Instructing Jurors on Detecting Deception:
Is the Jury Still Out?

By

Jessica Woskett, BA, PG Dip. Psych., MCrin

Submitted in partial fulfilment of the requirements for the degree of
Doctor of Psychology (Forensic)

Deakin University
July, 2019
I am the author of the thesis entitled:

**Instructing Jurors on Detecting Deception: Is the Jury Still Out?**

submitted for the degree of Doctor of Psychology (Forensic)

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

*I certify that I am the student named below and that the information provided in the form is correct*.

Full name: Jessica Woskett

Signed: [Signature Redacted by Library]

Date: 12 January 2020
DEAKIN UNIVERSITY

CANDIDATE DECLARATION

I certify the following about the thesis entitled:

**Instructing Jurors on Detecting Deception: Is the Jury Still Out?**

submitted for the degree of **Doctor of Psychology (Forensic)**

a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.

b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.

d. That any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.

e. All research integrity requirements have been complied with.

‘I certify that I am the student named below and that the information provided in the form is correct’

**Full Name:** Jessica Woskett

**Signed:** [Signature Redacted by Library]

**Date:** 28 July 2019
Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my primary supervisor, Emeritus Professor Donald Thomson. I am thankful to you for your sense of humor that always featured in our discussions, your enduring patience, and your ability to support me in a way that not only fostered my intellect but resilience. Above all, I am thankful to you for sharing your resounding knowledge and expertise. Your sage words and teachings will undoubtedly follow me over the course of my career.

I would also like to thank my associate supervisors past and present, Dr Clint Gurtman, Professor Andy Day and Professor Ian Coyle. Dr Clint Gurtman, I am grateful to you for your warm encouragement and motivation throughout this project and above all, your collegial approach. I value our shared interests in all things forensic and eagerly look forward to many a future encounters. I extend my deepest appreciation to Professor Andy Day for his pragmatic approach to research, his insightful perspectives and words of wisdom that were always timely. Your unwavering support has meant the world to me – it has not only kept me on course but of sound mind. I will be forever grateful for your contributions. To my oldest mentor, Professor Ian Coyle. The lessons I have gained from you will remain with me forever. The abundance of expertise and knowledge that you so willingly shared have shaped me to be the professional I am today. Thank you for believing in me.

I also extend sincere thanks to Dr Nicole Moulday and Associate Professors Matthew Fuller-Tyszkiewicz, Linda Byrne, and Nicki Dowling for their unreserved assistance, encouragement and motivational support throughout this project.

To my peers, it was a pleasure to experience this journey along-side you. I value our shared passion and desire to be the best that we can be in our respective fields of practice. To my forensic cohort, I look forward to our paths crossing in unexpected ways.
Lastly, to my family, thank you for your incredible patience and unconditional support. Words cannot express my appreciation for the way you have supported me to achieve my dream.
Table of Contents

Deakin University Candidate Declaration ........................................................................... iii

Acknowledgements ........................................................................................................ iv

Table of Contents ........................................................................................................... vi

List of Tables ................................................................................................................ xii

List of Figures ................................................................................................................ xiii

Abstract ......................................................................................................................... xv

Chapter 1. Thesis Overview ............................................................................................ 18

Chapter 2. Introduction .................................................................................................. 19

2.1 What is Deception? .................................................................................................. 19

2.2 What is the Problem? .............................................................................................. 21

2.2.1 Looks Can be Deceiving: Relying on witness demeanour to determine witness credibility .................................................................................................................. 21

2.2.2 Accuracy in Detecting Deception: Fact or fiction? ........................................... 24

2.3 Why is this Important? ........................................................................................... 29

2.3.1 Dangerous Decision Theory: First impressions matter ....................................... 29

2.3.2 Judicial Directions: A case of the blind leading the blind and the need to safeguard legal decision-making ....................................................................................... 32

2.3.3 Techniques in Detecting Deception .................................................................. 37

2.3.4 Training in Detecting Deception: Does practice make perfect? .................... 41
Chapter 3. Rationale and Contribution of Research

Chapter 4. Aims and Hypotheses

4.1 Aims

4.2 Hypotheses

Chapter 5. Method

5.1 Research Design

5.2 Participants

5.3 Materials

5.3.1 Expert Witness DVD

5.3.2 Written Judicial Direction

5.3.3 Oral Judicial Direction DVD

5.3.4 Combined Expert Witness and Judicial Direction DVD

5.4 Measures

5.5 Procedure

5.5.1 Participant Recruitment

5.5.2 Questionnaire

5.5.3 Advice on Indicators of Deception: The experimental conditions

5.6 Data Analytic Strategy

5.6.1 Data Screening and Cleaning

5.6.2 New Variables

5.6.3 Statistical Techniques
Chapter 6. Study One: Results..............................................................................................................69

6.1 Does Advice Impact Jurors’ Perceptions of Cues? Assessing Shifts from Baseline:
   Analyses within-groups..................................................................................................................70

6.1.1 Reliable Deception Cues........................................................................................................70

6.1.2 Unreliable Deception Cues .....................................................................................................73

6.1.3 Reliable Cues versus Unreliable Cues ...................................................................................77

6.2 Does Advice Impact Jurors’ Perceptions of Cues? Assessing Differences Between
   Advice and No Advice: Analyses between-groups. .....................................................................81

6.3 What Combination of Nonverbal Cues are Considered Most Important?” .................84

6.3.1 Top Five Most Important Cues .............................................................................................84

6.3.2 Most Commonly Preferred Cues ...........................................................................................88

6.4 Does Exposure to Advice Influence Juror Confidence in Detecting Deception? ..........91

Chapter 7. Discussion.......................................................................................................................92

7.1 Does Advice Impact Jurors’ Perceptions of Nonverbal Cues of Deception? .................92

7.1.1 Assessing Shifts from Baseline: Analyses within-groups.................................................92

7.1.2 Assessing Differences Between Advice and No Advice: Analyses between-
    groups. ........................................................................................................................................101

7.1.3 Overall Conclusions .............................................................................................................104

7.2 What Combination of Nonverbal Cues do Jurors Considered Most Important?........111

7.3 Does Advice Influence Juror Confidence in Detecting Deception? ...............................112

7.4 Chapter Summary....................................................................................................................114
Chapter 8. Study Two ........................................................................................................... 116

8.1 Introduction .................................................................................................................. 116

8.1.1 Delayed Presentation of Evidence: Only a matter of time? ............................... 116

Chapter 9. Aims and Hypotheses ..................................................................................... 118

9.1 Aim ............................................................................................................................... 118

9.2 Hypothesis ................................................................................................................... 118

Chapter 10. Results ......................................................................................................... 119

10.1 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor? Assessing
Shifts from Baseline: Analyses within-groups. ................................................................. 120

10.1.1 Reliable Deception Cues ....................................................................................... 120

10.1.2 Unreliable Deception Cues .................................................................................... 123

10.1.3 Reliable versus Unreliable Cues When Delay is a Factor .................................... 127

10.2 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor? Assessing
Differences Between Advice and No Advice: Analyses between-groups ...................... 131

10.3 What Combination of Nonverbal Deception Cues are Considered Most Important Over
Time? ...................................................................................................................................... 134

10.3.1 Top Five Most Important Cues ............................................................................. 134

10.3.2 Most Commonly Preferred Cues ........................................................................... 138

10.4 Does Exposure to Advice Influence Jurors’ Confidence Over Time? ...................... 141

Chapter 11. Discussion .................................................................................................... 142
11.1 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor? .......... 142

11.1.1 Assessing Shifts from Baseline: Analyses within-groups................................. 142

11.1.2 Assessing Differences Between Advice and No Advice: Analyses between-
   groups. .................................................................................................................. 147

11.2 Overall Conclusions ............................................................................................... 150

11.2.1 Perceptions of Deception Over Time When No Instruction is Given .............. 150

11.2.2 The Effect of Instruction on Perceptions of Deception Over Time............... 152

11.3 What Combination of Nonverbal Deception Cues do Jurors Considered Most
   Important? .............................................................................................................. 154

11.4 Does Advice Influence Jurors’ Confidence Over Time? .................................... 156

11.5 Chapter Summary ................................................................................................. 157

Chapter 12. General Discussion .................................................................................. 162

12.1. Aim One: Does the Provision of Advice Influence Jurors’ Opinions of Reliable Indicia
   of Deception? .......................................................................................................... 164

12.2. Aim Two: Does the Provision of Advice Correct Jurors’ Misconceptions of Nonverbal
   Indicia of Deception? ............................................................................................ 165

12.3. Aim three: To Determine Whether the Effect of Advice on Improving Jurors’
   Perceptions Persist Over Time............................................................................. 167

12.4. Overarching Conclusions: Do Instructions to Jurors on Nonverbal Diagnostics of
   Deception Matter? ............................................................................................... 170
List of Tables

Table 1. Study 1 Demographics

Table 2. Study 2 Demographics

Table 3. Nonverbal Deception Cues Assessed by Participants

Table 4. Shifts in Mean Scores for Mock Juror Perceptions from Pre-Test (Q1) to Post Test (Q2) by Cue-Type and Source of Advice

Table 5. Post-Test Mean Scores of Mock Juror Perceptions of Reliable and Unreliable Cues by Source of Advice

Table 6. Between-Group Comparisons of Mean Differences in Mock Juror Perceptions by Cue-Type and Source of Advice

Table 7. Shifts in Mean Scores of Mock Juror Perceptions from Baseline (Q1) to After a One-Week Delay (Q2) by Cue-Type and Source of Advice

Table 8. Post-Test Mean Scores of Mock Juror Perceptions of Reliable and Unreliable Cues by Source of Advice

Table 9. Between-Group Comparisons of Mean Differences in Mock Juror Perceptions by Cue-Type and Source of Advice When Delay is a Factor
List of Figures

Figure 1. Flow and Response Rates of Participants in Study 1 and Study 2.

Figure 2. Mean Scores for Reliable Cues by Advice and Time-to-Assessment.

Figure 3. Mean Scores for Unreliable Cues by Source of Advice and Time-to-Assessment.

Figure 4. Mock Juror Perceptions at Baseline (Q1) and Post-Advice (Q2) by Cue-Type and Source of Advice.

Figure 5. Total Mean Scores for the Reliable and Unreliable Subscales at Baseline (Q1) and Post-Advice (Q2) for Mock Jurors Who Received No Advice (Control).

Figure 6. Nonverbal Cues Ranked Most Important Before and After Advice.

Figure 7. Percentage Change in Nonverbal Cues Ranked Most Important After Advice.

Figure 8. Most Commonly Endorsed Nonverbal Cues Before and After Advice.

Figure 9. Percentage Change in Nonverbal Cues Most Commonly Endorsed After Advice.

Figure 10. Mean scores for the Reliable Subscale by Source of Advice at Baseline (Q1) and After a One-Week Delay (Q2).

Figure 11. Mean Scores for the Unreliable Subscale by Source of Advice at Baseline (Q1) and After a One-Week Delay (Q2).

Figure 12. Mock Juror Perceptions at Baseline (Q1) and After a One-Week Delay (Q2) by Cue-Type and Source of Advice.

Figure 13. Total Mean Scores for the Reliable and Unreliable Subscales at Baseline (Q1) and After a One-Week Delay (Q2) for Mock Jurors Who Received No Advice (Control).
Figure 14. Nonverbal Cues Ranked Most Important Before and After Advice When Delay Was a Factor.

Figure 15. Percentage Change in Nonverbal Cues Ranked Most Important After Advice When Delay Was a Factor.

Figure 16. Most Commonly Endorsed Nonverbal Cues Before and After Advice When Delay Was a Factor.

Figure 17. Percentage Change in Nonverbal Cues Most Commonly Endorsed After Advice When Delay Was a Factor.
Abstract

Legal doctrine dictates that the triers of fact in court cases (i.e. the jury, or a judge in a judge-alone trial) can reliably assess the trustworthiness of a witness by interpreting, among other things, nonverbal demeanour evidence. This faith is misplaced. No matter how learned the triers of fact, and irrespective of their training (O’Sullivan & Ekman, 2004), most people operate at about chance level in detecting deception (Bond & DePaulo, 2006; Gemma, Schertler & Bull, 2019; Morgan, Rabinowitz, Hilts, Weller & Coric, 2013; Wright, Berry & Bird, 2012; Vrij & Turgeon, 2018) and further, jurors’ initial impressions have been found to carry disproportionate weight (Porter & ten Brinke, 2009). This venerated belief can, and does, lead to unsafe and dangerous legal decisions (Baker, Porter, ten Brinke & Mundy, 2016; Denault & Dunbar, 2017; Korva, Porter, O’Conner, Shaw & ten Brinke, 2013; Porter & ten Brinke, 2009, 2010; Wilson & Rule, 2019).

Judicial directions to the triers of fact about witness credibility have a long history (see Alexander v The Queen, 1981; Domican v R, 1992; Longman v The Queen, 1989). However, there is simply no evidence that such directions have any meaningful effect on the decisions of jurors (Spivak, Ogloff & Clough, 2011). This fallibility has prompted interest in providing education to jury members. Nonetheless, although there have been numerous attempts to train jurors in techniques to assess deception such training has no import with respect to decision-making in real courtrooms. Primarily, training has no place in a court of law because training involves, by definition, feedback (Driskell, 2012), which cannot be given in an actual trial. In an actual trial, the only scientific information that can assist the triers of fact in assessing witness demeanour is that which can be adduced by way of expert evidence. Adducing evidence does not and cannot, as a matter of law and practice, involve feedback.

The current program of research comprises two studies that aim to advance work in this area by testing specific tenets of Dangerous Decisions Theory (DDT; Porter & ten Brinke, 2009) which posits that initial judgments of trustworthiness based on nonverbal behaviour form rapidly
and are resistant to change, even in the face of conflicting evidence (Baker, et al, 2016; Porter ten Brinke 2009; 2010). The first study investigates the pervasiveness of initial, baseline perceptions of nonverbal deceptive behaviour, and examines whether or not expert evidence and/or judicial directives on reliable/unreliable indicators of deception can improve the reliability of jurors’ initial perceptions.

In an actual trial it is not uncommon for members of a jury to consider evidence adduced by an expert at a later stage in proceedings, requiring that triers of fact retain the testimony proffered by the expert. This delay introduces decay and interference effects that can undermine the retention of information. However, the effect of this has not previously been investigated in this context. A unique aspect of the current research therefore, is examining the impact of delay on jurors’ perceptions of trustworthiness/deception. Specifically, the second study seeks to investigate the effect of a delay between the presentation of expert evidence and/or judicial directions on juror decision-making regarding perceptions of deception indicia. This delayed consideration of advice could be expected to occur in a trial where the expert evidence or judicial instruction presented earlier in a trial has to be relied upon later when applied to other evidence such as a factual witness.

