behaviour, rather than the pursuit of direct influences of television viewing on behaviour which seem to dominate the existing CSI Effect literature. This approach was advocated by Kim et al. (2009) who conducted the first CSI Effect study to investigate the interaction between variables using path analysis. Indeed some authors argue that most psychological enquiries would benefit from an indirect effect approach to analysis, as there are unlikely to be many truly direct effects of cognition or attitudes on observable behaviour (MacKinnon, 2008). Therefore the approach taken in this thesis to investigate the indirect effect of attitudes on juror decisions, via the mediator variable perceived strength of evidence, provides a more psychologically sound explanation for the relationship between these variables.

Finally, this thesis has attempted to shift the focus from identifying particular television programmes which are associated with jurors’ beliefs, to instead measuring the attitudes themselves rather than their hypothesised source. This thesis argues
that this is a more psychologically robust approach compared to measuring television viewing habits as the independent variable, as it is not the television shows viewed which are likely to impact directly on behaviour, but rather the internalisation of the themes and attitudes portrayed in the media about forensic science which are most likely to predict juror behaviour. Indeed this argument is supported throughout the thesis, as the only reliable predictors to emerge across the studies are the attitudes measured by the FEEBS.

9.4.2 Weaknesses of the Research Design

The most relevant limitations inherent in research employing a mock juror paradigm have been outlined in Chapter two, and the aim of the present section is to consider these weaknesses in the context of this thesis, and how some of these might be addressed in future research.

Mock juror participants are asked to assume the role of a juror and make decisions which are designed to simulate those which occur in real criminal cases involving real jurors. There are a number of limitations in these studies which threaten the validity of the findings and the ability to generalise these to situations involving actual jurors. The first is the use of simulated trial transcripts for the presentation of the evidence and relevant case material to participants. The fact that the studies within this thesis used written testimony excerpts, presented in an online format, is obviously significantly different from the way information is communicated to real jurors in trials. Although some authors argue that most studies do not find a main effect of presentation medium on mock juror decision making (Bornstein, 1999), the drastic deviation from how trials are conducted in real life requires researchers to
consider their results cautiously, and seek to improve the realism of the trial simulation in future studies.

Apart from the presentation style differences between the studies in this thesis and real trial situations, there remains a more serious issue which is practically impossible to overcome in the mock juror research paradigm. In real trials, the decisions made by juries have serious and life-changing consequences which cannot be replicated in any mock juror experiments. It is not known precisely how these lack of consequences in mock juror studies impact on the decision making process, but it is important to not lose sight of this major practical issue when interpreting mock juror research findings.

Another issue of ecological validity which is inherent in the studies presented within this thesis is the lack of jury deliberations in these studies. Deliberations were not conducted in this thesis, as the aim of the studies was to investigate the impact of attitudes on individual juror perceptions of case evidence. However, it is acknowledged that in real cases verdicts are not reached based on individual juror perceptions, although these will play a role in the deliberation process. Jury deliberation in real criminal cases is a crucial element to consider in mock jury research, therefore the findings presented in this thesis should not be generalised to predictions about actual post-deliberation verdicts in real case scenarios. It is therefore important that future research attempts to investigate the interaction between the attitudes discussed within this thesis and the deliberation process. In addition, comparisons of mock juries consisting of homogenous and heterogeneous jurors with respect to attitudes measured by the FEEBS would shed some light on how
these attitudes might determine the outcome of group decision making about cases involving weak or absent forensic evidence.

With regard to the trial scenarios used in these studies, they have tested the predictive validity of the FEEBS in cases with only weak forensic DNA evidence. The decision was made to proceed with only DNA evidence in order to hold the type of evidence constant throughout these initial FEEBS studies. However there are many other, commonly used sources of forensic evidence (such as those outlined in Chapter three, e.g. fingerprints, ballistics, hairs and fibres, etc.) which should also be investigated in relation to the FEEBS. Therefore future research should be conducted which includes different types of forensic evidence across different crime types in order to determine whether the predictive ability of the FEEBS is consistent across different evidence scenarios.

Finally, although it was considered a strength of this thesis to include a variety of sample types across the various studies (e.g. students, non-students, jury venire persons), it is also important to note that there may be cross-cultural issues which have not been investigated directly in the analyses presented. In particular this refers to the fact that the final chapter presents findings from a sample of jury venire members recruited at a courthouse in Chicago, Illinois, whereas the rest of the samples consisted of jury-eligible members of the UK population. It would have been ideal to have had a jury sample from a court in England and Wales in order to maintain consistency in the country of residence of participants. However, despite attempts to access such a sample this was not achieved. Therefore the opportunity to access a
jury sample in Chicago was considered too valuable to dismiss, although it is acknowledged that this may have introduced cross-cultural confounds to the research.

9.5 Future directions

The development and initial testing of the FEEBS presented within this thesis represents what may well be only the early stages of such research. There are a number of different directions that future research could take to further investigate this topic. As mentioned in the previous section further studies utilising trial scenarios with various types of forensic evidence (e.g. fingerprints, ballistics, footwear impressions, etc.) would contribute to determining how well the attitudes measured by the FEEBS generalise to predicting decisions about a wide range of sources of evidence. In addition, and also mentioned in the previous section, studies involving the deliberation of mock juries would assist in better understanding how these attitudes might impact on the group decision making of actual juries in real trials.

The findings in Chapter seven produced some inconsistencies regarding the sexual assault scenario. This seems consistent with some previous research which has attempted to predict juror behaviour in rape trials using pre-trial attitude measures (e.g. Kassin & Wrightsman, 1983; Lecci & Myers, 2002). Such findings provide justification for further research into the relationship between juror attitudes and decision making in sexual offence scenarios, and perhaps more sophisticated statistical modelling techniques could be employed to better understand the causes of these apparently inconsistent research findings. For example, by incorporating measures such as ‘levels of rape victim blaming’, or ‘rape myth acceptance’ into the research design it may be possible to identify some interactions between these types of
attitudes and those measured by the FEEBS which may better predict verdict preference in these types of scenarios.

Another way to utilise more complicated statistical models to further understand juror verdict preferences is to consider how other personality constructs may have an impact on perceptions of forensic evidence. For example, in a very recent study by Mancini (2011) need for cognition (NC) was proposed as an important psychological element of the juror decision making process. In particular that study investigated whether NC interacted with viewing habits of forensic television programmes to predict verdicts in a sample of mock jurors. The author concluded that high NC jurors are more likely to find support for their initial hypothesis in ambiguous evidence, however it is not clear how the ambiguity of the forensic evidence was defined or manipulated. In terms of future directions for research into the FEEBS and juror decisions about forensic evidence, it may be beneficial to consider NC as a possible moderating variable.

Potential future research with the FEEBS is not limited strictly to juror decision making, and particularly in more inquisitorial legal systems it is possible that expert decision makers (e.g. panels of judges) may also hold some of the attitudes measured by the FEEBS which may have an impact on their decision making processes. This may be a particularly important area to investigate, as it is often the case that these expert decision makers play a more active role in seeking out forensic evidence, as well as deciding on its admissibility, and therefore, unlike jurors, they are increasingly likely to be exposed to very weak or ambiguous forensic evidence (which may not be deemed admissible and heard by a jury in an adversarial courtroom) (Broeders, 2003). It may
therefore be particularly interesting to determine how their attitudes about forensic science evidence (as measured by the FEEBS) might impact on their expert decision making process.