In this thesis, findings are reported that suggest that jurors’ perceptions of nonverbal indicia may be enhanced with instructions on reliable/unreliable cues of deception and that these effects may be both immediate and in the long-term with the greatest gains in accuracy found for advice adduced by an expert witness. A number of implications for theory and forensic practice were identified with logical and important steps worthy of future investigation highlighted. The contribution of this program of research, from both a theoretical and applied perspective, lies in the mixed support for DDT that attests initial judgements are resistant to change even in the face of conflicting evidence. In doing so, this thesis presents a strong argument for enhancing the professional practice of forensic psychology in educating triers of fact on behavioural, nonverbal diagnostics of deception.
“Lies, my dear boy, can easily be recognized. There are two kind of them: those with short legs, and those with long noses. Your kind have long noses.”

Chapter 1. Thesis Overview

The current program of research comprises two studies that investigate whether jurors’ perceptions of nonverbal cues of deception could be enhanced with education through the provision of advice, whether the source of that advice is relevant, and whether or not any enhancement of jurors’ perceptions of deception hold over time.

The thesis begins with an overview of the literature on detecting deception. Firstly, terms are defined and the reader is orientated to the problems associated with interpreting demeanour evidence when assessing witness credibility, and why the problem is an important one. In chapter three, the rationale and contribution of research is set out with a strong argument made for the value of scientific replication. The background to the thesis concludes at chapter four where the aims and hypotheses for Study one are presented. A general method is detailed at chapter five eliminating, where possible, duplication and repetition between studies. The results for Study one are presented in chapter six with a discussion of the findings thereafter. Study two is introduced at chapter eight with the aim and hypothesis outlined in chapter nine with results, thereafter. An analysis of the findings follow in the discussion at chapter eleven. The thesis concludes with a general, overarching discussion at chapter twelve. Here, the findings of both studies are interpreted together with implications for theory and forensic practice considered. Lastly, the strengths and limitations of the thesis are noted along with areas worthy of future research.
Chapter 2. Introduction

2.1 What is Deception?

At first glance deception might be described as the act of lying: the telling of an untruth or an intentionally false statement ("Lie [Def. 2]", n.d.). This definition however, overlooks a number of complexities. To tell a lie implies that incorrect, misinformation has been exchanged but not all exchanges of inaccurate information imply a lie or the intent to deceive. For example, information may be inaccurate and unreliable because of a false belief or false memory (Lanley & Loftus, 2013; Loftus & Pickrell, 1995; Schacter & Loftus, 2013; Vrij, Granhag, & Porter, 2010). Consider a witness’s testimony in court: on the one hand testimony may be inaccurate because the witness is lying with the intent to deceive, or because their reconstruction of events are wrong. Although the information is inaccurate and unreliable in both circumstances, reconstruction is a normal part of memory process (Tulving & Thomson, 1979), that due to post-event contamination may result in a false belief (Loftus, 1975; Loftus & Hoffman, 1989). Although false beliefs, false memories and misremembering may undermine the accuracy of information exchanged between one or more parties, they do not, by that fact alone, involve deceit.

The terms lying and deception will be used interchangeably throughout as related concepts pertaining to the antithetical, truthfulness. The terminology adopted is consistent with that employed by seminal authors in the field (DePaulo et al., 2003; Vrij, 2000; Frank & Svetieva, 2013; Levine, 2014a; Knapp et al., 2016; also see Kalbfleisch & Docan-Morgan, 2019 for an overview). The Oxford Online English Dictionary defines deception as “the action of deceiving someone” with deceit characterised as the “action or practice of deceiving someone by concealing or misrepresenting the truth” ("Deceit [Def. 1]", nd; "Deception [Def. 1]", nd). This definition encapsulates the act of making a false statement (a lie) but also captures the act of concealment. When considering the role of deception, the act of concealment is important – especially so when withholding, concealing, or refusing information has advantage, either directly or indirectly.
For the purposes of the current research deception is defined as the intention to mislead. This stance is consistent with prevailing doctrine on deception and reflects sentiments espoused by key contributors such as DePaulo and her colleagues (2003) who define deception as “a deliberate attempt to mislead others” (p.74) and Vrij (2008) who attests deception as “a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue” (p.14).

Detection of deception has received considerable attention with respect to potentially assisting police investigations (see Ioannou & Hammond, 2015; Vrij & Granhag, 2007; Vrij & Semin, 1996; Vrij & Semin, 2004), and more recently, national security operations in trying to assess intent of suspects (Alison, Alison, Noone, Elntib, Waring & Christiansen, 2014; Clemens, Granhag & Strömwall, 2013; Granhag & Mac Giolla, 2014; Levine, 2014b, 2015; Weinberger, 2010). Yet, little consideration has been given to detection of deception within curial proceedings (for recent commentary see Denault & Dunbar, 2019; Denault & Jupe, 2018; Denault, Jupe, Dodier & Rochat, 2018; Leach et al., 2016). In a trial by jury, the role of the tribunal is to reach a verdict by considering the evidence introduced at trial and the directions of the presiding judge. As the triers of fact, the aim of the jury is to uncover the truth through establishing the facts of the case. However, when contests of credibility are at play this task becomes increasingly difficult, and the detection of deception becomes of great import.

In considering oral evidence adduced through witness testimony (witnesses of fact or opinion), the jury must determine the credibility of the witness and the weight to be placed on the testimony presented. In evaluating witness credibility one area jurors must consider is the truthfulness of a witness: as judges of fact, they must decide whether a witness has told the truth or testified falsely. In doing so, jurors evaluate, among other things (i.e. accuracy and bias) witness demeanour: relying on the behaviour and appearance of a witness to detect deception. Although witness credibility can be considered to involve a number of factors (for example the reliability,
INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION

accuracy, or expertise of a witness), for the purposes of the current research, credibility is understood in terms of the truthfulness, trustworthiness and believability of a witness.

2.2 What is the Problem?

2.2.1 Looks Can be Deceiving: Relying on witness demeanour to determine witness credibility

In their role as triers of fact, jurors’ decision-making is largely informed by evidence adduced through the testimony of experts and witnesses of fact. A difficulty posed however, is in evaluating conflicting testimony and then considering its impact on the truthfulness of a witness (Denault & Dunbar, 2019; Porter, ten Brinke, & Gustaw, 2010). Legal doctrine assumes that triers of fact evaluate demeanour by observing, among other things, nonverbal behaviour in assessing the veracity of a witness (Morales v Artuz, 2002; Qureshi, 2014). The Western Australian (WA) and New South Wales (NSW) Bench Books posit that members of the jury “assess a persons’ credibility or trustworthiness on their perceptions of the person’s demeanour” (Equality before the Law Bench Book, WA, 2009, p. 300; Equality before the Law Bench Book, NSW, 2014, p. 313). Similar sentiments are echoed in the Victorian Bench Book which advocates for “the need to challenge a witness’ credibility or examine his or her demeanour” if a fair trial is to be ensured (Open Courts Bench Book, Vic, 2014 at 4.3.3).

In R v. François (1994) McLachlin J. of the Canadian Supreme Court observed: “the jury must decide whether it believes the witness’s story in whole or in part. That determination turns… on the demeanour of the witness and the common sense of the jury”. This observation implies that the truthfulness of a witness rests in the interpretation and evaluation of witness demeanour. Similarly, in 2012, the Supreme Court of Canada ruled in R v. S that an important factor in trial fairness is the deceit indicated by facial cues of witnesses, and that observations of the witnesses’ face while testifying is “too deeply rooted in the criminal justice system to be set aside” (para, 27). As Coyle and Thomson (2014) note, assessing the demeanour of a witness means, in practice
“attending to verbal and nonverbal indicia of truthfulness… [and] where competing versions of events are given, the assessment of such indicia are of great import” (p. 478).

The task of assessing witness credibility is deemed the “bread and butter” of jurors (Porter & ten Brinke, 2009, p. 120), most likely because it is presumed “common sense” (Vrij & Turgeon, 2018, p. 232). So much so that assessments of credibility are a function reserved for the sole province for the jury whereby adducing evidence on witness credibility is prohibited by the credibility rule (Evidence Act 2008, Vic, s.102). This position was espoused in R v. Marquard (1993), in which the Canadian Supreme Court unambiguously noted “that the ultimate conclusions as to the credibility or truthfulness of a particular witness is for the trier of fact, and is not the proper subject of expert opinion” (at 228). Although originating from British common law roots, Canadian case law may be used as an authority for Australian cases (Lefler, 2002). Such a direction as that stated in R v Marquard (1993) assumes that a common sense approach is reliable. However, this approach ignores the fact that assessing the truthfulness of a witness is a highly complex and unreliable task and hence, verdict deliberations may be fatally flawed if based on jurors’ subjective, intuitive interpretations of nonverbal behaviour that are invalid and unreliable (Baker et al., 2016; Bennett, 2015; DePaulo, et al, 2003; Porter & ten Brinke, 2010; Sporer & Schwandt, 2007; Snook, McCardle, Fahmy, & House, 2017; Vrij, Hartwig & Granhag, 2019; Vrij & Turgeon, 2018).

Demeanour evidence refers, in part, to a witness’s behavioural cues observed while testifying, including facial expressions, eye contact, attitude, body language, pauses in speech, hesitation, sincerity, use of gestures, candour, tone of voice, dress, grooming, and level of confidence (Ogden, 2000). McKimmie, Masser and Bongiorno (2014) studied the effect of witness demeanour on mock jurors’ perceptions by investigating behavioural stereotypes of deception on convincing and unconvincing witness testimony. They found that jurors were less influenced by convincing testimony when a witness displayed behavioural stereotypes of deception than when a witness displayed behavioural stereotypes of trustworthiness (McKimmie et al., 2014). This
finding suggests that jurors give more weight to the demeanour of a witness than the strength of their testimony. However, the study confounded the use of reliable and unreliable indicia of deception: when testifying, witnesses displayed both accurate stereotypes of deception (i.e. high pitched voice), and inaccurate stereotypes of deception (i.e. gaze aversion). As a result, McKimmie et al., (2014) conducted a follow-up study to test whether the evaluation of witness testimony was influenced by the presence of reliable cues, unreliable cues or both. It is of some note that mock jurors were primarily influenced by unreliable, inaccurate behavioural stereotypes of deception such as gaze aversion. And thus, paradoxically, they concluded that instructing jurors to pay attention to witness demeanour may undermine assessments of witness credibility (McKimmie et al., 2014).

The potential for error is unequivocal when unreliable indicia of deception/truthfulness guide legal decision-making (Baker et al., 2016; Kramer & Van Volkom, 2018; Minzner, 2008). An egregious example of this is found in the Supreme Court of Queensland (2006) Equal Treatment Bench Book wherein it is averred: “An impressive witness according to Anglo-Australian culture will look his or her questioner in the eye and answer questions confidently and clearly” (p.75). This statement suggests that eye contact, confidence and speaking unclearly are reliable markers of truthfulness. They are not (DePaulo et al., 2003; Vrij, 2008).

The problem of relying on demeanour evidence to infer witness credibility is that the weight attributed to witness demeanour ignores the literature consensus that shows triers of fact typically perform no better than chance in accurately discriminating truth-telling from lying (O’Sullivan & Ekman, 2004), however; they think they do (Porter & ten Brinke, 2009). On the contrary, triers of fact fall prey to unreliable indicia of deception when making a judgement. These errors in thinking also hold for members of the judiciary (Denault & Dunbar, 2017, 2019; Granhag & Strömwall, 2004; Vrij & Mann, 2004). Not only do the adjudicature hold the same misplaced beliefs as laypersons (Burns, 2016) but, at times, their limited scientific knowledge renders them vulnerable
to the dangers of pseudoscience (Denault, et al., 2019; Denault & Jupe, 2018). Especially concerning is when judicial members receive training in nonverbal indicators of deception that lacks any basis in empirical evidence (Denault, in press; Denault & Dunbar, 2017; Denault & Jupe, 2019). All this points to the fact that judicial decision-making processes are not immune from basic human biases (Baker et al., 2016; Bennett, 2015; Burns, 2016; Kramer & Van Volkom, 2018; Porter & ten Brinke, 2009).

2.2.2 Accuracy in Detecting Deception: Fact or fiction?

To reiterate, people are poor at detecting lies with most people’s ability to detect deception only slightly better than chance (Bond & DePaulo, 2006; Colwell, James-Kangal, Hiscock-Anisman & Plelan, 2015; Hauch, Sporer, Michael & Meissner, 2016; Hartwig & Bond, 2014; Morgan, et al, 2013; Wright, Berry & Bird, 2012; Vrij, 2008). A meta-analysis of 206 studies on veracity judgements found that the average rate of performance in detecting deception was 54% (Bond & DePaulo, 2006). When comparing ability to correctly discriminate lies from truth, Bond and DePaulo (2006) found only 47% of people correctly identified lies, and 61% accurately detected statements that were truthful. Furthermore, their meta-analysis revealed that detection accuracy was poorer when based on nonverbal indicia when compared to verbal indicia (Bond & DePaulo, 2006).

Two schools of thought seek to explain this poor performance. First, lie-detection is erroneous because people tend to rely on invalid cues when judging deception (Levine & McCormack, 2014; Vrij et al., 2019). Second, valid and reliable indicia of deception are weak, making detection of deception difficult and unreliable (Hartwig & Bond, 2011; Strömwall, Granhag, & Hartwig, 2004; Masip, 2017; Vrij et al., 2010; Vrij & Turgeon, 2018). Nonetheless, these two perspectives are not mutually exclusive. There is some preliminary evidence to suggest that both explanations have validity. That is, people pay attention to both incorrect and correct cues but the subtle differences between lying and truth-telling render valid cues hard to detect
(McKimmie et al., 2014; Vrij et al., 2010). Accordingly, objective truthfulness is often easily missed, misinterpreted or overlooked due to subtle differences and/or wrong techniques (Coyle & Thomson, 2014).

Contrary to common knowledge, professionals experienced in trying to detect deception are no better at discriminating truth-tellers from liars (Bond & DePaulo, 2006). Based on the premise that experience in dealing with liars positively influences skills in detection, trained professionals are often presumed to be better at detecting deception than laypersons (Vrij & Mann, 2001). A considerable body of work has investigated demeanour-based detection accuracy among police investigators, custom and immigration officials, security and intelligence personnel, psychologists and judges (Akehurst, Kohnken, Vrij & Bull, 1996; Bogaard & Meijer, 2018; Delmas, et al, 2019; DePaulo, Kashy, Kirkendol, Wyer & Epstein, 1996; DePaulo & Pfeifer, 1986; Elkman & O’Sullivan, 1991; Mann & Bull, 2004; Mann & Vrij, 2006; Mann, Vrij & Bull, 2006; Masip & Herrero, 2015b; Wright & Wheatcroft, 2017; Wright, Wagstaff & Wheatcroft, 2015).

While some individual evidence suggests that police, federal judges and forensic psychologists perform significantly better than chance (Ekman, O'Sullivan, & Frank, 1999) group data indicates differently (Sternglanz, Morris, Morrow, & Braverman, 2019). In a meta-analysis, Aamodt and Custer (2006) found no relationship between detection accuracy and experience ($k = 13$, $d = -0.16$), or between detection accuracy and education ($k = 3$, $d = 0.06$). In fact, their results indicated that the ability to detect deception did not differ between professionals (55.5%) and laypersons (54.2%). This finding is consistent with a larger meta-analysis conducted by Bond and DePaulo (2006) that showed experts did not outperform laypersons in differentiating liars from truth-tellers ($k = 20$, $d = -0.03$).

A common error among professionals is their tendency to overestimate their ability to detect deceit (Vrij et al., 2010). In a review of 24 studies on lie-detection abilities among professionals (predominantly police) Vrij (2008) found a total average accuracy rate of 55.9%. In
seven of eight comparative studies Vrij’s (2008) analysis found no differences in detection accuracy between police and laypersons. Indeed, one study revealed laypersons outperformed police in detecting lies (Vrij, 2008). At the time of Vrij’s (2008) review no one study found police to be superior to laypersons in discriminating deception from truth-telling. Similarly, O'Sullivan and Ekman (2004) believed some professional groups would outperform others in detecting deception. Nonetheless, they found that most professions – police officers, agents from the Central Intelligence Agency and Federal Bureau of Investigation, lawyers, university students, psychotherapists, judges and forensic psychologists – performed little better than chance. Only one group, the United States Secret Service, did significantly better than chance in detecting deception with more than half obtaining accuracy scores of 70% or more. This difference, they claimed, can be explained by the absence of confirmatory bias that otherwise influences veracity judgements made by police; suspects are presumed to lie so investigators seek evidence to confirm their belief that the suspect is ‘guilty’.

A number of cognitive heuristics contribute to systematic errors and biases in veracity judgements (for an exhaustive account see Vrij, 2008). Most notably, truth-bias and confirmatory biases are among the most consistently reported (Kramer & Van Volkom, 2018; Colwell, 2005; Vrij et al., 2010). Truth-bias refers to the tendency to assess statements as truthful rather than deceptive and in part, is responsible for increased accuracy in detecting truth over lies (Bond & DePaulo, 2006; Levine, Park & McCornack, 1999; Zuckerman, DePaulo, & Rosenthal, 1981). This decision-making error was illustrated in the review by Bond and DePaulo (2006) which found 56% of statements were judged as honest and 44% were judged as deceptive despite an equal number of honest and deceptive accounts. The tendency to be credulous is attributed to the availability heuristic (Tversky & Kahneman, 1981); people are more inclined to assume someone is telling the truth because truthful encounters are experienced more routinely than deceptive ones (Vrij, 2008). Interestingly, this bias is reversed, or at least diminished among professions with experience in trying to detect deception (Masip, Alonso, Herrero & Garrido, 2016; Meissner, 2002). Accordingly
it can be said that our experiences influence our expectations and our expectations shape the way we interpret information.

Commonly referred to as confirmatory bias, errors occur when appropriate corrections are not made to one’s initial position despite the emergence of conflicting evidence (Tversky & Kahneman, 1981; Vrij, 2008). This cognitive heuristic has been shown to affect decision-making ability among triers of fact with evidence to suggest that jurors discount new information (i.e. evidence or judicial direction) because of a predisposition to adhere to their initial hypotheses (Baker et al., 2016; Festinger, 1957; Korva et al., 2013; Porter & ten Brinke, 2009). The consequence of confirmatory bias is that jurors’ initial impressions formed early in a trial due to misconceptions about what constitutes indicators of truthfulness and deception may be fatal to a proper reception of all the evidence later in the trial (Baker et al., 2016; Porter & ten Brinke, 2009).

Judgements of deception are strongly associated with stereotypes that are both unreliable and pervasive across cultures (Villar, Arciuli, & Paterson, 2013; Vrij et al., 2019). In an international study coordinated by the Global Deception Research Team (2006) stereotypes about liars were examined across 75 different countries and 43 different languages. Despite being an unreliable indicator, gaze aversion was identified as the most dominant stereotype of deception across all cultures accounting for 71.5% of beliefs reported (DePaulo et al., 2003; Global Deception Research Team, 2006). This finding is attributed to deception stereotypes based on false beliefs learned through observation and/or experience (Global Deception Research Team, 2006; Hurley, Griffin, & Stefanone, 2014; Villar et al., 2013; Vrij et al., 2010). Moreover, the uniformity of these stereotypes suggest some partial support that false beliefs are widespread and shared among both laypersons and erudite professionals (Vrij et al., 2010, 2019).

The most popular beliefs about how deceptive people behave are associated, sometimes incorrectly, with the view that lying is “bad” (Vrij et al., 2010). Based on this premise, deceivers are expected to behave nervously with increased fidgeting, grooming gestures, and overall
restlessness perceived as indicators of anxiety (Greely & Illes, 2007; Vrij & Turgeon, 2018). Similarly, lying is presumed to be associated with feelings of shame with avoided eye contact and reduced engagement indicating withdrawal and distancing (DePaulo et al., 2003; Vrij et al., 2010) yet these along with the overwhelming majority of deception stereotypes are flawed (DePaulo et al., 2003; Global Deception Research Team, 2006). According to DePaulo and colleagues (2003), a large number of behaviours are stereotypically attributed to deception with only a few found to be correlated with deceit. In their 2003 study, they showed that 78% (118 of 158) of all nonverbal cues examined in the literature had no relationship with deceit (DePaulo et al., 2003).

Although the interpretation of behavioural indicia is highly subjective, and is typically misinformed by false beliefs about how liars look and behave (Strömwall et al., 2004; Vrij et al., 2010), nonverbal behaviour has been shown to shape impressions of people, including personality traits, gender roles, sexual orientation, romantic involvement, relationship potential, status, and competence (Vrij et al., 2010). In fact, the effects of facial appearance alone on decision-making are pervasive (Bull, 2006; Berry & Zebrowitz-McArthur, 1988). Impressions of physical attractiveness have been shown to influence mock juror deliberations with attractive defendants less likely to be convicted, punished or held criminally responsible than unattractive defendants, and attractive defendants are perceived as happier, more likeable and trustworthy than unattractive defendants (Abel & Watters, 2005; Abwender & Hough, 2001; Darby & Jeffers, 1988; Patry, 2008; Wuensch & Moore, 2004). Equally, baby-faced defendants are less likely to receive harsher judicial penalties than people with mature faces (Zebrowitz & McDonald, 1991). And these assessments form rapidly and effortlessly.

In a series of studies conducted by Willis and Todorov (2006) initial impressions based on facial appearance were explored. They examined three exposure intervals: 100ms, 500ms and 1,000ms for trustworthiness, likeability, competence, aggressiveness and attractiveness. Based on a sample of 245 randomly assigned students they found that minimal exposure time (100ms) was
sufficient for impressions based on facial appearance and that these impressions remained virtually unchanged overtime. In addition, they demonstrated that confidence in initial judgements increased with time (Willis & Todorov, 2006, p. 596-597). Contrary to expectation, their strongest findings were for judgements of trustworthiness over attractiveness. Specifically, impressions of trustworthiness formed the quickest and an individual’s confidence in their initial judgement of trustworthiness was boosted with time (Willis & Todorov, 2006).

In summary, unguided judgements of trustworthiness based on demeanour evidence such as nonverbal behaviour and facial characteristics are unreliable. People are poor at discriminating truthtellers from liars and drawing on professional experience and stereotypes of deceptive behaviour as the means to assess credibility is flawed. This is important because in a trial by jury the demeanour of a witness is attributed significant weight in aiding the jury in its determination of witness credibility. In addition, assessments of credibility may be further compounded by initial impressions that form rapidly and are perceived with greater confidence over time. This is exacerbated when judgements of trustworthiness are based on nonverbal behaviour as these impressions are unlikely to change, even in the face of conflicting evidence (Porter & ten Brinke, 2009).

2.3 Why is this Important?

2.3.1 Dangerous Decision Theory: First impressions matter

Dangerous Decision Theory (DDT) is a theory which seeks to explain how psychological processes contribute to flawed judgements of deception that can, and do, lead to unsafe judicial outcomes (Baker et al., 2016; Porter & ten Brinke, 2009, 2010). It proposes that nonverbal behaviour exhibited by a defendant’s face plays a critical role in initiating a serious of unsafe decisions concerning credibility and ultimately culpability – judgements of witness trustworthiness occur rapidly and are subjectively internalised as intuition (Baker et al., 2016). However, just because judgements of trustworthiness are automatic does not mean that they are necessarily
accurate. The first component of the theory thus predicts that intuitive assessments of trustworthiness based on nonverbal behaviour may be unreliable.

Porter and ten Brinke (2009) support this argument by citing prior research that examines the accuracy of initial judgments based on the faces of two groups differing in trustworthiness (Porter, England, Juodis, ten Brinke, & Wilson, 2008). The two target groups comprised Nobel Peace Prize winners and humanitarians and criminals from America’s most wanted list. Participants viewed a total of 34 faces for either 1/10th of a second or 30 seconds and were asked assess the trustworthiness of each face. While the mean judgement accuracy did not differ by exposure, initial judgements of untrustworthiness were less accurate (mean of 48.8%) than trustworthy judgements (Porter et al., 2008). Following ratings of trustworthiness, participants were advised of the two target groups and asked to predict group membership. Judgement accuracy for group membership was found to be just above chance. Based on these findings Porter and colleagues (2008) concluded that while intuitive evaluations contribute to judgements of trustworthiness on a small scale, judgement errors are common.

In addition to judgement errors that bias decision-maker’s assessment of witness truthfulness, DDT proposes that these same biases may also affect the interpretation and integration of information presented at trial. Indeed, in a courtroom environment assessments of evidence by jurors and/or judges may be inherently flawed if the interpretation of evidence adduced is biased by initial, ill-informed assessments of trustworthiness (Baker et al., 2016; Porter & ten Brinke, 2009). Porter, ten Brinke and Gustaw (2010) provide preliminary evidence to support this. They examined participant’s ability to assimilate evidence on an individual’s alleged involvement in a crime along with their initial impression of the accused. With a sample of 80 student participants, the study used a 2 (severity) x 2 (trustworthiness) mixed design where two crime types (severe and non-severe) and two impressions (trustworthy and untrustworthy) were tested. Participants were randomly assigned to either the severe-crime condition (a violent murder and a sexually motivated
murder) or the non-severe crime scenario (a car theft and a fraud). Presentation of crime vignettes and defendant photographs were counterbalanced so that each vignette was equally associated with trustworthy and untrustworthy images. The images evaluated were adopted from an earlier, pilot study that had participants rate 20 photographs of Caucasian males depicting neutral expressions. Trustworthiness had been determined by a previous sample of respondents on a one (not at all) to seven (highly) scale for each image. After rating photographs of the accused, completing a distractor task, and reading the crime vignette, participants were presented with evidence associated with the crime. Participants were exposed to five neutral pieces of evidence, five increasingly incriminating pieces of evidence and one exonerating piece of evidence. Participants were instructed to indicate their verdict (guilty/not guilty) and confidence in their decision after each piece of evidence was presented. This procedure was then repeated for the counterbalanced scenario.

Porter and ten Brinke (2010) concluded support for DDT as a model for biased legal decision making on the basis that less evidence was necessary to convict an untrustworthy looking defendant for the same crime than was required for a person perceived as trustworthy. Further, they demonstrated that guilty verdicts were reached on fewer pieces of neutral evidence than was required for the same crime when the person was perceived to be trustworthy. However, the effects of exonerating (conflicting) evidence on decisions of guilt were not significant. While the researchers drew attention to their “pattern of means” in support of the premise that potential jurors are unlikely to accept the innocence of defendant judged as untrustworthy, even in the face of influential exonerating evidence, this proposition cannot be concluded with confidence (Porter & ten Brinke, 2010, p. 487).

This distinction is an important one given that a central tenant of DDT is that once impressions are formed they are unlikely to change, even in the face of conflicting evidence (Baker, et al., 2016; Porter & ten Brinke, 2009). Drawing on the principle of cognitive dissonance and the
imperviousness of cognitive heuristics to intervention (Coyle & Thomson, 2014), the theory predicts that human biases are responsible for negatively impacting evaluations of truthfulness but also the consideration of evidence, especially so in forensic settings. This bias is exacerbated by courtroom practices “despite the lack of validity associated with [juror’s] intuitive assessments of trustworthiness… they are encouraged by judges” (Porter & ten Brinke, 2009, p. 62).

2.3.2 Judicial Directions: A case of the blind leading the blind and the need to safeguard legal decision-making.

Written directives to the jury are the favoured method by which judges inform jurors about the law in each case, and how the law should be applied to the facts presented. Judicial directions to the tribunal of fact on the issue of witness credibility have a long history. So much so that judges have been known to disregard the utility of psychological knowledge on the basis that human behaviour is considered to be a matter of ‘common sense’ and within the common experience of the jury. In R v. Marquard (1993), the court concluded that determining the truthfulness of a witness is common sense: “Credibility is a matter within the competence of laypeople. Ordinary people draw conclusions about whether someone is lying or telling the truth on a daily basis” (p. 248). Although common sense is an inevitable feature of judicial decision-making, judicial use of common sense as to the reasonableness and normality of behaviour can present an avenue through which faulty thinking, error and discrimination influence the law (Burns, 2016). Bagaric and McConvill (2005) argue that empirical evidence that assists the court in establishing truth ought to be admissible, and that exclusionary rules (i.e. the credibility rule and common knowledge principles) should be abolished in favour of a ‘relevancy’ test (Frye v. The United States, 1923). To ascertain the truth, Bagaric and McConvill (2005) suggest that this approach would be a safer, more reliable one than relying on “‘common sense’, judicial hunches or other non-rational sentiment” (p.13).
Notwithstanding this, Australian courts are divided on their stance regarding the admissibility of expert psychological opinion that talks to common sense and the issue of credibility. With the development of a Uniform Evidence Law (s. 80) that saw the recent abolishment of the Common Knowledge Rule in Victoria, the ACT and some states of the Commonwealth (Evidence Act of the Australian Capital Territory, 2011; Evidence Act of New South Wales, 1995; Evidence Act of Tasmania, 2001) the courts are less likely to reject opinion evidence when the facts in consideration need the help of an expert (Australian Law Reform Commission, New South Wales Law Reform Commission, & Victorian Law Reform Commission, 2006). Judges have routinely commented on areas within the ambit of psychological knowledge such as memory, eyewitness identification, witness demeanour and the credibility of child witnesses without the benefit of expert evidence on these topics (Coyle & Field, 2013). It is not surprising then, that judicial directions on these matters have been inaccurate. This misplaced advice presents an inherent legal problem given that juries are required to follow the ‘directions’ set forth by a trial judge (Coyle & Field, 2013, p.99; and for a dedicated analysis of illustrative cases of “judicial expert evidence” and the resultant issues see p.94-103). This problem is of significant import when it comes to matters of witness demeanour, particularly when judicial directions that fall within the domain of expert psychological knowledge are concerned, and expert evidence has not been adduced due to outmoded legal doctrines.