It is also possible that the FEEBS might be useful in understanding the decision making processes of victims of crime, in particular their reported levels of satisfaction with the police response to their reported victimisation. Citizens play a crucial role in law enforcement, and without some level of public confidence in the police it is likely that there would be very limited exchange of information which is critical if crimes are to be reported, responded to, and solved effectively (Brandl & Horvath, 1991). It is therefore important that police agencies understand the factors contributing to victim satisfaction, and previous research suggests that the prior expectations of the victim play a significant role in predicting their level of satisfaction with the police (Coupe & Griffiths, 1999). This might be best explained theoretically by the concept of expectancy disconfirmation, mainly reported by marketing and organisational psychology literature (Oliver, 1980).

Expectancy disconfirmation is rooted in social psychology and organisational behaviour theory, and has been used for decades to successfully predict consumer satisfaction (Cadotte, Woodruff, & Jenkins, 1987; Phillips & Baumgartner, 2002) and more recently to consider the factors contributing to victim satisfaction with police responses (Reisig & Chandek, 2001; Robinson & Stroshine, 2005). None of the existing research however, considers victims’ attitudes and subsequent expectations concerning forensic science, and therefore what evidence victims believe police ought to be recovering and analysing in relation to the offence reported. Anecdotal evidence
suggests that scenes of crime officers (SOCO) regularly encounter unrealistic
expectations from crime victims, but no empirical research exists to determine what
impact this may have on the level of reported victim satisfaction. The FEEBS might be
a useful measure of attitudes about forensic science which could be used to
investigate victim satisfaction in an expectancy disconfirmation paradigm in future
research.

9.6 CONCLUSIONS

The studies presented within this thesis have demonstrated that jury-eligible
members of the public possess variable levels of forensic awareness about common
forensic science techniques, and the majority of people report that television and
movies are the main source of their knowledge about forensic science and the criminal
justice system. Contrary to some previous literature, this thesis also concluded that in
the absence of case context participants were able to distinguish between forensic
evidence items which were varied in probative value based on the mobility and
relevance of the evidence. However contextual effects were found when participants
evaluated the same evidence types framed in a trial narrative, and these effects were
particularly problematic when the evidence was of a weak or ambiguous probative
standard.

In order to further explain the significant disagreement amongst participants
regarding the probative value of weak forensic evidence, the Forensic Evidence
Evaluation Bias Scale (FEEBS) was developed. This scale measured attitudes about
forensic science, which are hypothesised in the CSI Effect literature to be the cause of
problematic juror decisions in recent years. The FEEBS was consistently shown
throughout this thesis to have two distinct clusters of attitudes, which were conceptualised as two subscales (representing pro-prosecution, and pro-defence attitudes about forensic science). Although scores on the FEEBS were not found to be reliably linked to viewing habits (an assumption made in the CSI Effect literature), they were conceptually similar to the two main manifestations of the CSI Effect hypothesised in the previous literature.

The FEEBS subscales were used to predict mock juror verdict preferences across a range of crime types, including murder, robbery, and sexual assault using a range of participant samples. The FEEBpp was found to predict the perceived strength of weak forensic DNA evidence, which was confirmed to be a mediator variable for probability of guilt estimates. The FEEBpd was found to have a similar mediation relationship with probability of guilt for cases in which the prosecution did not present any forensic evidence, through the perceived strength of the circumstantial evidence. The predictive validity of the FEEBS was superior to the JBS, GBJW, and other commonly used voir dire questions, which is supported by research from the attitude-behaviour literature which suggests that specific attitudes are better predictors of behaviour than more general attitude measures. Finally, the use of negative evidence witnesses was investigated, and found to be potentially useful for managing unrealistic expectations that jurors might have for forensic evidence in the prosecution’s case.

This thesis has contributed to the existing literature by investigating the evaluation of probative value by using a systematic manipulation of mobility and relevance of evidence, rather than focusing only on the evaluation of probabilistic testimony. It has also improved our understanding of pre-trial attitudes about forensic
science, and the impact that these beliefs might have on the evaluation of weak forensic evidence at trial. In doing so, it has also improved upon the methodology commonly employed within the CSI Effect literature by moving away from focusing on viewing habits as the independent variable and instead highlighting the importance of internalisation of attitudes commonly portrayed in forensic fiction. Finally, this thesis has challenged the use of some current pre-trial questioning techniques, and argues that empirically-derived attitude measures (such as the FEEBS) are superior predictors of juror behaviour.

Regardless of whether one believes that the jury system is an ideal way to decide the fate of defendants at trial, the fact remains that juries are likely to continue to be the hallmark of adversarial legal systems in a number of countries. Given that this is the case, legal and psychological researchers should continue to develop research programmes which aim to improve the ability of jurors to reach fair and legally correct decisions. This becomes an increasingly important area of research, as scientific technology continues to produce more complex evidential analyses alongside a pop culture fascination with the police use of forensic science. It is hoped that research, such as that presented within this thesis, will continue to improve our understanding of juror decision making about complex evidence and ultimately contribute to the goal of minimising miscarriages of justice and maximising legally and logically sound decision making in the criminal courts.
REFERENCES


Bornstein, B. (1999). The ecological validity of jury simulations: Is the jury still out? 
*Law and Human Behavior, 23,* 75-91. doi: 10.1023/A:1022326807441


juror: The psychology of juror decision making (pp. 65-83). Cambridge: Cambridge University Press.


Juries Act 1974 S.12 (1)(b)


R v Adams [1998] 2 Cr App R 467

R v George (Barry) [2007] EWCA Crim 2722; (2007) 151 S.J.L.B. 1498 (CA (Crim Div))