The efficacy of judicial directions has been the subject of review by the Victorian Law Reform Commission (2009). This review promoted reform in the way directions were determined and communicated that saw the enactment of the Jury Directions Act (Vic, 2013, 2015) and now Jury Directions and Other Acts Amendment Act (Vic, 2017). While the Act provides guidance on what matters of law and evidence should be directed to a jury and when, the extent to which the triers of fact comprehend judicial directions has been largely overlooked.
One study conducted in the United Kingdom (UK) by Thomas (2010), has investigated jurors’ understanding of judicial instructions. A significant strength of the study was its sampling technique: 797 actual jurors who had been called for jury service but were later relinquished of their role were sampled from across the UK. The study tested jurors’ perceived ability and actual ability in understanding judicial directions. The research also examined whether or not jurors’ understanding of directions improved when provided in writing. On perceived ability, jurors were asked how easy or difficult they felt it was to understand a judge’s oral instruction. On actual ability, jurors were asked to identify two legal questions that they were explicitly directed to answer by the judge.

Thomas (2010) found that while the majority of jurors believed that they could easily understand directions from a judge, less than a third of jurors actually understood the judicial direction given. This finding is consistent with extant research confirming juror’s inability to understand judicial instruction (see Charrow & Charrow, 1979; Dumas, 2000; Lieberman & Sales, 1997; Mueller-Johnson, Dhami & Lundrigan, 2018; Steele & Thornburg, 1988). In contrast, when judicial directions were provided in writing, jurors’ comprehension was greater (48%) than those who only heard the instruction (31%). Indeed, while the written and oral direction contained the same content, Thomas’s (2010) procedure suggests that the written summaries were distributed to each juror at the start of the judge’s legal directions. It is not clear therefore, whether jurors who received the written direction were, by way of design, exposed to both the written and oral instruction at the same time. At issue is the generalisability of their finding; if jurors were able to read the written summary along with listening to the oral instruction the combined effect of presentation modality on comprehension ought to be considered.

In recent times, the legal fraternity has started to recognise limits with respect to the assessment of witness demeanour. Numerous judgements, both local and international, caution against attributing undue weight to witness demeanour (for local authorities see Abalos v. Australia.

This position is reflected in the leading Australian legal authority Fox v. Percy (2003) wherein the High Court of Australia warned against the fallibility of courts in overstating witness credibility in the face of objective evidence. Specifically, the ruling draws attention to the “dangers of too readily drawing conclusions about truthfulness and reliability solely or mainly from the appearance of witnesses”. Therein, Kirby J. cited, with approbation, the words of Lord Justice Atkins who in 1924 in Société d’Avances Commerciales (Société Anonyme Egyptienne) v. Merchants’ Marine Insurance Co (The "Palitana") remarked: "... I think that an ounce of intrinsic merit or demerit in the evidence, that is to say, the value of the comparison of evidence with known facts, *is worth pounds of demeanour*” (at 32, emphasis added). Kirby J. also made the pungent observation that “judges have become more aware of scientific research that has cast doubt on the ability of judges (or anyone else) to tell truth from falsehood accurately on the basis of such appearances” (at 31).

The provision for reliable, psychological evidence on human behaviour would protect against jurors relying on misplaced common sense when assessing witness demeanour, but also, would mitigate the need for members of the judiciary to proffer psychology from the bench (Coyle & Field, 2013). The decision set out in Aytugrul v. The Queen (2012) “leaves the door open” for
psychological evidence to be adduced to assist the tribunal of fact (Coyle & Thomson, 2014, p.478). And, if applied, could be used to “dispel myths and misapprehensions that otherwise might play an unacknowledged role in juror deliberations” (Freckleton, 1993, p. 129). Bagaric and McConvill (2005) support the principles articulated in Aytugrul v. The Queen declaring:

Rules of evidence [need] to be based not on the intuition of lawyers, but rather on the knowledge of experts in the area. We should be listening to behavioural scientists, not more lawyers – they are the ones that got us into this hole in the first place (p.12).

Although judicial directions to the triers of fact have a long history regarding witness credibility (i.e. Alexander v The Queen, 1981; Domican v R, 1992; Longman v The Queen, 1989), there is simply no evidence that such directions have any meaningful effect on improving the accuracy of the decisions of jurors. In fact, the efficacy of judicial directions on juror decision-making has been the subject of lengthy empirical debate (Blankenship, Luginbuhl, Cullen, & Redick, 1997; Baguley, McKimmie & Masser, 2017; English & Sales, 1997; Luginbuhl, 1992; Ogloff, Nadjovski-Terziovski, Spivak & Clough, 2011; Ogloff & Rose, 2005; Rose & Ogloff, 2001; Severance & Loftus, 1982; Spivak, Ogloff, & Clough, 2018). This limitation has prompted interest in providing education to jury members in the form of expert evidence and/or judicial evidentiary instructions (e.g. Coyle & Thomson, 2014; Goodman-Delahunt, Cossins & O’Brien, 2011; Ribbers & Hennebery, 2018). Although there have been numerous attempts to train venire jurors in techniques to assess deception, these models reflect little resemblance to the constraints within which a trial operates. This will be demonstrated in the next section where a brief overview of the key approaches to detecting deception and a critique of training in detecting deception is provided.
2.3.3 Techniques in Detecting Deception

The array of methods aimed to detect deception includes both specialised and non-specialised techniques (Vrij, 2008). Collateral information aside, these methods encompass three key approaches: psychophysiological measures, verbal-based methods and nonverbal methods.

**Psychophysiological measures**: These measures include the polygraph, ‘lie-detector test’, and the evolution of functional neuroimaging. At its core, these measures assume different psychophysiological responses when a person is lying and when a person is telling the truth (Raskin & Kircher, 2014). Physiological measures such as the polygraph presume heightened biological arousal in response to the threat of detection (US National Academy of Science 2003). This premise is informed by the theory of emotional arousal that posits increases in stress, fear and guilt indicate markers of deception, such as an increase in breathing rate, perspiration, heart rate and blood pressure (Ekman & Friesen, 1969; Greely & Illes, 2007; Zuckerman et al., 1981).

Neuropsychological perspectives however, are based on the assumption that deceptive behaviour is associated with changes in executive functioning (Vrij & Gannis, 2014). Theoretically, these changes are a consequence of deception, as a process, placing increased taxing loads on cognitive functioning (Buller & Burgoon, 1996; DePaulo et al., 2003; Vrij, 2008; Vrij & Gannis, 2014). Although still in their infancy, functional neuroimaging techniques aim to discriminate deceptive from genuine behaviour by assessing differences in brain activity (Granhang, Vrij, & Verschuere, 2015; Johnson, 2014).

Nevertheless, psychophysiological techniques for detecting deception have been found unsuitable for use in courts both in Australia and internationally (Bizzi, et al, 2009; Greely & Illes, 2007; Hahm, et al, 2009). Specifically, polygraph evidence is not considered of evidentiary utility in Australian courts. In *R v. Murray* (1981) the District Court of New South Wales deemed polygraph evidence inadmissible on the grounds that, among other things, the technique is “devoid of any scientific basis … with no probative value” and that it is “within the province of the jury to
determine on the facts, presented to them by witnesses who perceived them by the exercise of their physical senses”. This position was cemented by the recent West Australian Supreme Court authority, *Mallard v The Queen* (2003), a renowned case in Australian law that in 2005 saw the High Court of Australia overturn the conviction against *Mallard* acknowledging that the inadmissibility of polygraph evidence had been robustly dealt with by the Supreme Court of Western Australia (Freckelton, 2004; *Mallard v The Queen*, 2005). Apart from these decisions, the application of specialised technology for the detection of deception cannot as a matter of law and practice be adopted in Australian courts. Apart from the practical absurdity of using psychophysiological methods in a courtroom setting, technology or opinion evidence that talks to the identification of guilt usurps the role of the jury. These very issues were expressed in the High Court of Australia:

> The trial judge was right to approach with caution any attempt to call evidence which could have the effect of usurping the jury's function in reaching their ultimate conclusion as to whether a witness was telling the truth or not. … the assessment of credibility is a matter for the tribunal of fact (here the jury). In the present state of science it may not be usurped by technology. Nor may it be assumed by witnesses, including expert witnesses, offering their opinion on the accuracy, consistency and believability of the testimony in question, however derived (*Farrell v. The Queen*, 1998 at 27).

**Verbal-based techniques:** Verbal lie-detection tools have become increasing popular in assessing the veracity of statements (for example, witness statements and police interview recordings). Verbal lie-detection methods are based on the underlying assumption that it is more demanding to make a convincing impression when lying than when telling the truth (Buller & Burgoon, 1996; Driskell & Driskell, 2019). The concepts of cognitive load and impression management whereby the process of inventing a story that is believable – such as a false alibi – requires greater thought and awareness than being able to rely on actual experience (Zuckerman,
Depaulo & Rosenthal, 1981; and see Levine et al., 2018; Vrij, et al., 2017; Vrij et al., 2008; Vrij & Gannis, 2014; and Vrij et al., 2011 for detailed explanations of the cognitive load approach pertaining to verbal deception detection). The most common approaches used in detecting verbal deception that have been established in the literature include Statement Validity Analysis (SVA), Reality Monitoring (RM), and Scientific Content Analysis (SCAN). According to Vrij (2008, 2015), the validity and utility of verbal methods do not satisfy the legal thresholds of admissibility articulated in Daubert v. Merrell Dow Pharmaceuticals (1993). Although the guidelines set out in Daubert are not formalised as a standard of proof in Australia, the principles are expressed in Australian case law: Makita Pty Ltd v. Sprowles (2001). Nonetheless, Vrij’s (2008, 2015) blanket rejection of verbal methods overplays the argument, at least against Criteria Based Content Analysis, a component of SVA that has shown some considerable promise in accurately detecting deception.

**Nonverbal methods:** Unlike psychophysiological methods and the analysis of verbal content, nonverbal behaviours can be discerned without the aid of equipment and/or specialised techniques (Vrij et al., 2010). Nonverbal indicia of deception encompass vocal, para-verbal cues and physical behaviours (Sporer & Schwandt, 2006; Vrij, 2008). Ultimately, nonverbal indicia are responsible for guiding overall impressions of demeanour (Bogaard & Meijer, 2018; McKimmie et al., 2014). When nonverbal and verbal content are conflicting, that is a person’s behaviour contradicts what they are saying or vice a versa, assessments of truthfulness tend to rely on nonverbal indicia (Levine & McCormack, 2014; Masip & Herrero, 2015b; Vrij et al., 2010). Indeed, the identification and interpretation of behavioural indicators are highly subjective and are, typically, mislead by false beliefs about how liars behave (Strömwall et al., 2004; Vrij et al., 2010). As noted, common misconceptions among jurors include beliefs that shifting one’s posture, averting one’s gaze, scratching, touching or self-grooming, repeated blinking, relaying longer stories and increased rates of speaking are correlated with deception but, these have been shown to
be unreliable indicators of deception (Akehurst et al., 1996; DePaulo et al., 2003; Duke, 2007; Mann & Vrij, 2006; Mann, Vrij, & Bull, 2004; Strömwall et al., 2004; Vrij et al., 2010).

The literature reports that the most reliable nonverbal indicia of deception include: change in vocal pitch with liars found to speak in a higher pitched voice than truth-tellers; reduced body movements and gesticulations (‘illustrators’) with liars making fewer hand, arm, and feet movements than truth-tellers; speech latency with longer pauses characteristic of deceit compared to truthfulness; and responses containing fewer details, less contextual information, indirect answers and fewer spontaneous corrections common to deceptive statements over truthful ones (DePaulo et al., 2003; Hartwig & Bond, 2011, 2014; Levine, 2018; Sporer & Shwandt, 2006, 2007; Strömwall et al., 2004; Villar et al., 2013; Vrij, 2008; Vrij et al., 2019; Zuckerman et al., 1981). Nevertheless, where behavioural markers of deception have been shown to be valid, the average observed effect size for detection accuracy is small (\(d = .25\)) (Levine, 2018; Vrij et al., 2019).

According to Hartwig and Bond (2014) relying on cues in isolation to one another may provide some explanation. That is, overall displays of behaviour consist of multiple cues and therefore “signals of deception are manifested in constellations rather than single cues” (Hartwig & Bond, 2014, p. 667). In their recent meta-analysis, Hartwig and Bond (2014) examined this idea to see whether multi-cue analyses of deception could improve detection. They found that deception can be better predicted from multiple cues than a single cue, but that the strongest cue contributes to the bulk of detection while other cues contribute less (Hartwig & Bond, 2014).

Concerning the validity of behavioural correlates of deception, key criticisms amount to issues of specificity and variability (Levine, 2018). Further, the need to establish individual baseline performance to discriminate usual from unusual behaviour undermines the generalisability of behavioural indicia of lying (Porter, 2007, 2010). Although the research concludes some indicia are more reliable than others, the fact remains that techniques relying on verbal, paraverbal or
behavioural indicators do not allow for intra and inter-variability with people acting differently in different settings and at different times (Levine, 2018; Levine, 2015; Vrij, 2008; Vrij et al., 2019).

2.3.4 Training in Detecting Deception: Does practice make perfect?

Traditionally, researchers maintained that training enhanced the ability to detect lies. However, the results were unimpressive with less than 5% gains observed in overall detection accuracy following training (Frank & Feeley, 2003; Hauch, et al, 2016; Sternglanz et al., 2019; Vrij, 2008; Vrij, et al., 2019; Vrij, Mann, & Leal, 2013). These discouraging findings inspired research interest to refine training models to improve performance (Dunbar et al., 2018; Levine, Feeley, McCornack, et al., 2005; Miller, et al., in press). In a short space of time, the body of research on training in detecting deception has grown considerably with recent meta-analyses having quantified small to medium effects on the overall impact of training on detection accuracy (Driskell, 2012; Frank & Feeley, 2003; Hauch et al., 2016).

The largest, most recent review, was conducted by Hauch and colleagues (2016) who demonstrated a small, pooled effect, of 30 studies (published and unpublished) that implemented control-group designs. Conversely, Driskell (2012) and Frank and Feely’s (2003) work identified moderate training effects for detection accuracy across 16 and 11 studies respectively. Albeit that the moderate effect sizes may be inflated, both meta-analyses neglected a number of applicable studies that were available at the time of analysis and the same control groups were used repeatedly by both authors resulting in problematic dependent effect sizes (Hauch et al., 2016). Notwithstanding these methodological and statistical issues, some of the variability in results may be explained by differences in research design, training content delivered (e.g. Levine et al., 2005), and training approaches used.

With respect to experimental design, to satisfy inclusion into Hauch et al.’s (2016) meta-analysis the study had to include a control-group. However, the vast majority of these only involved a post-test design thereby assuming equality between experimental and control groups.
before training (Shadish, Cook, & Campbell, 2002). In the absence of a pre-test – and only six studies included this – baseline comparisons between groups could not be determined. This approach is common to deception research and is further confounded when quasi-experimental designs are employed. By way of example, consider research conducted by Shaw, Porter, and ten Brinke (2013) who trained forensic specialists (experimental group) in nonverbal deception indicia only to use non-forensic specialists – namely university students – as their control. As such, participant allocation to the experimental and control groups were not random but rather, pre-determined by the differing participant characteristics. And further, differences in performance could not be attributed to the intervention because group differences prior to training were not considered.

**Training content:** Hauch and colleagues (2016) concluded that verbal-based training content resulted in larger effects than relying on combined, multichannel methods (i.e. nonverbal and/or paraverbal). As a consequence, they advocate for the use of verbal methods in developing future models in detecting deception to the exclusion of combined approaches (Hauch et al., 2016). While the results for verbal techniques are promising, the use of statement-based analysis in court ignores the practicalities of a trial. It is simply impractical to expect jurors to further analyse testimony against the presence/absence of 19 content-based criteria. These criteria, developed by Steller and Köhnken (1989), are referred to as Criteria-Based Content Analysis (CBCA). Originally, the tool was designed to assess the testimonies of victims of child sexual abuse but has since been empirically validated for use with adult populations (Amadoa, Arcea, Farina & Vilarino, 2016). The technique is one of the most extensively used for evaluating the veracity of a witness’s testimony, and has been admitted as evidence in a number of courts world-wide (Vrij, 2008). The tool evaluates written statements against 19 factors that are grouped in two: cognitive factors (criteria 1 through 13), and motivational factors (criteria 14 to 18). The technique contends that memories of truthful experiences differ in content and quality to fictitious accounts (Amadoa et al., 2016), because lying is more cognitively demanding than telling the truth (Vrij, 2008). Despite its
efficacy, asking jurors to critically evaluate 19 criteria would not be feasible in a trial setting. Not only would jurors require specialist training to use the technique, but such training would likely be a financial burden to the courts, and logistically prohibitive.

A common deficiency among studies investigating the use of combined methods in detecting deception is the use of unreliable techniques and unreliable indicia. This oversight is problematic for a number of reasons. Most notably, these studies are cited as the body of evidence against the use of nonverbal indicia of deception. The seminal study cited to corroborate claims that nonverbal indicia results in poor deception detection was that conducted by Kassin and Fong (1999). Their findings indicated a significant negative relationship between training and detection accuracy. However, little commentary has dealt with the fact that the training approach adopted was the popular but unreliable Reid Technique (Docan-Morgan, 2007; Mann et al., 2004; Masip & Herrero, 2015a). Kassin and Fong (1999) reflect that “the Reid technique may not be effective – and indeed, may be counterproductive – as a method of distinguishing truth and deception” (p. 512).

It is not surprising that nonverbal indicators are criticised as poor discriminators between truth-tellers and liars when invalid cues are tested in training models (i.e. Geiselman, 2013). Driskell (2012) sums this up best: “Clearly if training is designed to draw attention to behavioral patterns reflecting deception, then it is important that those cues trained are in fact reliable indicators of deception” (p.725). This position tends to be overlooked when claims are made that nonverbal approaches are unreliable in detecting deception. Then, there are studies that claim the provision of ‘empirically validated’ indicia resulting in positive benefits but don’t report them (i.e. Matsumoto, Skinner, & Frank, 2012; Porter, Juodis, ten Brinke, Klein, & Wilson, 2010; Shaw et al., 2013). This means that the validity of training content used cannot be compared or evaluated.

**Type of training:** Feedback is a core component of training. Theoretically, feedback serves to positively or negatively reinforce behaviour (Driskell, 2012). Training models that
incorporate the provision of information, the opportunity for practice and feedback are considered superior to information-only interventions by some (Driskell, 2012; Levine et al., 2005; Porter et al., 2010; Powell, 2008; Powell, Fisher, Wright, Brewer, & Williams, 2005; Powell, Wright, & Clark, 2010), but not by others (Hauch et al., 2016). Both bogus feedback and accurate feedback have been investigated. Hauch and her colleagues (2016) suggest that neither type of feedback has any effect on detection accuracy and Levine and his colleagues (2005) found that although training enhances awareness, it does not provide the depth of knowledge required. In the field of forensic interviewing, practice and feedback are argued as necessary agents of change for effective training (Powell, 2008; Powell et al., 2005). Indeed, it is concluded that the absence of practice and individualised feedback may contribute to poor performance despite training (Powell et al., 2005). Although skill acquisition is advanced through practice and feedback the provision for “critical feedback… in cognitively challenging skills” for triers of fact is not possible (Powell et al., 2010, p. 218; Powell, 2008). In an actual trial the only scientific information that can aid the triers of fact in assessing witness demeanour can be adduced by expert evidence. But as a matter of law and practice, this evidence does not and cannot involve giving jurors practice or feedback in distinguishing liars from truth tellers. Rather, the provisions of expert advice or directions to jurors are the best mechanisms available within the constraints of a jury trial.

In sum, there have been numerous attempts to train jurors in techniques to assess deception. However, these training models have no place with respect to jury decision-making during a trial. Despite their promise, models employing verbal-based techniques, practice and feedback approaches to training and pre-post control-group experimental conditions cannot be utilised within the constraints of a trial. This is not to say that potential jurors couldn’t be trained in detecting deception before being empanelled albeit that logistics and cost implications suggest it is a remote prospect. Of course, expert evidence could be introduced (i.e. reliable verbal techniques) but it would be vulnerable to rejection by the court on the grounds that such knowledge lies within the common sense of the jury.
Introducing expert evidence on indicators of deception to aid jurors in their assessment of witness credibility was investigated in a study conducted by Coyle and Thomson (2014). Using a pre and post-test mixed design the authors investigated juror opinions on what factors (indicia of deception) were considered important when determining the truthfulness of a witness in court. Mock jurors were asked to score 31 items according to how much each factor would help them in determining the truthfulness of a witness in court. Each of the items tested has been well-established in the research to be diagnostic/not diagnostic of deception. Mock jurors completed the questionnaire pre-intervention to gather their baseline knowledge on nonverbal indicia of deception and again post-intervention to measure any change from baseline. The intervention took the form of expert advice on reliable indicia of deception. Coyle and Thomson (2014) observed significant differences in the behavioural indicators relied upon by jurors following education by way of expert witness testimony and judicial direction. With the provision of reliable information on nonverbal indicia, jurors shifted to relying on more valid behavioural indicators in assessing witness credibility. Based on their results Coyle and Thomson (2014) suggest that well-entrenched misconceptions about deception indicia can be corrected by way of judicial direction or the proffering of expert testimony. They also argue that providing advice on the right techniques can significantly shift perceptions on what cues discriminate lying from truth-telling.

It is important to consider the strength of the evidence reported before arriving at any overall conclusions about the efficacy of promoting change in juror behaviour due to education on detection of deception. This conclusion requires a careful assessment of the quality of the research design before it can be assumed that any shifts in indicia are a direct result of expert evidence on deception indicia. Indeed, a strength of the design adopted by Coyle and Thomson (2014) was the random allocation of participants to the experimental groups to help to overcome problems associated with selection bias (Haynes, 1992), and the use of a pre/post-test design as recommended by Shadish and colleagues (2002) to facilitate baseline comparisons.
With respect to methodological deficiencies, Coyle and Thomson (2014) identify two aspects of their design that would benefit from revision. Both modifications relate to experimental condition one whereby participants received a written judicial instruction on deception indicia. The first deals with removing the opportunity for participants to copy answers when completing the follow-up questionnaire. The authors note that although participants assigned to this condition were instructed to turn over the written direction after completing the baseline questionnaire they had the opportunity to crib when filling out the questionnaire the second time. It is possible, they suggest, that some participants cribbed by checking their answers against the written judicial direction. The second shortcoming relates to the presentation modality used: they suggest that providing a written judicial direction may have had more impact than had the direction been provided orally. They refer to related research into the efficacy of judicial directions that show written directions are twice as effective as oral ones (e.g., Thomas, 2010). In the light of Coyle and Thomson’s (2014) unexpected results that judicial directions had the same effect as expert evidence in shifting factors relied upon by jurors, the analysis of presentation modality is worthy of further investigation.

The evidence presented by Coyle and Thomson (2014) to indicate that advising jurors on indicia of deception can influence the factors relied upon were mean differences from pre to post-test. However, in doing so they only reported shifts on items indicative of deception (12 of a possible 31 items). Although the mean differences shifted significantly on these items, any changes observed on items not indicative of deception were not reported and are therefore unknown: Coyle and Thomson (2014) predicted increased scores on items indicative of deception, but no prediction was made for items not indicative of deception. Although changes on items not indicative of deception would not disturb the unique factor pattern identified by Coyle and Thomson (2014) this analysis may have provided a measure of validity for the intervention. For the items missing from analysis, it could be argued that any change observed from pre-test to post-test would be smaller with jurors relying less on unreliable indicia post-intervention. Indeed, it
may be reasoned that an overall decrease in the weight attributed to non-diagnostic items by jurors may be possible and, if found, would provide strong evidence that advising jurors on indicia of deception aids in determinations of credit. Had Coyle and Thomson (2014) reported on the nature of change for items not indicative of deception it may have provided greater insight into the value of presenting evidence on reliable/unreliable indicia of deception. As it stands shifts to relying on valid indicia by jurors cannot be excluded as a generalised effect associated with presenting evidence and the long-term impact of providing advice to jurors on detecting deception remains unanswered.

Indeed, if DDT is accurate – and the jury is still out on this – then knowledge gleaned about indicia of deception would not be applied by decision-makers in their final determinations of witness credibility (Coyle & Thomson, 2014). If this holds true (as the theory predicts), then introducing expert evidence on detecting deception would be of little utility in an adversarial trial: by the time evidence on indicia of deception could be presented by the defence it would be too late – the triers of fact would have already formulated their impression and there would be no value in adducing evidence. And, if this is found to be correct it would call into question the entire foundation on which jury trials operate. These issues are the focus of the current research.
Chapter 3. Rationale and Contribution of Research

The dialogue on science's replicability crisis contends that systematic reforms in scientific practice are paramount on the basis that science's hallmark function of self-correction is undermined by deviations from the truth-seeking principle (Ioannidis, 2012; Pashler & Harris, 2012). Ioannidis (2012) purports that timely, objective replication mechanisms are critical for upholding high levels of scientific credibility. His summation describes this predicament best:

In the absence of replication efforts, one is left with unconfirmed (genuine) discoveries and unchallenged fallacies, where in many areas of psychological science, perpetuated and unchallenged fallacies may compromise the majority of the circulating evidence (p. 645).

The 'outpouring of concern' that psychological science is facing a 'replicability crisis' suggests that science won’t necessarily correct itself unless direct (not just conceptual) replications are performed (Pashler & Harris, 2012). In a series of papers published by Perspectives on Psychological Science, the prevalence of replication studies performed following 'positive' findings (true & false positives) across science is said to be very low, in the order of 1% - 5% (Ioannidis, 2012). Until recently, replication rates in psychology were unknown with Makel and colleagues (2012) reporting the first systematic examination of prevalence rates for replications in psychology. Their work examined papers on replications in psychology published since 1900 in the 100 most-cited psychology journals. From this, they found 1.6% of psychological research included the term ‘replication’ but only 68% of those represented actual replications. This results in an overall replication rate of 1.07% \((n = 342)\) (Makel, Plucker, & Hegarty, 2012). Nevertheless, this replication rate accounts for all types of replications.

The distinction between a direct replication and a conceptual replication is an important one. Direct replications refer to studies that repeat all relevant elements (i.e. conditions and method) of an original study and are adopted to test the validity and robustness of new findings
INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION

(Koole & Lakens, 2012; Schmidt, 2009). On the other hand, conceptual replications employ different procedures to the original study to test the generality of findings purported thereby adding to our theoretical understanding of an effect (Nosek, 2014; Pashler & Harris, 2012). However unlike direct replications, conceptual replications cannot disconfirm a study’s original findings (Koole & Lakens, 2012; Pashler & Harris, 2012). Put simply, identifying false positives is unlikely without direct replications (Nosek & Lakens, 2014). Despite this, conceptual replications account for the majority (81.9%) of all replications published in psychological science with only 14% of replications recognised as direct attempts (Makel et al., 2012). Accordingly the vast majority of psychological science comprises discoveries that have no record of published replication efforts (Ioannidis, 2012).

The value of direct replication has been at the forefront of debate with claims that behavioural science research is compromised by a culture of enquiry that promotes inflated effects and false-positive errors (Fiedler, Kutzner, & Krueger, 2012; Ioannidis, 2012). Moreover, the attention attributed to false positives overshadows a subtler dilemma: the prospect of false negatives – neglecting to discover and confirm correct hypotheses (Coyle, Campbell, Thomson, & Woskett, 2016; Fiedler et al., 2012). Given the demonstrated bias against publication of negative findings, science can’t self-correct against false negatives (Ferguson & Heene, 2012; Francis, 2012; Giner-Sorolla, 2012). The omission of negative findings means that when a non-effect is concluded (incorrectly) the hypothesis is cast aside and the opportunity to generate alternative hypotheses is lost (Fiedler et al., 2012). Fielder and his colleagues (2012) assert: “…overlooking alternative hypothesis [false negatives] renders all scrutiny intended to reduce α errors worthless. False positives can be corrected through replication whereas false negatives are less likely to be detected, corrected, and understood” (p. 662).