R v Hoey [2007] NICC49


**APPENDIX A**

**General forensic awareness questionnaire items**

Positively worded items (agreement indicates a correct response):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA can be used to positively identify someone</td>
<td>Williams &amp; Johnson (2008)</td>
</tr>
<tr>
<td>DNA can be recovered from discarded chewing gum at a crime scene</td>
<td>Forensic Science Service (2006b)</td>
</tr>
<tr>
<td>DNA can be recovered from blood stains at a crime scene</td>
<td>Forensic Science Service (2006b)</td>
</tr>
<tr>
<td>Fingerprints can be used to positively identify someone</td>
<td>White (2004)</td>
</tr>
<tr>
<td>Fingerprints can be destroyed by an offender in order to avoid detection</td>
<td>White (2004)</td>
</tr>
<tr>
<td>Gun shot residue can be used to determine the manufacturer of a firearm cartridge</td>
<td>Forensic Science Service (2007c)</td>
</tr>
<tr>
<td>Detectable amounts of gun shot residue can be transferred to a suspect by contact with armed police officers</td>
<td>Forensic Science Service (2007c)</td>
</tr>
<tr>
<td>Examination of a bullet from a crime scene can determine the type of weapon used in the shooting (calibre, model of the firearm)</td>
<td>Forensic Science Service (2006c)</td>
</tr>
<tr>
<td>A bullet can be successfully traced to the individual weapon from which it was fired</td>
<td>Forensic Science Service (2006c)</td>
</tr>
<tr>
<td>Footwear impressions can be used to successfully link crime scenes with the same perpetrator</td>
<td>Forensic Science Service (2006a)</td>
</tr>
<tr>
<td>Footwear impressions can be successfully recovered from clothing and bodies</td>
<td>Forensic Science Service (2006a)</td>
</tr>
<tr>
<td>Tiny fragments of glass recovered on a suspect’s clothes can be successfully matched to broken glass at a crime scene</td>
<td>Forensic Science Service (2007a)</td>
</tr>
</tbody>
</table>
Negatively worded items (disagreement indicates an accurate response):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A DNA profile can only be recovered from large amounts of DNA (e.g. visible to the naked eye)</td>
<td>Williams &amp; Johnson (2008)</td>
</tr>
<tr>
<td>DNA cannot be recovered from a surface that someone has only touched</td>
<td>Forensic Science Service (2006b)</td>
</tr>
<tr>
<td>Regardless of the circumstances, the best possible evidence to find at a crime scene is DNA</td>
<td>Williams &amp; Johnson (2008)</td>
</tr>
<tr>
<td>Fingerprints can be successfully recovered from any type of surface</td>
<td>White (2004)</td>
</tr>
<tr>
<td>Only complete fingerprints are useful for identifying a person</td>
<td>White (2004)</td>
</tr>
<tr>
<td>Gun shot residue can be successfully linked to a particular fired bullet</td>
<td>Forensic Science Service (2007c)</td>
</tr>
<tr>
<td>Gun shot residue typically remains on the hands of the person who fired a gun, even if they have washed their hands to avoid detection</td>
<td>Forensic Science Service (2007c)</td>
</tr>
<tr>
<td>A bullet is only useful for determining what type of gun was fired if it is in good condition (e.g. not damaged)</td>
<td>Forensic Science Service (2006c)</td>
</tr>
<tr>
<td>It is impossible to tell whether a particular person is the regular wearer of a pair of shoes</td>
<td>Forensic Science Service (2006a)</td>
</tr>
<tr>
<td>Regardless of the quality of a footwear impression at a crime scene, a conclusive match with a particular shoe is very unlikely</td>
<td>Forensic Science Service (2006a)</td>
</tr>
<tr>
<td>If a suspect breaks a window at a crime scene, glass fragments will remain on their clothing until the clothing is washed</td>
<td>Forensic Science Service (2007a)</td>
</tr>
</tbody>
</table>
Evidence evaluation items

Low mobility/high relevance items (strongest evidence category)

- DNA from a bloodstain at a crime scene matches a suspect’s DNA
- A suspect’s fingerprint is found on a table in a burglary scene
- Shoeprints are left in blood at the scene of a murder

High mobility/high relevance items (somewhat strong evidence category)

- A DNA sample from saliva on a beer bottle found inside a burglary scene
- DNA from sweat recovered from a mask left at a burglary scene
- Fingerprints are recovered from a discarded crisp pack in a stolen vehicle

Low mobility/low relevance items (somewhat weak evidence category)

- A husband is suspected of killing his wife, and his fingerprints are found on the murder weapon, which is his hunting knife
- Fingerprints are found on the outside of the door of a stolen vehicle
- Shoeprints are found outside a burglary scene

High mobility/low relevance items (weakest evidence category)

- DNA recovered from discarded chewing gum found on the pavement outside a burglary scene
- DNA recovered from a discarded cigarette end found outside a stolen car
- Fingerprints are recovered from a tool discarded near the scene of a burglary (the tool may have been used during the burglary)
APPENDIX C

Case context conditions – witness statements

Text in **bold** appeared in the low mobility/high relevance DNA evidence condition (strongest evidence)

Text in *italics* appeared in the high mobility/low relevance DNA evidence condition (weakest evidence)

Text in _underline_ appeared in the high mobility/high relevance DNA evidence condition (moderate evidence)

Case Summary

At 22:46 on the evening of November 22, 2008 a 999 operator received a call from a man who reported hearing screams from his neighbour’s house and witnessed a man run from the house and flee the scene in a light-coloured four door car. The witness entered the house and found the victim, Miss Jones, lying unconscious in the front hallway covered in blood. Police and ambulance responded to the call and Miss Jones was pronounced dead at the scene. An autopsy confirmed that she died as a result of multiple stab wounds inflicted by a large hunting-style knife.

At 23:25 the suspect (Mr. Evans) was stopped by police for speeding and was suspected of driving under the influence of alcohol. Police brought Mr. Evans in for questioning in the murder of Miss Jones, due to the proximity to the crime scene and the fact that Mr. Evans’ vehicle was consistent with the description of the car, seen by
the eyewitness, leaving Miss Jones’ house. A subsequent search of Mr. Evans’ vehicle recovered some stolen property which was traced to houses in Miss Jones’ neighbourhood which had recently been burglarised.

*Instructions to the jury*

As a member of the jury it is your job to read the evidence presented by each witness and consider whether the defendant, Mr. Evans, is guilty of murder. You should vote guilty only if the evidence has proven the guilt of the defendant beyond a reasonable doubt.

*Scenes of Crime Examination*

Scenes of Crime officers who examined the crime scene recovered a number of blood samples from the immediate area surrounding the victim’s body. The blood samples were analysed and all matched the victim’s DNA except for one sample. This sample was found to be a match to the suspect, Mr. Evans.

Scenes of Crime Officers who examined the crime scene were not able to recover any fingerprint or DNA evidence from inside the crime scene or from the body of the victim. A wider search of the property recovered a discarded cigarette end from the front garden. The forensic laboratory was able to recover a DNA profile from the cigarette end, and this profile was found to be a match with the suspect, Mr. Evans.

Scenes of Crime Officers who examined the crime scene were not able to recover any fingerprint evidence from inside the crime scene or forensic evidence on the body of the victim. A search of the property recovered a cigarette end crushed into the carpet inside the front hall of the house. The forensic laboratory was able to recover a DNA
profile from the cigarette end, and this profile was found to be a match with the suspect, Mr. Evans.

Despite a thorough search of the crime scene and surrounding areas, as well as the suspect’s vehicle, the murder weapon was not recovered.

Eyewitness Testimony – Victim’s Neighbour (Mr. Smith)

On the night of November 22 at about 10:40pm I was outside of my house getting something out of my car when I heard screaming coming from Miss Jones’ house next door. I looked up and saw a man run out the front door of the house to a car that was parked on the street in front of the house. He got into the car and drove away quickly. The car was a light colour, like white or cream or grey, and was definitely a four-door model.

It was really dark, and everything happened so quickly, so I didn’t get a good look at the man. He was wearing a dark coloured jacket, and dark trousers. He was also wearing a hat, so I could not see his hair. He appeared to be of average physical build.

Once he had driven away I ran across the front garden of Miss Jones’ house and entered the house through the front door, which had been left wide open when the man left. As soon as I entered the front door I could see a body lying in the front hall area. I ran over to the body and realised that it was Miss Jones. There was a lot of blood on the floor and on her clothes. I immediately called 999 from my mobile. The ambulance and police arrived shortly after.
**Testimony of Arresting Officer**

On the evening of November 22 I was on traffic patrol about ten miles from the location where Miss Jones was murdered. At approximately 23:25 I stopped a vehicle for speeding. The driver of this vehicle was the accused, Mr Evans. When I approached his vehicle and engaged in a conversation with the driver he seemed agitated and initially I suspected that he may have been under the influence of alcohol.

A basic search of the vehicle recovered a number of items, such as mobile phones and iPods, which we later were able to determine were stolen items from a number of local houses which had been burglarised in the previous months.

I brought Mr Evans in to the police station for questioning, as I was aware of the relatively close proximity to the murder scene. Mr Evans claimed that he was driving home from a local Pub, where he had been, alone, since earlier in the evening. In regards to the stolen items found in his car, Mr Evans claimed that he had purchased the items from an acquaintance and was not aware of the fact that they were stolen.

I investigated Mr Evans’ alibi and the bartender at the pub was able to confirm that Mr Evans had indeed been in the pub on the evening of November 22, but he was not able to say for certain what time Mr Evans had left that night.

**Testimony of Forensic Pathologist**
I performed the post mortem examination on the victim in this case, Miss Jones. I determined the cause of death to be multiple stab wounds to the victim’s chest made by a large hunting-style knife. I also noted other injuries on the victim’s body such as bruises and other superficial wounds which, in my professional opinion, are consistent with a physical struggle prior to her death.

**Testimony of Forensic Laboratory Technician**

Analysis of the forensic DNA evidence recovered from the crime scene in this case was carried out at an accredited laboratory, and I conducted the DNA profile analyses myself. I have been a laboratory analyst for ten years and have carried out hundreds of similar analyses throughout my career.

The material submitted to our laboratory which was recovered from the scene of this crime was compared with DNA extracted from the suspect, Mr. Evans. The results of the DNA comparison concluded that the sample from the crime scene and the sample from Mr Evans were indeed a match. The chance that the Mr Evans’ DNA would match the crime scene sample by coincidence if he were not the source of the DNA is approximately 1 in 1,000,000.

**Summary of the Prosecution’s Arguments**

The victim’s neighbour (Mr Smith) who called the police reported seeing a man run from Miss Jones’ house and leave quickly in a light coloured four-door car, which is
consistent with the type of car that Mr. Evans was stopped in later that evening. The suspect told police that at the time of the murder (approx 10:45pm) he was at a local pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening, but he could not confirm what time he left the pub.

The stolen goods found in Mr. Evans’ car link him to a number of burglaries in the victim’s neighbourhood, and surrounding area. We believe that Mr. Evans had broken into Miss Jones’ home on the evening of Nov 22 with the intention of stealing property and was confronted by Miss Jones’. This led to a struggle during which the defendant stabbed the victim, resulting in her death.

The DNA profile obtained from the blood sample taken from the scene, near the victim’s body, matches Mr. Evans’ DNA profile. This demonstrates that Mr. Evans was involved in a struggle with the victim, and therefore proves that he is guilty of murdering Miss Jones.

The DNA profile obtained from the discarded cigarette end found in the front garden of the crime scene location matches Mr. Evans’ DNA profile. This demonstrates that Mr. Evans was present at the crime scene and specifically where the eyewitness reported seeing the murderer run to his car. This therefore further proves that he is guilty of murdering Miss Jones.

The DNA profile obtained from the discarded cigarette end found in the front hall of the crime scene location matches Mr. Evans’ DNA profile. This demonstrates that Mr.
Evans was present inside the crime scene. This therefore further proves that he is guilty of murdering Miss Jones.

**Summary of the Defence Arguments**

The victim’s neighbour (Mr Smith) testified that he did not get a clear look at the man who ran from the scene, and could therefore not identify Mr. Evans as the man he saw. Although Mr. Evans was in fact driving a light coloured four-door car at the time of his arrest, there was nothing irregular or specific about his car which could confirm that it was the vehicle the neighbour saw leaving the scene.

The stolen goods found in Mr Evans’ car were not stolen by him, but he purchased these items from an acquaintance and was not aware that they were stolen.

Mr. Evans also told police that at the time of the murder (approx 10:40pm) he was at a local pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening.

The only clear evidence against Mr Evans is the DNA profile obtained from the crime scene. It is entirely possible that another person could be the source of the blood sample, even if the profile ‘matches’ Mr. Evans.

*Mr. Evans does smoke cigarettes and so there is nothing strange about finding a discarded cigarette end with his DNA profile available on it. The fact that the cigarette was found in front garden of the crime scene location does not provide proof that Mr. Evans is guilty of the murder of Miss Jones. The cigarette may have been discarded*
previous to the evening of the murder, and could have easily been ‘tracked’ into the
garden on someone’s shoe. The presence of this evidence can conceivably be explained
without requiring Mr. Evans to be guilty of the murder of Miss Jones.

Mr. Evans does smoke cigarettes and so there is nothing strange about finding a
discarded cigarette end with his DNA profile available on it. The fact that the cigarette
was found in front hall of the crime scene location does not provide proof that Mr.
Evans is guilty of the murder of Miss Jones. The cigarette may have been discarded
previous to the evening of the murder, and could have easily been ‘tracked’ into the
scene on someone’s shoe. The presence of this evidence can conceivably be explained
without requiring Mr. Evans to be guilty of the murder of Miss Jones.

Mr. Evans has been wrongfully accused of the murder of Miss Jones as a result of
being ‘in the wrong place, at the wrong time’ and therefore the real murderer is still at
large.
APPENDIX D

**Items included in initial version of the scale**

1. Science is the most reliable way to identify the perpetrators of crimes
2. There are many fewer wrongful convictions in recent years because of forensic science evidence
3. Forensic evidence always eventually identifies the guilty person
4. It would be very difficult to commit a crime without being caught with forensic science
5. Scientific evidence does not lie
6. If forensic evidence suggests a defendant is guilty, this should be enough to convict them even if other evidence (e.g. eyewitness, alibi) suggests otherwise
7. Every criminal leaves some physical evidence behind at every crime scene
8. Every crime can be solved with forensic science
9. The real strength of scientific evidence is that it is not affected by human error
10. One big advantage to scientific evidence – as opposed to other types of evidence – is that it always provides a conclusive answer
11. Forensic evidence is unreliable because errors can occur in the laboratory when the evidence is being analysed
12. If there is no forensic evidence presented in a particular criminal case, then the jury should not convict
13. Police should not charge someone with a serious crime unless forensic evidence is available to prove their guilt
14. Forensic scientists who testify in court are often only presenting evidence to support one side (e.g. prosecution or defence) because they have been hired by that side.

15. It is possible that scientific evidence can identify an innocent person as the perpetrator of a crime.

16. If no forensic evidence is recovered from a crime scene, it means the investigators did not look hard enough.

17. It is easy for police to tamper with physical evidence.

18. Some forensic science techniques are not based on ‘hard science’ and therefore should not be used in the courtroom.

19. Witness testimony is not enough on its own to convict a defendant of a serious crime.

20. DNA evidence should always be presented in court as evidence in sexual assault cases.

21. If no forensic evidence is recovered from a crime scene, the defendant is probably innocent of the crime.

22. Crime-related television programmes provide a realistic look at what happens during criminal investigations.

23. DNA is always the best type of evidence for proving the guilt of a defendant.

24. I look forward to serving as a juror, as it would be interesting and exciting.

25. If police have forensic evidence which suggests a suspect is guilty, then that suspect will usually confess.

26. Most people brought to trial are guilty of the crime with which they are being charged (JBS/PC).
27. Generally, the police make an arrest only when they are sure about who committed the crime (JBS/PC)

28. If the majority of evidence – but not all of it – suggests that the defendant committed the crime, then the jury should vote not guilty (JBS/RD)

29. Too many obviously guilty people escape punishment because of legal technicalities (LAQ/A)

30. No one should be convicted of a crime on the basis of circumstantial evidence, no matter how strong such evidence is (LAQ/AA)

31. It is better for society that several guilty people be freed than one innocent person wrongfully imprisoned (LAQ/E)
APPENDIX E

Robbery scenario materials

**Bold** text refers to content only included in the DNA evidence condition

*Italic* text refers to content only included in the no-DNA evidence condition

Summary

At 10:46pm on the evening of November 22, 2008 an emergency services operator received a call from a man who reported witnessing a robbery on Market Street, outside a pub called the Black Horse. The witness (Mr. Smith) was walking east on Market street and then turned left on High Street and moments later he heard a woman scream. He quickly ran back onto Market Street and saw a man struggling with a woman (Miss Jones), who was yelling for help. As Mr. Smith approached them, the perpetrator ran off with Miss Jones’ handbag and jumped into a car parked further up the street and drove quickly from the scene.