The contribution of the current research is twofold. First, the aim to reproduce results previously observed by Coyle and Thomson (2014) will add to dearth of replications in
psychological science that have not only been at the forefront of debate in academic circles but that have more recently gained momentum and increasing attention in popular science forums (Luntz, 2015; Wells, 2015). Indeed, the reproducibility attempted by Study one is only a small component of the overall body of work accounting for a third of the entire experimental design. Hence, the second focus of the current research is to expand the design and extend the interventions and reporting of measures under investigation. In doing so the current research aims to resolve the putative methodological and reporting deficiencies identified within the Coyle and Thomson (2014) design, but also make a significant original contribution to what we know about the effects of delay on jurors’ determination of evidence presented at trial.

The contribution of the current program of research is significant, from both a theoretical and applied perspective. If it is demonstrated that providing information in the form of expert evidence and/or judicial directions to jurors causes a shift towards employing valid nonverbal indicia of deception, and that this shift is maintained over time, then aspects of DDT (that initial judgements are formed quickly and resistant to change) may be questioned. With this would come profound implications for the presentation of expert evidence relating to the issue of witness credibility that may ultimately, have the potential to re-shape jury-trial proceedings to make safer decisions when witness credibility is at issue. In the words of Granhag and Strömwall (2004) “professionals who need to assess veracity on a regular basis will, in the long run, make more correct veracity assessments and fewer mistakes if armed with scientifically based knowledge” (p.328).
Chapter 4. Aims and Hypotheses

4.1 Aims

In an attempt to replicate previous research conducted by Coyle and Thomson (2014), Study one aims to:

1) Confirm/disconfirm that expert evidence and/or judicial direction on diagnostics of deception can influence mock jurors to rely on more valid indicia of deception when determining the truthfulness of a witness in court;

2) Extend the replication by examining whether or not the impact of advice can correct jurors’ misconceptions of unreliable indicators of deception.

4.2 Hypotheses

For Study one, two directional hypotheses are proposed. It is hypothesised that:

1) Jurors’ perceptions of reliable nonverbal cues of deception will improve with the provision of advice; and

2) Jurors’ perceptions of unreliable nonverbal cues (i.e. misconceptions) will be corrected with advice.

To test these two hypotheses, increases will be expected in mean scores on perceptions of reliable cues (i.e. the reliable subscale), and decreases in mean scores on perceptions of unreliable cues (i.e. the unreliable subscale) are anticipated.

Other variables that may be associated with influencing jurors’ opinions of nonverbal cues of deception are also explored. These include presentation modality of advice and participant confidence. Previous research suggests both written evidentiary judicial instructions (Coyle &
Thomson, 2014; Thomas, 2010) and oral evidentiary directions (Goodman-Delahunty et al., 2011; Ribbers & Henneberg, 2018) are effective in shaping jurors’ perceptions regarding scientific knowledge. Accordingly, directional hypotheses (as noted above) are adopted for the modalities of advice under investigation. Conversely, perceived confidence, although established in related research as a factor confounding detection deception accuracy (i.e. the “overconfidence effect”; DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997), and erroneously influencing jurors’ assessment of eyewitness identification evidence (Charman, Wells, & Joy, 2011; Deffenbacher, 1980; Wells, Lindsay, & Ferguson, 1979), it has not been studied with respect to advising jurors on deception detection. Accordingly, no a-priori hypotheses for the effect of advice on perceived confidence is proposed.
Chapter 5. Method

The research methodology for Study one and Study two were similar in design. To avoid duplication and repetition of content, a general method is described with differences in procedure highlighted between the two.

5.1 Research Design

The current research used a parallel group, pre-and post-test design. Two separate mixed models were adopted allowing for within-group and between-group comparisons. Both repeated measures employed a three-way $5 \times 2 \times 2$ model where the first factor, a between-subject factors was source of advice (with five levels: expert witness, written judicial direction, oral judicial direction, expert/judge combined, or no advice); the second factor, a within-subject factor, was type of deception cue (with two levels: reliable cues and unreliable cues); and the third factor, a within-subject factor was time-to-assessment (pre-and post-test).

The two studies differed with respect to the time-to-assessment factor. In Study one there was no delay between the time participants received advice and their assessment of indicia, thereby representing ‘immediate advice’. By contrast, Study two investigated the impact of a one-week delay between the time respondents received advice and their assessment of indicia thus resulting in ‘delayed advice’. The delay-to-assessment therefore was between the time of receiving the advice and participants’ ratings of nonverbal cues.

The dependent variable was the mean importance of deception cues rated by mock jurors in determining the truthfulness of a witness in court. Changes in importance ratings (e.g. mean difference) from pre- to post-test were investigated on both reliable and unreliable cues across both studies.
5.2 Participants

The demographic particulars of Study one and Study two were remarkably alike. Although separate studies, for simplicity, the demographics of both studies are reported together. Participants were recruited as mock jurors. Throughout the thesis the terms jurors and mock jurors are used throughout. For ease of reference it is noted that hereon in when the results of the thesis are reported and interpreted, jurors refers to mock jurors.

A sample of 173 mock jurors was recruited for Study one and a pool of 177 mock jurors was collected for Study two. The mean age of participants in Study one was 38.96 years (SD = 9.79) and 39.56 years (SD = 10.67) in Study two. Ages ranged from 18 to 70 years across both studies with females comprising the sample majority: 91.3% in Study one, and 76.7% in Study two. Level of education completed was predominantly tertiary with 78.6% (Study one) and 79.3% (Study two) having completed, or partially completed a tertiary qualification. High school or TAFE was completed, or partially completed by 21.4% of participants in Study one and 20.7% in Study two. The likeness between samples are illustrated in Table 1 (Study one) and Table 2 (Study two).

Participants were predominantly recruited within Victoria for Study one (93.1%, n = 161) and Study two (85.5%, n = 100). Other States and Territories represented in Study one were: Queensland (1.73%, n = 3), South Australia (1.73%, n = 3), New South Wales (1.16%, n = 2), and the Australian Capital Territory (n = 1). For Study two, the remainder of participants resided in: South Australia (6.0%, n = 7), New South Wales (5.1%, n = 6), and Queensland (3.4%, n = 4).

A small number of participants from Study one (3.5%, n = 6) had previous training in detecting deception compared to no participants from Study two. Half of the participants who received training in Study one were in the control (n = 3) with the remaining three participants distributed across expert witness, oral judicial direction and combined expert/judge conditions.
The representativeness of the current samples with known statistics of Australian jurors was sought. Given the prevalence of Victorians represented in both samples, demographic profiles of Victorian jurors were preferred, however, no published data was found. Nonetheless, statistics obtained from Juries Victoria (2019) revealed that the mean age of jurors in Victoria was 46.99 years with a sex split of 44.89% female and 55.11% male. In a report published by the Victorian Law Reform Commission (2014), Victorian juries are considered to approximate the Victorian community in relation to age. The mean age of the current samples was relatively consistent with the profile of Victorian jurors. Not surprisingly, the breakdown of gender was not representative of members serving on juries in Victoria. The level of education of Victorian jurors is unknown and so no conclusions can be inferred regarding the representativeness of the largely tertiary qualified samples obtained.
Table 1

*Study 1 Demographics.*

<table>
<thead>
<tr>
<th>Study Type</th>
<th>n</th>
<th>Age</th>
<th>Sex</th>
<th>Education</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Male % (n)</td>
<td>Female % (n)</td>
</tr>
<tr>
<td>Expert Witness</td>
<td>20</td>
<td>38.21</td>
<td>8.95</td>
<td>10.0 (2)</td>
<td>90.0 (18)</td>
</tr>
<tr>
<td>Written Judicial Direction</td>
<td>32</td>
<td>41.06</td>
<td>12.45</td>
<td>6.3 (2)</td>
<td>93.8 (30)</td>
</tr>
<tr>
<td>Oral Judicial Direction</td>
<td>34</td>
<td>37.68</td>
<td>10.54</td>
<td>17.6 (6)</td>
<td>82.4 (28)</td>
</tr>
<tr>
<td>Expert/Judge Combined</td>
<td>22</td>
<td>40.00</td>
<td>8.63</td>
<td>13.6 (3)</td>
<td>86.4 (19)</td>
</tr>
<tr>
<td>Control (no advice)</td>
<td>65</td>
<td>37.57</td>
<td>8.31</td>
<td>1.6 (1)</td>
<td>98.4 (63)</td>
</tr>
<tr>
<td>Total</td>
<td>173</td>
<td>38.96</td>
<td>9.79</td>
<td>8.1 (14)</td>
<td>91.3 (158)</td>
</tr>
</tbody>
</table>
### Table 2: Study 2 Demographics

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Sex</th>
<th>Education</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Expert Witness</td>
<td>19</td>
<td>40.89</td>
<td>10.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>31.6 (6)</td>
<td>68.4 (13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>26.3 (5)</td>
<td>73.7 (14)</td>
<td></td>
</tr>
<tr>
<td>Written Judicial Direction</td>
<td>31</td>
<td>36.90</td>
<td>9.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.7 (3)</td>
<td>90.3 (28)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.7 (3)</td>
<td>90.3 (28)</td>
<td></td>
</tr>
<tr>
<td>Oral Judicial Direction</td>
<td>29</td>
<td>42.48</td>
<td>12.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17.2 (5)</td>
<td>82.8 (24)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24.1 (7)</td>
<td>75.9 (22)</td>
<td></td>
</tr>
<tr>
<td>Expert/Judge Combined</td>
<td>18</td>
<td>42.56</td>
<td>10.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>22.2 (4)</td>
<td>77.8 (14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.3 (6)</td>
<td>66.7 (12)</td>
<td></td>
</tr>
<tr>
<td>Control (no advice)</td>
<td>20</td>
<td>35.26</td>
<td>8.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>47.4 (9)</td>
<td>52.6 (10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.8 (3)</td>
<td>84.2 (16)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>117</td>
<td>39.56</td>
<td>10.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.3 (27)</td>
<td>76.7 (89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.7 (24)</td>
<td>79.3 (92)</td>
<td></td>
</tr>
</tbody>
</table>
5.3 Materials

The four conditions of advice investigated by the current research (expert witness, written judicial direction, oral judicial direction, and expert witness and judicial direction combined) required three stimuli: two Digital Versatile Discs (DVDs) and one transcript. The stimulus for each advice condition is detailed in turn and, for reference, copies are attached in Appendix C.

5.3.1 Expert Witness DVD

The expert witness DVD was an existing resource that was used in the original study by Coyle and Thomson (2014). It presented evidence by an expert witness on indicia that have been shown to be reliable, to one extent or another, in detecting deception. The expert witness was a full professor in forensic psychology with more than three decades of experience in proffering expert testimony. The duration of the expert’s evidence lasted 10 minutes and 54 seconds and was filmed in a mock courtroom.

5.3.2 Written Judicial Direction

The written transcript portrayed a judge’s direction to a jury that cited research on reliable and unreliable indicators of deception. The transcript was an existing resource developed by Coyle and Thomson (2014).

5.3.3 Oral Judicial Direction DVD

The second DVD was developed for the current research. It presented a male in his early sixties who portrayed a presiding judge directing a jury on reliable and unreliable cues of deception. The duration of the direction lasted 1 minute and 54 seconds and mirrored the content transcribed in the aforementioned transcript. The presiding judge wore formal legal regalia including ceremonial robe and wig consistent with that of a Supreme Court Chief Justice. The setting resembled a bench trial with a non-descript backdrop and judge’s gavel in the foreground.
5.3.4 Combined Expert Witness and Judicial Direction DVD

The final stimulus was adapted for the current research by merging the evidence of the expert witness and judicial direction into one continuous film clip. Each form of advice was introduced by way of a title page stating the role of the presenter such as “Sworn evidence of Emeritus Professor Donald Thomson. Given by Videolink”. The duration of advice in the combined intervention lasted 12 minutes and 54 seconds. The combination of expert witness and judicial direction was chosen over other permutations on the basis that advice by an expert and a judge is the most likely of situations to occur in a court of law.

5.4 Measures

Two measures were assessed: number of reliable cues of deception and number of unreliable cues of deception that mock jurors rated important in evaluating the truthfulness of a witness in court. The two measures investigated nonverbal deception cues that have been validated by empirical research and have been well established in the literature as diagnostic (reliable cues) and not diagnostic (unreliable cues) of deception.

Participants were instructed to rate each cue according to how important they perceived it to be in determining the truthfulness of a witness giving evidence in court. Both measures should provide evidence of the effect of advice to which participants were exposed (increases in the number of reliable cues and decreases in number of unreliable cues). This expanded on the single measure reported by Coyle and Thomson (2014) who only reported reliable items indicative of deception.

Both measures were continuous variables. Reliable cues reflected items indicative of deception and were represented by 12 of the 31 factors. These were item numbers: 2, 3, 6, 11, 14, 17, 19, 21, 25, 28, 29, and 30. The reliable scale was created by calculating the summed mean score across the 12 items indicative of deception.
Unreliable cues reflected the 19 items not diagnostic of deception. These were questionnaire items: 1, 4, 5, 7, 8, 9, 10, 12, 13, 15, 16, 18, 20, 22, 23, 24, 26, 27, and 31. The unreliable scale was developed by calculating the summed mean score across the 19 factors not indicative of deception. The 31 items rated in the questionnaire are listed in Table 3, with reliable cues highlighted grey.

Table 3
Nonverbal Deception Cues Assessed by Participants

Participants were asked to rate the importance of each factor relied upon when determining the truthfulness of a witness giving evidence in court.

| 1. Confidence | 17. Responses that lack logical structure |
| 2. Pauses | 18. Pressed lips |
| 3. High-pitched voice | 19. Repetition of certain words and phrases |
| 4. Fidgeting | 20. Blinking |
| 5. Restless behaviour | 21. Responses that contain very few details |
| 6. Implausible responses | 22. Tenseness |
| 7. Postural changes | 23. Self-grooming |
| 8. Speech disturbances/fillers (i.e., ‘ah’s’ and ‘umms’) | 24. Nervous behaviour |
| 10. Self-manipulators (i.e., hand movements that involve relieving a bodily need like scratching) | 26. Gaze aversion |
| 11. Short responses | 27. Crossed arms |
| 12. Mumbling | 28. Illustrators (i.e. hand movements that accompany and illustrate the content of the speaker’s responses) |
| 13. Covering the mouth | 29. Delayed responses |
| 15. Stuttering | 31. Responses that contain a lot of detail |
| 16. Interrupting the questionnaire | |

Note: The reliable indicia that comprise the reliable subscale are highlighted in grey and the factors that remain white reflect the unreliable subscale.