At 11:25pm the suspect (Mr. Evans) was stopped by police for speeding and was suspected of driving under the influence of alcohol. Police brought Mr. Evans in for questioning in the robbery of Miss Jones, due to the proximity to the crime scene and the fact that Mr. Evans’ vehicle was consistent with the description of the car, seen by the eyewitness, leaving the area where Miss Jones was attacked. A subsequent search of Mr. Evans’ vehicle recovered a number of mobile phones and iPods, as well as £500 in cash.

Despite a thorough search of the suspect’s vehicle, Miss Jones’ handbag and personal belongings were not recovered.
Instructions to the jury

As a member of the jury it is your job to read the evidence presented by each witness and consider whether the defendant, Mr. Evans, is guilty of robbery. You should vote guilty only if the evidence has proven the guilt of the defendant beyond a reasonable doubt.

Eyewitness Testimony (Mr. Smith)

“On the night of November 22 at about 10:40pm I was walking home from a bar on Market Street where I had met with some friends. I turned left on High Street and then I heard a woman scream. It sounded like the screaming was coming from Market Street, where I had just been, so I turned around and ran back around the corner. I saw a man and woman struggling on the opposite side of the road, and the woman was yelling for help. I ran toward them, and the man saw me and managed to grab the woman’s handbag before running off. He got into a parked car and drove away very quickly. The car was a light colour, like white or cream or grey, and was definitely a four-door model.

It was really dark, and everything happened so quickly, so I didn’t get a good look at the man. He was wearing a dark coloured jacket, and dark trousers. He was also wearing a woolly hat and had a scarf around his neck which was covering his face. He appeared to be of average physical build and height.

Once he had driven away I made sure Miss Jones was not hurt, and then I called the police. The police arrived shortly after and I told them what I had seen.”
Testimony of responding Officer

“At approximately 22:55 on November 22 I responded to an emergency call from a witness to a robbery. This call was made by Mr. Smith. When I arrived at the scene on Market Street I questioned both Mr. Smith and Miss Jones about the robbery. Miss Jones told me that she had been walking along Market Street on her way to a bar to meet some friends, when she was approached by a man wearing a knit hat and a scarf partially covering his face. He told her to give him her handbag, and when she refused he grabbed her and tried to forcibly take her handbag. Miss Jones’ told me that during the struggle she managed to hit the perpetrator and scratch him on the side of the face.

She yelled for help, but the man managed to take her handbag before running off and driving away in a car parked further up the street. Both Mr. Smith and Miss Jones described the car as a four-door model, and a light colour, although it was very dark and difficult to tell the model of the car from their position.

I conducted a search of the scene and surrounding area and I was not able to find any useful evidence in the immediate area. Later that evening however, while on foot patrol, I found a discarded knit hat one block from the scene of the robbery, and also two handbags and purses lying nearby. This hat was identified by the victim and witness as being consistent with the type of hat worn by the perpetrator. Miss Jones’ personal belongings were not among the items found near the hat”.

I conducted a search of the scene and surrounding area and I was not able to find any useful evidence in the immediate area.”
Testimony of arresting Officer

“On the evening of November 22 I was on traffic patrol about five miles from the location where Miss Jones was robbed. At approximately 23:25 I stopped a vehicle for speeding. The driver of this vehicle was the accused, Mr. Evans. When I approached his vehicle and engaged in a conversation with the driver he seemed agitated and initially I suspected that he may have been under the influence of alcohol.

A basic search of the vehicle recovered a number of items, such as mobile phones and iPods, as well as £500 in cash in the glove compartment of the vehicle. I also found a knit hat in Mr Evans’ car, which was later identified by the victim and witness as being consistent with the hat worn by the person who robbed Miss Jones.

I brought Mr Evans in to the police station for questioning, as I was aware of the relatively close proximity to the robbery scene. In addition Mr. Evans had a number of scratches on his face as well as a swollen lower lip, which he told me he got in a minor fight earlier that evening.

Mr Evans claimed that he was driving home from a local Pub, where he had been, alone, since earlier in the evening. In regards to the items found in his car, Mr Evans claimed that he had purchased the items from an acquaintance and that the cash was payment that he had received for some casual construction work that he had recently done for a friend.

I investigated Mr Evans’ alibi and the bartender at the pub was able to confirm that Mr Evans had indeed been in the pub on the evening of November 22, but he was not able to say for certain what time Mr Evans had left that night.”
**Testimony of Forensic Laboratory Analyst**

“Analysis of the forensic DNA evidence recovered from the wool hat found discarded near the scene of the robbery was carried out at an accredited laboratory, and I conducted the DNA profile analyses myself. I have been a laboratory analyst for ten years and have carried out hundreds of similar analyses throughout my career.

The material submitted to our laboratory which was recovered from the wool hat was compared with DNA extracted from the suspect, Mr. Evans. The results of the DNA comparison concluded that the sample from the crime scene and the sample from Mr Evans were indeed a match. The chance that the Mr Evans’ DNA would match the crime scene sample by coincidence if he were not the source of the DNA is approximately 1 in 1,000,000.”

*Summary of the prosecution’s arguments:*

The man who called the police reported seeing a man struggling with Miss Jones’ who ran to a parked car and drove away quickly in a light coloured four-door car, which is consistent with the type of car that Mr. Evans was stopped in later that evening.

Mr. Evans also told police that at the time of the robbery (approx 22:40) he was at a pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening, but he could not confirm what time he left the pub.

The mobile phones and iPods found in Mr. Evans’ car, along with a large quantity of cash, are what we might expect to find in the possession of someone who commits
thefts and robberies. Indeed there have been a number of similar, recent incidents in this neighbourhood. **The wool hat found near the robbery scene contained Mr. Evans’ DNA and was also consistent with the type of hat the perpetrator was wearing when the robbery took place.** In addition to this evidence, the injuries which Mr. Evans had on his face and lip were consistent with the struggle described by Miss Jones when she gave the police her statement. **Finally, the hat found in Mr. Evans’ car matches the description of the perpetrators hat given in statements by both the victim and the witness.**

The prosecution alleges that Mr. Evans is the man who robbed Miss Jones and after fleeing the scene he threw his wool hat and some other stolen handbags and items from his car one block from where the robbery occurred, possibly to avoid being apprehended with these items in his possession.

**Summary of the defence arguments:**

The eye witness (Mr. Smith) testified that he did not get a clear look at the man who attacked Miss Jones and ran from the scene, and could therefore not identify Mr. Evans as the man he saw.

Although Mr. Evans was in fact driving a light coloured four-door car at the time of his arrest, there was nothing irregular or specific about his car which could confirm that it was the vehicle the witness saw leaving the scene.

Mr. Evans told police that the items found in his car were not stolen by him, and that he purchased these items from an acquaintance. **When questioned about the wool hat found near the robbery scene with his DNA on it, Mr. Evans explained that he**
had recently donated some clothing to a charity shop, and this wool hat was among the donated clothing items. *When questioned about the scratches on his face and his swollen lip, Mr. Evans explained that he had been in a minor fight earlier in the evening. Although Miss Jones claimed to have injured her attacker in a similar way, there is no proof that she inflicted Mr. Evans’ injuries.*

Mr. Evans also told police that at the time of the murder (approx 22:40) he was at a pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening.

The defence argues that Mr. Evans has been wrongfully accused of the robbery of Miss Jones, and that there is insufficient physical evidence to indicate otherwise.

**Sexual assault scenario materials**

**Bold** text refers to content only included in the DNA evidence condition

**Italic** text refers to content only included in the no-DNA evidence condition

**Summary**

At 22:46 on the evening of November 22, 2008 an emergency services operator received a call from a man who reported a sexual assault on Market Street, in an alleyway near a pub called the Black Horse. The witness (Mr. Smith) was walking east on Market street after leaving a nearby bar. He turned left on High Street and moments later he heard a woman scream. He quickly ran back onto Market Street.
and saw a man run past him and jump into a car parked further up the street and drove quickly from the scene. Mr. Smith carried on walking along Market Street and found the victim, Miss Jones, in an alleyway yelling for help.

At 23:25 the suspect (Mr. Evans) was stopped by police for speeding and was suspected of driving under the influence of alcohol. Police brought Mr. Evans in for questioning in the sexual assault of Miss Jones, due to the proximity to the crime scene and the fact that Mr. Evans' vehicle was consistent with the description of the car seen leaving the area where Miss Jones was attacked.

*Instructions to the jury*

As a member of the jury it is your job to read the evidence presented by each witness and consider whether the defendant, Mr. Evans, is guilty of sexual assault. You should vote guilty only if the evidence has proven the guilt of the defendant beyond a reasonable doubt.

*Eyewitness Testimony (Mr. Smith)*

"On the night of November 22 at about 10:40pm I was walking home from a bar on Market Street where I had met with some friends. I turned left on High Street and a few moments later I heard a woman scream. It sounded like the screaming was coming from Market Street, where I had just been, so I turned around and ran back around the corner. Just as I got back onto Market Street a man ran by me, got into a parked car and drove away very quickly. The car was a light colour, like white or cream or grey, and was definitely a four-door model."
It was really dark, and everything happened so quickly, so I didn’t get a good look at the man. He was wearing a dark coloured jacket, and dark trousers. He was also wearing a woolly hat and had a scarf around his neck which was covering his face. He appeared to be of average physical build and height.

I carried on walking along Market Street and found Miss Jones standing in an alleyway crying and yelling for help. I called the police and they arrived shortly after to take our statements.”

Testimony of responding Officer

“At approximately 22:55 on November 22 I responded to an emergency call regarding a sexual assault. This call was made by Mr. Smith. When I arrived at the scene on Market Street I questioned both Mr. Smith and Miss Jones about the attack. Miss Jones told me that she had been walking along Market Street on her way to a bar to meet some friends, when she was approached by a man wearing a knit hat and a scarf partially covering his face. He grabbed her and dragged her into an alleyway where he sexually assaulted her.

She yelled for help and tried to fight him off, and he eventually ran off and drove away in a car parked further up the street. Mr. Smith described the car as a four-door model, and a light colour, although it was very dark and difficult to tell the model of the car from his position.

I conducted a search of the scene and surrounding area and I was not able to find any useful evidence in the immediate area. I took Miss Jones to the hospital for examination and further questioning. In a further statement Miss Jones told me that
she had managed to scratch her attacker on the face during the assault. Later, while I was on foot patrol in the area I found a wool hat similar to the hat worn by the perpetrator, discarded one block from the scene of the assault. This hat was identified by the victim and witness as being consistent with the type of hat worn by the perpetrator.”

Testimony of arresting Officer

“On the evening of November 22 I was on traffic patrol about five miles from the location where Miss Jones was assaulted. At approximately 23:25 I stopped a vehicle for speeding. The driver of this vehicle was the accused, Mr. Evans. When I approached his vehicle and engaged in a conversation with the driver he seemed agitated and initially I suspected that he may have been under the influence of alcohol.

I brought Mr Evans in to the police station for questioning, as I was aware of the relatively close proximity to the assault scene. In addition Mr. Evans had a number of scratches on his face as well as a swollen lower lip, which he told me he got in a minor fight earlier that evening. During a basic search of the suspect’s car, a knit hat was recovered which was later identified by Mr. Smith and Miss Jones as being consistent with the hat worn by Miss Jones’ attacker.

Mr Evans claimed that he was driving home from a local Pub, where he had been, alone, since earlier in the evening. I investigated Mr Evans’ alibi and the bartender at the pub was able to confirm that Mr Evans had indeed been in the pub on the evening
of November 22, but he was not able to say for certain what time Mr Evans had left that night.”

Testimony of Forensic Laboratory Analyst

“Analysis of the forensic DNA evidence recovered from the wool hat found near the scene of the assault was carried out at an accredited laboratory, and I conducted the DNA profile analyses myself. I have been a laboratory analyst for ten years and have carried out hundreds of similar analyses throughout my career.

The material submitted to our laboratory which was recovered from the wool hat was compared with DNA extracted from the suspect, Mr. Evans. The results of the DNA comparison concluded that the sample from the crime scene and the sample from Mr Evans were indeed a match. The chance that the Mr Evans’ DNA would match the crime scene sample by coincidence if he were not the source of the DNA is approximately 1 in 1,000,000.”

Summary of the prosecution’s arguments:

The witness who called the police reported seeing a man run past him and drive away quickly in a light coloured four-door car, which is consistent with the type of car that Mr. Evans was stopped in later that evening.

Mr. Evans also told police that at the time of the assault (approx 22:40) he was at a pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the
assault told police that the suspect was in the bar that evening, but he could not confirm what time he left the pub.

The wool hat found near the scene of the attack contained Mr. Evans’ DNA and was also consistent with the type of hat the perpetrator was wearing when Miss Jones was sexually assaulted. In addition, the injuries Mr. Evans had on his face when he was brought in for questioning were consistent with the struggle described by Miss Jones when she gave the police her statement. In addition, when police stopped Mr. Evans’ he had in his possession a woolly hat just like the one worn by Miss Jones’ attacker.

The prosecution alleges that Mr. Evans is the man who sexually assaulted Miss Jones and after fleeing the scene he threw his wool hat from the car near where the assault occurred, possibly to avoid being apprehended with this item in his possession.

Summary of the defence arguments:

The witness (Mr. Smith) testified that he did not get a clear look at the man who attacked Miss Jones and ran from the scene, and could therefore not identify Mr. Evans as the man he saw.

Although Mr. Evans was in fact driving a light coloured four-door car at the time of his arrest, there was nothing irregular or specific about his car which could confirm that it was the vehicle the witness saw leaving the scene.
When questioned about the wool hat found near the scene of the assault, Mr. Evans explained that he had recently donated some clothing to a charity shop, and this wool hat was among the donated clothing items.