5.5 Procedure

5.5.1 Participant Recruitment

Participants were recruited via a snowballing, crowdsourcing technique. Appropriate research portals available through the University and online, social media channels (i.e. Facebook) were implemented to recruit a convenient sample of the general public. Research flyers were distributed throughout cafes, restaurants, movie theatres, domestic airports, sporting clubs, gaming
venues, supermarket and community noticeboards and retail stores. The research flyers detailed the nature of the study and listed the research website (https://www.howtospatiaiar.info). Because the study was online, crowdsourcing facilitated distribution of the survey to large and varied audiences.

Participants completed the survey using the online platform Qualtrics. In line with jury-duty requirements, participants had to satisfy three inclusion criteria. Mock jurors had to be: 18 years or older, fluent in English and an Australian citizen. Ethics approval for the two studies was obtained through Deakin University’s Human Research Ethics Committee, the details of which were reported for participants in the Plain Language Statement (PLS).

Adopting Faul and colleagues (2007) a-priori calculation using G*Power 3 analysis, for one-tailed, matched pairs analyses with $\alpha$ set at .05 and power $(1 - \beta)$ was determined at .80, $n$ was required to be 164 to detect a small effect ($d = .20$). To detect a medium effect size ($d = .50$) an $n$ of 28 was needed and to detect a large effect size ($d = .80$) a sample of $n = 12$ was recommended. For more abundant caution a sample of 150 participants was sought for each study allowing for 30 participants per cell.

5.5.2 Questionnaire

The survey comprised three questions. Question one listed 31 items indicative of truth-telling/deception. Participants were asked to rate each item as to its importance in determining whether someone was truthful when giving evidence in court. Respondents had to score each factor on a six-point rating scale. The presentation order of factors was randomised to enhance internal validity and to protect against participants identifying those items that were or were not indicative of deception. Question one was developed by Coyle and Thomson (2014) and replicated for the current project. Two new questions were added to the current research extending the instrument to a 33-item survey. The details of question two and three are outlined below.
Question two asked participants to rank five of 31 items that they considered the most important in determining the truthfulness of a witness. Respondents had to score each factor in order of importance from most important (1) through to least important (5). The third question asked participants to rate on a five-point Likert scale from strongly agree to strongly disagree how confident they were in their perceived ability to determine if someone was lying or not. Please see Appendix B for a copy of the questionnaire.

Before participants commenced the survey, they were provided with a PLS that outlined the background, aim and research procedures. The PLS explained to participants that they were recruited as members of a mock jury. The PLS stipulated that participation was both voluntary and anonymous and that participants were free to withdraw at any point during the survey. Anonymous participation removed any preconceived expectation or obligation to participate by those known to the researchers (i.e. social media networks). Access to a computer or mobile device and Internet connection was required to complete the online survey. Participants who consented to partake in the study selected the ‘continue’ button at the end of the PLS and were re-directed to a new page to start the survey.

The purpose of the questionnaire was twofold. First, it was to obtain mock jurors’ baseline perceptions of cues considered indicative of deception. This was achieved by participants completing the first, pre-intervention questionnaire (Q1). The second aim was to measure shifts in perceptions on cues considered important after jurors received advice on deception detection. This was achieved by participants completing a second, post-intervention questionnaire (Q2).

The pre-and post-intervention questionnaires completed by participants were one in the same: both Q1 and Q2 were identical in structure, content and instructions. This allowed for changes to be assessed from pre-to post-test on the two measures: reliable cues and unreliable cues. To investigate the effect of advice on reliable and unreliable cues considered important by mock jurors the mean scores of importance were compared from pre-to post-test where pre-test scores
indicated baseline opinion and any shift in scores at post-test reflected change in opinion after exposure to advice.

5.5.3 Advice on Indicators of Deception: The experimental conditions

After completing the pre-intervention questionnaire participants were randomly assigned to one of five experimental conditions where advice on reliable and unreliable indicators of deception was presented by a: (1) expert witness; (2) written judicial direction; (3) oral judicial direction; (4) expert/judge combined; or (5) no advice at all for those in the control.

Advice conditions one (expert witness) and two (judicial transcript) replicated the experimental groups investigated by Coyle and Thomson (2014). Conditions three, four and five extended upon the original design by comparing written and verbal judicial advice as well as the impact that repetitive information (expert witness and judicial advice combined) may have on influencing legal decision-making (Kovera, Borgida, Gresham, Gray, & Regan, 1997).

Participants exposed to no advice on detecting deception formed the control group. Participants in the control group completed Q1 and then Q2 immediately thereafter under the guise that the study was testing the strength of their opinions about deceptive behaviour. Participants in the control group took approximately 10 minutes to complete both questionnaires.

Aside from the control (no advice), participants were randomly assigned to an advice condition and were instructed that, as members of a jury, they would be receive advice on reliable and unreliable indicators of deception. For the three conditions in which advice was presented via DVD, participants were instructed to press play to watch the advice. Subjects exposed to the transcript of judicial advice were required to read the direction carefully before proceeding. In-built timers were incorporated into each advice group to prevent subjects from skipping ahead of the intervention before it finished. It took participants in the advice groups approximately 20 minutes to complete the studies.
5.6 Data Analytic Strategy

Data screening and recoding practices carried out on the research sample are described below. An overview of the statistical analyses conducted and the internal consistency of the two measures (reliable and unreliable cues) are also defined.

5.6.1 Data Screening and Cleaning

A total of 636 pre-intervention questionnaires were commenced. Participants were randomly allocated across the five advice conditions: $n = 128$ to expert witness; $n = 121$ to judicial transcript; $n = 128$ to judicial direction; $n = 122$ to expert and judicial advice combined; and $n = 137$ to the controls. Of these, 290 participants completed the post-intervention questionnaire (Study one and two combined), resulting in an overall response rate of 45.6%. Figure 1 depicts the progress of all surveys undertaken with the corresponding response rates of both Study one and Study two reported.

\[\text{Figure 1. Flow and Response Rates of Participants in Study 1 and Study 2}\]

Of the 338 participants randomly allocated to Study one, $n = 165$ surveys were invalid due to incomplete responses and so were removed from analysis. This resulted in a 51.2% response rate for Study one. In Study two, 298 participants completed the pre-intervention questionnaire.
After the seven-day delay, 164 participants (55%) returned to the post-intervention questionnaire and 134 did not. After data cleaning, a total of 117 valid questionnaires were retained for analysis achieving a 39.3% response rate and a 60.7% attrition rate for Study two.

Post-hoc power analyses were determine for the different effect sizes found across studies. Power was calculated at group-level using G*Power (Faul, Erdfelder, Lang & Buchner, 2007). Whole of sample power for the respective studies was not appropriate because analyses were conducted at group level. A summary of the post-hoc calculations follow.

For Study one, post-hoc analyses for one-tailed, within-group comparisons revealed power ($1 - \beta = .80$) ranging from .21 to .40 for the small effects obtained. Power was calculated at .61 to .82 for the medium effects and the large effect was obtained with power at .91. For one-tailed, between-group analyses post-hoc analyses indicate that power ranged from .39 to .42 for small effects and .83 to .88 for the large effects found.

Post-hoc analyses for Study two showed that for the one-tailed, within-group comparisons the small effects achieved minimal power (.18 to .20). For the medium effects, power was determined at .63 and for the large effect power was calculated at .87. The one-tailed, between-group analyses achieved power between .34 to .36 for the small effects and .57 to .58 for the medium effects obtained.

5.6.2 New Variables

Control group: The introduction of a control group extended the current research design from the Coyle and Thomson (2014) study. The control reported mock juror baseline perceptions of items indicative of deception as rated on the first survey (Q1) and again on the second survey (Q2), with no intervening advice. As such the control added a fifth level of advice; no advice. Controls were recruited for both Studies one and two. Controls were sought with the aim to identify
potential retest effects as a consequence of simple questionnaire re-administration (Scharfen, 2018).

Mean importance: To facilitate comparisons of the mean for deception cues rated important, each reliable item of deception (12 items) was transformed into a summed mean score of importance, generating a new variable, reliable cues. The same transformation was performed on unreliable items (19 cues) creating a new scale, unreliable cues. The two new variables enabled the effect of advice, and delayed advice, to be summarised by a total mean difference for each participant.

5.6.3 Statistical Techniques

The research data was not continuous, but rather, ordinal in nature with categorical variables only collected for demographic information such as age, sex, education, and training.

Five statistical techniques were undertaken to identify differences between the effect of immediate advice and delayed advice on cues of deception perceived important by mock jurors. The statistical analyses included crosstab frequency distributions, Wilcoxon-matched pair signed-rank tests, independent paired-samples t-tests, Mann-Whitney two sample tests and Kruskal-Wallis for comparing multiple independent samples.

Descriptive statistics were used to identify proportional differences and allowed for direct group comparisons. Frequencies also provided for a comprehensive summary of demographic items and deception cues rated most important by jurors.

Wilcoxon-matched pair signed-rank tests (Wilcoxon tests) were conducted for two reasons. First, non-parametric methods were necessitated by non-normal distributions identified across both samples (Bradley, 1978, 1982; Hubbard, 1978; Singer, 1979; Sprent & Smeeton, 1989; please see results for further details). Second, the analyses reflected within-group repeated measures. Accordingly, Wilcoxon tests were selected to analyse differences on the dependent variable (mean
importance) across the three independent variables (time-to-assessment and cue-type), and was achieved by assessing whether the summed mean score of importance for cue type deviated from pre-intervention (Q1) to post-intervention (Q2).

A comment regarding normality: it has been well established that provided the violations of normality are not excessive, parametric techniques can be adopted (Bradley, 1978, 1982; Hubbard, 1978; Singer, 1979; Sprent & Smeeton, 1989). However, in the case of the current thesis, departures from normality occurred in a number of situations, not least sample size rendering parametric inferences problematic (Sprent & Smeeton, 1989). Further, the sample sizes were inconsistent between conditions. The alternative, to the transform data, was deemed inappropriate because the distributions included both skewed and kurtotic spreads that ranged from moderate to severe across all conditions (Tabachnick & Fidell, 2007). As such, comparing the results to the original study of Coyle and Thomson (2014) would have made for a more difficult, and potentially “misleading, interpretation” (Leech & Onwuegbuzie, 2002, p.6; Sprent & Smeeton, 1989). Furthermore, given the small samples and non-normal distributions, the Wilcoxon test adopted has been established to be the most powerful test for those distributions (Imam, Mohammad & Abanyam, 2014). Further to the above, it is pertinent to emphasis that the research data herein was not continuous but ordinal in nature necessitating the use of non-parametric techniques despite the violated assumptions of normality.

Wilcoxon tests were reported in the Coyle and Thomson (2014) study. Replicating the Coyle and Thomson study calls for consistency with respect to the analytic technique adopted (assumptions met herein), as well as for the statistical metrics reported to ensure potential differences in effects of advice on juror opinions are not a consequence of differing analyses. Accordingly, mean scores and mean differences are reported in the current studies. That said, it is acknowledged that non-parametric data, ordinarily, necessitates the reporting of medians and interquartile ranges in place of means and standard deviations (Field, 2013). However, as
explained, to maintain a direct rather than conceptual replication (Nosek & Lakens, 2014), means were adopted and standard deviations were reported sparingly. The use of paired-samples t-tests were limited to one analysis per study where the parametric equivalent to Wilcoxon tests was appropriate.

Mann-Whitney two sample tests (Mann-Whitney test) and Kruskal-Wallis tests were required for non-parametric, between-group analyses for independent samples. Mann-Whitney tests compared mean differences between the different levels of advice and the control group of no advice for both reliable and unreliable cue types. The Kruskal-Wallis tests was adopted when comparisons across multiple independent groups was required.

One-tailed tests were fixed when directional differences were anticipated, and two tailed tests were set for exploratory analyses. For all analyses, significant or not, effect sizes were calculated using Cohen’s $d$ (Cohen, 1962; 1988). Significance was determined at $\alpha = .05$ with marginal significance noted where $\alpha > .05$ but $< .07$. 
Chapter 6. Study One: Results

For the purpose of the following analyses, the results of Study one are structured as follows. First, the effect of advice on mock jurors’ perceptions of cues considered indicative of deception are reported. The within-group analyses are stated initially with the results of the replication of Coyle and Thomson’s (2014) study examined to determine whether advice can influence perceptions of reliable cues. Following the replication of Coyle and Thomson (2014), the impact of advice on mock jurors’ perceptions of unreliable cues is investigated. The within-group analyses conclude by comparing shifts in perception between reliable and unreliable cues following advice. Second, to establish whether source of advice influenced mock juror perceptions over and above no advice, between group analyses were conducted comparing shifts in opinion across each level of advice with the control. Third, the influence of advice on individual indicators of deception ranked most important to least important are assessed. Lastly, the chapter concludes with an analysis of mock juror confidence to ascertain whether receipt of advice influences perceived confidence in detecting deception.

Normality: The distribution of data for each condition of advice was non-normal as indicated by skewness and kurtosis scores and histogram graphs (Field, 2013; Hubbard, 1978; Sprent & Smeeton, 1989; Tabachnick & Fidell, 2007). Inspection of boxplots revealed outliers. Extreme values were determined using the rule: Interquartile Range x 1.5. At group level, 18 extreme values were removed from analysis of reliable cues (expert witness n = 3; written direction n = 3; oral direction n = 5; expert and judicial direction combined n = 1; control n = 6). For unreliable cues 27 extreme cases were excluded (expert witness n = 4; written direction n = 3; oral direction n = 6; expert and judicial direction combined n = 2; control n =12). As such, the cell counts and mean difference calculations differ across analyses of reliable and unreliable subscales.

Internal consistency: The consistency of the two scales, reliable cues and unreliable cues, was assessed using Cronbach’s (1951) alpha. Both the reliable deception scale (12 items α = .81)
and unreliable deception scale (19 items $\alpha = .91$) were found to have good and excellent levels of internal consistency, respectively.