When asked about the scratches on his face and his swollen lip, Mr. Evans explained that he had been in a minor fight earlier in the evening. Although Miss Jones claimed to have injured her attacker in a similar way, there is no proof that she inflicted Mr. Evans’ injuries.

Mr. Evans also told police that at the time of the assault (approx 22:40) he was at a pub by himself and had left the pub approx 5 minutes before he was stopped by police. The bartender at the pub where Mr. Evans claimed to be at the time of the assault told police that the suspect was in the bar that evening.

The defence argues that Mr. Evans has been wrongfully accused of assaulting Miss Jones, and that there is insufficient physical evidence to indicate otherwise.
APPENDIX F

Trial materials for Study 6

Murder scenario – No forensic evidence

Summary

At 22:46 on the evening of November 22, 2008 a 911 operator received a call from a man who reported hearing screams from his neighbour’s house and witnessed a man run from the house and flee the scene in a light-coloured four door car. The witness entered the house and found the victim, Miss Jones, lying unconscious in the front hallway covered in blood. Police and ambulance responded to the call and Miss Jones was pronounced dead at the scene. An autopsy confirmed that she died as a result of multiple stab wounds inflicted by a large hunting-style knife.

At 23:25 the suspect (Mr. Evans) was stopped by police for speeding and was suspected of driving under the influence of alcohol. Police brought Mr. Evans in for questioning in the murder of Miss Jones, due to the proximity to the crime scene and the fact that Mr. Evans’ vehicle is consistent with the description of the car, seen by the eyewitness, leaving Miss Jones’ house. In addition, Mr. Evans had visible scratches on his face, which suggested that he had been involved in a struggle or fight recently. A subsequent search of Mr. Evans’ vehicle recovered some stolen property which was traced to houses in Miss Jones’ neighbourhood which had recently been burglarized.
Scenes of Crime Officers who examined the crime scene were not able to recover any fingerprint or DNA evidence from inside the crime scene or from the body of the victim. Despite a thorough search of the crime scene and surrounding areas, as well as the suspect’s vehicle, the murder weapon was not recovered.

Eyewitness Testimony – Neighbour (Mr. Smith)

On the night of November 22 at about 10:40pm I was outside of my house getting something out of my car when I heard screaming coming from Miss Jones’ house next door. I looked up and saw a man run out the front door of the house to a car that was parked on the street in front of the house. He got into the car and drove away quickly. The car was a light colour, like white or cream or grey, and was definitely a four-door model.

It was really dark, and everything happened so quickly, so I didn’t get a good look at the man. He was wearing a dark coloured jacket, and dark trousers. He was also wearing a hat, so I could not see his hair. He appeared to be of average physical build.

Once he had driven away I ran across the front garden of Miss Jones’ house and entered the house through the front door, which had been left wide open when the man left. As soon as I entered the front door I could see a body lying in the front hall area. I ran over to the body and realised that it was Miss Jones. There was a lot of blood on the floor and on her clothes. I immediately
called 911 from my cell phone. The ambulance and police arrived shortly after.

Testimony of arresting Officer

On the evening of November 22 I was on traffic patrol about ten miles from the location where Miss Jones was murdered. At approximately 23:25 I stopped a vehicle for speeding. The driver of this vehicle was the accused, Mr Evans. When I approached his vehicle and engaged in a conversation with the driver he seemed agitated and initially I suspected that he may have been under the influence of alcohol. In addition to his behaviour, he also had a number of scratches on his face which seemed to me to indicate that he had been in a struggle with someone recently.

A basic search of the vehicle recovered a number of items, such as mobile phones and iPods, which we later were able to determine were stolen items from a number of local houses which had been burglarised in the previous months.

I brought Mr Evans in to the police station for questioning, as I was aware of the relatively close proximity to the murder scene. Mr Evans claimed that he was driving home from a local bar, where he had been, alone, since earlier in the evening. In regards to the stolen items found in his car, Mr Evans claimed that he had purchased the items from an acquaintance and was not aware of the fact that they were stolen. When I asked him about the scratches on his
face, he hesitated and then said that he had been playing with a friend’s cat and it must have scratched him.

I investigated Mr Evans’ alibi and the bartender at the bar was able to confirm that Mr Evans had indeed been in the bar on the evening of November 22, but he was not able to say for certain what time Mr Evans had left that night.

*Testimony of Pathologist*

I performed the post mortem examination on the victim in this case, Miss Jones. I determined the cause of death to be multiple stab wounds to the victim’s chest made by a large hunting-style knife. I also noted other injuries on the victim’s body such as bruises and other superficial wounds which, in my professional opinion, are consistent with a physical struggle prior to her death.

*Summary of the prosecution’s arguments:*

The neighbour who called 911 reported seeing a man run from Miss Jones’ house and leave quickly in a light coloured four-door car, which is consistent with the type of car that Mr. Evans was stopped in later that evening.

Mr. Evans also told police that at the time of the murder (approx 22:40) he was at a bar by himself and had left the bar approx 5 minutes before he was stopped by police. The bartender at the bar where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening, but he could not confirm what time he left.
The stolen goods found in Mr. Evans’ car link him to a number of burglaries in the victim’s neighbourhood. The prosecution alleges that Mr. Evans had broken into Miss Jones’ home on the evening of Nov 22 with the intention of stealing property and was confronted by Miss Jones. This led to a struggle during which the defendant stabbed the victim, resulting in her death. It is during this struggle that we believe Mr. Evan’s sustained the injuries to his face that evening.

**Summary of the defence arguments:**

Miss Jones’ neighbour testified that he did not get a clear look at the man who ran from the scene, and could therefore not identify Mr. Evans as the man he saw.

Although Mr. Evans was in fact driving a light coloured four-door car at the time of his arrest, there was nothing irregular or specific about his car which could confirm that it was the vehicle the neighbour saw leaving the scene.

Mr. Evans told police that the stolen goods found in his car were not stolen by him, and that he purchased these items from an acquaintance and was not aware that they were stolen.

Mr. Evans also told police that at the time of the murder (approx 22:40) he was at a bar by himself and had left the bar approx 5 minutes before he was stopped by police. The bartender at the bar where Mr. Evans claimed to be at the time of the murder told police that the suspect was in the bar that evening. This alibi confirms that Mr. Evans could not have been at Miss Jones’ house at the time the murder was committed.
The defence argues that Mr. Evans has been wrongfully accused of the murder of Miss Jones as a result of being ‘in the wrong place, at the wrong time’. They allege that the real murderer is still at large.

**Murder scenario – No forensic evidence plus negative evidence expert testimony**

**Summary**

At 22:46 on the evening of November 22, 2008 a 911 operator received a call from a man who reported hearing screams from his neighbour’s house and witnessed a man run from the house and flee the scene in a light-coloured four door car. The witness entered the house and found the victim, Miss Jones, lying unconscious in the front hallway covered in blood. Police and ambulance responded to the call and Miss Jones was pronounced dead at the scene. An autopsy confirmed that she died as a result of multiple stab wounds inflicted by a large hunting-style knife.

At 23:25 the suspect (Mr. Evans) was stopped by police for speeding and was suspected of driving under the influence of alcohol. Police brought Mr. Evans in for questioning in the murder of Miss Jones, due to the proximity to the crime scene and the fact that Mr. Evans’ vehicle is consistent with the description of the car, seen by the eyewitness, leaving Miss Jones’ house. In addition, Mr. Evans had visible scratches on his face, which suggested that he had been involved in a struggle or fight recently. A subsequent search of Mr.
Evans’ vehicle recovered some stolen property which was traced to houses in Miss Jones’ neighbourhood which had recently been burglarized.

Scenes of Crime Officers who examined the crime scene were not able to recover any fingerprint or DNA evidence from inside the crime scene or from the body of the victim. Despite a thorough search of the crime scene and surrounding areas, as well as the suspect’s vehicle, the murder weapon was not recovered.

Eyewitness Testimony – Neighbour (Mr. Smith)

On the night of November 22 at about 10:40pm I was outside of my house getting something out of my car when I heard screaming coming from Miss Jones’ house next door. I looked up and saw a man run out the front door of the house to a car that was parked on the street in front of the house. He got into the car and drove away quickly. The car was a light colour, like white or cream or grey, and was definitely a four-door model.

It was really dark, and everything happened so quickly, so I didn’t get a good look at the man. He was wearing a dark coloured jacket, and dark trousers. He was also wearing a hat, so I could not see his hair. He appeared to be of average physical build.

Once he had driven away I ran across the front garden of Miss Jones’ house and entered the house through the front door, which had been left wide open when the man left. As soon as I entered the front door I could see a body lying in the front hall area. I ran over to the body and realised that it was Miss
Jones. There was a lot of blood on the floor and on her clothes. I immediately called 911 from my cell phone. The ambulance and police arrived shortly after.

Testimony of arresting Officer

On the evening of November 22 I was on traffic patrol about ten miles from the location where Miss Jones was murdered. At approximately 23:25 I stopped a vehicle for speeding. The driver of this vehicle was the accused, Mr Evans. When I approached his vehicle and engaged in a conversation with the driver he seemed agitated and initially I suspected that he may have been under the influence of alcohol. In addition to his behaviour, he also had a number of scratches on his face which seemed to me to indicate that he had been in a struggle with someone recently.

A basic search of the vehicle recovered a number of items, such as mobile phones and iPods, which we later were able to determine were stolen items from a number of local houses which had been burglarised in the previous months.

I brought Mr Evans in to the police station for questioning, as I was aware of the relatively close proximity to the murder scene. Mr Evans claimed that he was driving home from a local bar, where he had been, alone, since earlier in the evening. In regards to the stolen items found in his car, Mr Evans claimed that he had purchased the items from an acquaintance and was not aware of the fact that they were stolen. When I asked him about the scratches on his
face, he hesitated and then said that he had been playing with a friend’s cat and it must have scratched him.

I investigated Mr Evans’ alibi and the bartender at the bar was able to confirm that Mr Evans had indeed been in the bar on the evening of November 22, but he was not able to say for certain what time Mr Evans had left that night.

Testimony of Pathologist

I performed the post mortem examination on the victim in this case, Miss Jones. I determined the cause of death to be multiple stab wounds to the victim’s chest made by a large hunting-style knife. I also noted other injuries on the victim’s body such as bruises and other superficial wounds which, in my professional opinion, are consistent with a physical struggle prior to her death.

Testimony of Forensic Expert (negative evidence testimony)

It is not unusual in cases like this for crime scene examiners to be unable to recover useful evidence from the crime scene. Although television shows like CSI suggest that forensic science evidence is always available from violent crime scenes, this is often not the case in real life. If the perpetrator in this case was wearing gloves when they committed the murder of Miss Jones, then this would explain the lack of fingerprints at the crime scene.

The fact that no DNA belonging to the murderer was recovered from this crime scene is also not surprising. There was no biological trace evidence recovered from this scene, and this is often the case when a crime is carried out quickly and other people,
like the neighbour and the police, have been present in the scene before forensic
examiners could search for DNA evidence.

*Summary of the prosecution’s arguments:*

The neighbour who called 911 reported seeing a man run from Miss Jones’ house and
leave quickly in a light coloured four-door car, which is consistent with the type of car
that Mr. Evans was stopped in later that evening.

Mr. Evans also told police that at the time of the murder (approx 22:40) he was at a
bar by himself and had left the bar approx 5 minutes before he was stopped by police.
The bartender at the bar where Mr. Evans claimed to be at the time of the murder told
police that the suspect was in the bar that evening, but he could not confirm what
time he left.

The stolen goods found in Mr. Evans’ car link him to a number of burglaries in the
victim’s neighbourhood. The prosecution alleges that Mr. Evans had broken into Miss
Jones’ home on the evening of Nov 22 with the intention of stealing property and was
confronted by Miss Jones’. This led to a struggle during which the defendant stabbed
the victim, resulting in her death. It is during this struggle that we believe Mr. Evan’s
sustained the injuries to his face that evening.

*Summary of the defence arguments:*

Miss Jones’ neighbour testified that he did not get a clear look at the man who ran
from the scene, and could therefore not identify Mr. Evans as the man he saw.

Although Mr. Evans was in fact driving a light coloured four-door car at the time of his
arrest, there was nothing irregular or specific about his car which could confirm that it
was the vehicle the neighbour saw leaving the scene.

Mr. Evans told police that the stolen goods found in his car were not stolen by him,
and that he purchased these items from an acquaintance and was not aware that they
were stolen.

Mr. Evans also told police that at the time of the murder (approx 22:40) he was at a
bar by himself and had left the bar approx 5 minutes before he was stopped by police.
The bartender at the bar where Mr. Evans claimed to be at the time of the murder told
police that the suspect was in the bar that evening. This alibi confirms that Mr. Evans
could not have been at Miss Jones’ house at the time the murder was committed.

The defence argues that Mr. Evans has been wrongfully accused of the murder of Miss
Jones as a result of being ‘in the wrong place, at the wrong time’. They allege that the
real murderer is still at large.
PRE-TRIAL QUESTIONS

FEEB Scale

1. Police should not charge someone with a serious crime unless forensic evidence is available to prove their guilt
2. If there is no forensic evidence presented in a particular criminal case, then the jury should not convict
3. If no forensic evidence is recovered from a crime scene, the defendant is probably innocent of the crime
4. If no forensic evidence is recovered from a crime scene, it means the investigators did not look hard enough
5. Every criminal leaves some physical evidence behind at every crime scene
6. Forensic evidence always eventually identifies the guilty person
7. Science is the most reliable way to identify the perpetrators of crimes
8. If forensic evidence suggests a defendant is guilty, this should be enough to convict them even if other evidence (e.g. eyewitness, alibi) suggests otherwise
9. The biggest advantage to scientific evidence – as opposed to other types of evidence – is that it always provides a conclusive answer
10. Every crime can be solved with forensic science
Additional voir dire questions:

1. Do you watch CSI (or similar programs such as Law and Order, NCIS, Bones, Etc)? (yes or no)
   a. If yes, how many times per week?

2. Forensic television programmes (such as CSI) provide a realistic view of the Criminal Justice System.
   a. Level of agreement (1-5)

3. Indicate how realistic you think the forensic science techniques are in shows like CSI.
   a. 5-point Likert scale (entirely unrealistic – completely realistic)

4. How much of your knowledge about the Criminal Justice System comes from television programs?
   a. 5-point Likert scale (none of my knowledge – all of my knowledge)

5. In what percentage of criminal cases would you expect forensic evidence to be part of the prosecution’s case?
   a. Expressed as a percentage (0-100)