6.1.1 Reliable Deception Cues

Baseline comparisons of mean ratings of reliable cues were undertaken to identify whether perceptions differed between groups before exposure to instructions. A Kruskal-Wallis test showed that differences in baseline scores of reliable cues between for all types of advice was non-significant, $\chi^2 (4) = 3.52, p = .48$ (two-tailed), with a mean rank of reliable cues of 69.74 for expert witness, 75.33 for written direction, 91.21 for oral direction, 79.21 for expert/judge combined and 74.77 for no advice.

The following section of results comprises the replication. To be consistent with the analysis performed by Coyle and Thomson (2014), within-group analyses (Wilcoxon tests) are stated first for reliable cues of deception for each condition of advice.

**Expert witness:** The mean score for reliable cues of deception increased from $M = 3.18$ at baseline to $M = 3.54$ after exposure to advice by expert witness. A Wilcoxon test revealed that the mean increase from Q1 to Q2 (M difference = .36), was statistically significant: $z = 2.05, p = .02$, $d = .58$, one-tailed.

**Written judicial direction:** When mock jurors were exposed to a written direction, the mean importance of reliable cues shifted from $M = 3.35$ at pre-advice to $M = 3.36$ post-advice and the negligible increase (M difference = .01) was non-significant, $z = .90, p = .45$, $d = .02$, one-tailed.
**Oral judicial direction:** When advice was presented as an oral direction, mean scores shifted from $M = 3.54$ at baseline to $M = 3.55$ post advice. The mean shift ($M$ difference = .01) on reliable cues was non-significant: $z = .27, p = .39, d = .02$, one-tailed.

**Expert and judge combined:** The mean score of reliable cues rated by mock jurors was $M = 3.37$ prior to advice and $M = 3.63$ post advice. A Wilcoxon test indicated that the mean increase on reliable cues ($M$ difference = .26) following exposure to both expert testimony and oral direction was statistically significant, $z = 2.30, p = .01, d = .60$, one-tailed.

**No advice:** The control group of no advice revealed that the mean rating of reliable cues was $M = 3.35$ on Q1 and $M = 3.29$ on Q2. The shift in mean scores from Q1 to Q2 ($M$ difference = -.06) was non-significant $z = -1.09, p = .14, d = .18$, one-tailed.

For the two types of advice that significantly enhanced jurors’ perceptions of reliable indicia (expert witness and expert/judge), the effectiveness of each source of advice was compared to determine whether one was more successful than the other in improving opinion. The results of a Mann-Whitney test revealed that the enhancement in perceptions of reliable indicia achieved with expert witness ($M$ difference = .38) and expert/judge ($M$ difference = .26), were non-significant, $u = -3.53, p = .73, d = .18$, two-tailed.

The pattern of mean scores for each source of advice at baseline, and again post advice, are shown in Figure 2. The mean scores ratings for reliable cues for the two time-to-assessment periods, (Q1: baseline and Q2: post-advice) and associated parameters (e.g. means of the mean differences denoted ‘M Diff’ between Q1 and Q2) are reported in Table 4 for each level of advice.
Figure 2. Mean Scores for Reliable Cues by Advice and Time-to-Assessment. Depicted here are the mean ratings for the reliable subscale, measured at baseline (Q1) and again post advice (Q2) for each source of advice including the control: no advice.
The following section, 6.1.2, mirrors the same format as above however, the analysis offers an extension to the replication by examining the results for the subscale of unreliable cues as an additional, separate independent variable.

### 6.1.2 Unreliable Deception Cues

A Kruskal-Wallis independent samples test confirms that despite differing starting points observed between the conditions of advice, the differences between baseline scores for unreliable cues were non-significant, $\chi^2 (4) = 2.92, p=.57$ (two-tailed), with a mean rank of unreliable cues of 72.00 for expert witness, 78.62 for written direction, 82.43 for oral direction, 65.28 for expert/judge and 69.54 for no advice.

**Expert witness:** There was a statistically significant mean decrease on unreliable cues (M difference = -.34) after mock jurors received advice by expert witness (Q2: M = 2.92), compared to before they received advice (Q1: M = 3.26), $z = -2.69, p = .01, d = .85$, one-tailed.

**Written judicial direction:** A mean decrease on unreliable cues (M difference = -.17) was found after mock jurors were presented with written judicial directions (Q2: M = 3.21) compared to before they received advice (Q1: M = 3.39). The impact of written directions on shifting mock jurors’ perceptions of unreliable cues was significant, $z = -2.24, p = .01, d =.45$, one-tailed.

**Oral judicial direction:** The mean score of unreliable cues at baseline was M = 3.44 and M = 3.52 after mock jurors received an oral direction. Although exposure to oral judicial directions saw an increase in perceptions of unreliable cues, the shift in perceptions was non-significant (M difference = .08), $z = .87, p = .19, d = .18$, one-tailed.

**Expert and judge combined:** Mock jurors rated unreliable cues with a mean score of M = 3.16 on Q1 and M = 3.01 on Q2. The impact of combined advice via expert testimony and judicial direction (M difference = -0.16) was marginally significant, $z = -1.53, p = .06, d = .32$, one-tailed.
No advice: The control group of no advice indicated that the mean rating of unreliable cues on Q1 and Q2 was $M = 3.26$ and $M = 3.18$, respectively. A Wilcoxon test found that the mean decrease ($M$ difference $= -.08$) from baseline was significant; $z = -2.74$, $p = .01$, $d = .37$, one-tailed.

To establish whether one type of advice was more effective than another in correcting misconceptions, the significant shifts in opinion achieved with expert evidence (the largest mean increase) was compared to shifts attained with written judicial instructions and expert/judge. The findings indicate that all three forms of advice were equally effective in correcting jurors’ beliefs regarding non-diagnostic cues of deception: the difference in improvement achieved by expert witness ($M$ difference $= -.34$) and written direction ($M$ difference $= -.17$), were non-significant, $u =1.40$, $p = .16$, $d = .41$, two-tailed; and the differences in opinion obtained by the expert ($M$ difference $= -.34$) did not significantly differ from those of the expert/judge ($M$ difference $= -.16$), $u =-1.21$, $p = .23$, $d = .42$, two-tailed.

The mean scores of importance for unreliable cues by source of advice are illustrated in Figure 3 and reported in Table 4 along with the mean scores for reliable cues already outlined.
Figure 3. Mean Scores for Unreliable Cues by Source of Advice and Time-to-Assessment.
Table 4
*Shifts in Mean Scores of Mock Juror Perceptions from Pre-Test (Q1) to Post-Test (Q2) by Cue-Type and Source of Advice*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>Reliable Cues</th>
<th></th>
<th></th>
<th>Unreliable Cues</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td></td>
<td>N</td>
<td>M</td>
<td>M diff</td>
</tr>
<tr>
<td>Expert Witness</td>
<td>17</td>
<td>3.18</td>
<td>3.54</td>
<td>.36</td>
<td>.62</td>
<td>.02*</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td>29</td>
<td>3.35</td>
<td>3.36</td>
<td>.01</td>
<td>.50</td>
<td>.45</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td>29</td>
<td>3.54</td>
<td>3.55</td>
<td>.01</td>
<td>.51</td>
<td>.39</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td>21</td>
<td>3.37</td>
<td>3.63</td>
<td>.26</td>
<td>.43</td>
<td>.01**</td>
</tr>
<tr>
<td>No Advice (control)</td>
<td>59</td>
<td>3.35</td>
<td>3.29</td>
<td>-.06</td>
<td>.33</td>
<td>.14</td>
</tr>
</tbody>
</table>

* p^- calculated using Wilcoxon-matched-pair signed-rank tests (one-tailed); * Significance at .05; **Significance at .01.

Note: Also calculated are the means of each mean difference (denoted ‘M diff’), the SD for the Wilcoxon test and effect sizes reported using Cohen’s d.
6.1.3 **Reliable Cues versus Unreliable Cues**

Baseline differences in mock jurors’ perceptions between reliable cues and unreliable cues were investigated. Comparisons between cue-type at baseline were undertaken to establish whether or not jurors perceived one type of cue to be any more indicative of deception, prior to advice, and whether perceptions deviated with advice and, if so, to what extent. Although no directional hypothesis were specified between opinions of reliable and unreliable, hypotheses one and two predicted that both cue-types would improve with advice. Consequently, increases in mean scores were expected for reliable cues and corresponding decreases in means scores were anticipated for unreliable cues. Hence the directional nature of the existing predictions, one-tailed tests were adopted. Although no directional hypotheses were made for baseline perceptions, one-tailed tests were considered appropriate in an attempt to avoid skewing interpretations of differences between cue-types from pre-to post-delay.

To perform this analysis cases were paired where outliers were absent on both the reliable and unreliable subscale. Accordingly, the cell sizes differ from those adopted for the independent analyses (i.e. 6.1.1 and 6.1.2). The results are reported by advice type, and are also depicted in Figures 4 and 5 along with Table 5.

**Expert witness:** At baseline, the difference in mean scores between cue-type for those randomly assigned to expert witness were non-significant (reliable cues: M = 3.20; unreliable cues: M = 3.26; M difference = .06), z = -.62, p = .27, d = .45, one-tailed. After receiving expert evidence, mock jurors rated reliable cues (M = 3.51) significantly higher than unreliable cues (M = 2.92; M difference = .59), z = 2.33, p = .01, d = .75, one-tailed.

**Written judicial direction:** Prior to advice, differences in mean scores between reliable (M = 3.40) and unreliable cues (M = 3.36) were non-significant for participants assigned to written directions, (M difference = .04), z = .58, p = .28, d = .10, one-tailed. Following receipt of written
instructions however, reliable cues (M = 3.37) were rated significantly higher than unreliable cues (M = 3.17; M difference = .20); z = 1.92, p = .03, d = .52, one-tailed.

**Oral judicial direction:** Differences between reliable (M = 3.56) and unreliable mean scores (M = 3.51) at baseline were non-significant for mock jurors randomly assigned to advice by way of oral judicial instructions: M difference = .05; z = .88, p = .44, d = .15, one-tailed. After advice, the difference in mock juror’s ratings of reliable cues (M = 3.58) and unreliable cues (M = 3.56) remained non-significant: M difference = .02; z = .63, p = .26, d = .04, one-tailed.

**Expert and judge combined:** At baseline, mock jurors rated reliable cues (M = 3.32) significantly greater than unreliable cues (M = 3.10; M difference = .22, z = 2.62, p = .001, d = .72, one-tailed. Post advice, reliable cues (M = 3.59) remained significantly higher than unreliable cues (M = 2.95; M difference = .64), z = 2.80, p = .01, d = 1.46, one-tailed.

Figure 4. Mock Juror Perceptions at Baseline (Q1) and Post-Advice (Q2) by Cue-Type and Source of Advice.
**No advice:** Mock jurors in the control group rated reliable (M = 3.32) and unreliable cues (M = 3.26) equally at baseline (M difference = .06) as confirmed by a Wilcoxon test, \( z = 1.12, p = .13, d = .15, \) one-tailed. However, on follow-up the difference between reliable cues (M = 3.27) and unreliable cues (M = 3.18) was significant: M difference = .09; \( z = 1.83, p = .03, d = .21, \) one-tailed. The pattern of findings for no instruction are displayed in Figure 5 below and the statistical parameters are summarised in Table 5.

*Figure 5.* Total Mean Scores for the Reliable and Unreliable Subscales at Baseline (Q1) and Post-Advice (Q2) for Mock Jurors Who Received No Advice (Control).
Table 5

*Post-Test Mean Scores of Mock Juror Perceptions of Reliable and Unreliable Cues by Source of Advice*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>N</th>
<th>M</th>
<th>M</th>
<th>M diff</th>
<th>SD</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Witness</td>
<td>16</td>
<td>3.51</td>
<td>2.92</td>
<td>.59</td>
<td>.79</td>
<td>.01**</td>
<td>.75</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td>27</td>
<td>3.37</td>
<td>3.17</td>
<td>.20</td>
<td>.47</td>
<td>.03*</td>
<td>.52</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td>25</td>
<td>3.58</td>
<td>3.56</td>
<td>.02</td>
<td>.63</td>
<td>.26</td>
<td>.04</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td>19</td>
<td>3.59</td>
<td>2.95</td>
<td>.64</td>
<td>.79</td>
<td>.01**</td>
<td>1.46</td>
</tr>
<tr>
<td>No Advice (control)</td>
<td>53</td>
<td>3.27</td>
<td>3.18</td>
<td>.09</td>
<td>.35</td>
<td>.03*</td>
<td>.21</td>
</tr>
</tbody>
</table>

\(p^\) calculated using Wilcoxon-matched-pair signed-rank tests; * Significance at .05; **Significance at .01.
6.2 Does Advice Impact Jurors’ Perceptions of Cues? Assessing Differences Between Advice and No Advice: Analyses between-groups.

Shifts in mock jurors’ perceptions of reliable and unreliable cues were compared to the control. This analysis differs from the previous analysis in that it makes a direct comparison of the effect of receiving advice (i.e. expert witness) against the effect of receiving no advice. Comparisons with the control were undertaken to determine whether any potential retest effects such as the simple repetition of the questionnaire (Scharfen et al., 2018) had any confounding effect on the results obtained. The analyses aimed to establish whether or not advice impacted shifts in opinion over and above no advice. The results for reliable and unreliable cues are reported concurrently and the summary statistics for the between-group analyses can be found in Table 6.

**Expert witness:** A between-group analysis revealed that the mean difference for reliable cues from Q1 to Q2 (M difference = .36) was significantly different when mock jurors were exposed to expert witness advice compared to mock jurors not exposed to advice (M difference = -.06), \( u = 2.45, p = .01, d = 1.02, \) one-tailed. For unreliable cues, a Mann-Whitney test revealed that when compared against the control, the decrease in mean importance was significantly different for jurors who received expert witness advice (M difference = -.34) compared to those who received no advice (M difference = -.08), \( u = -2.80, p = .001, d = .96, \) one-tailed.

**Written judicial direction:** A comparison against the control also indicated that reliable cues were not rated significantly differently by jurors exposed to written directions (M difference = .01) compared to jurors who received no advice (M difference = -.06), \( u = .47, p = .32, d = .18, \) one-tailed. A Mann-Whitney test showed that the difference in mean scores on unreliable cues was non-significant between jurors exposed to written instruction (M difference = -.17) and mock jurors exposed to no advice (M difference = -.08), \( u = -.91, p = .18, d = .30, \) one-tailed.
**Oral judicial direction:** Between-group analysis indicated that the mean difference on reliable cues for jurors exposed to advice by verbal direction (M difference = .01) compared to those who received no advice (M difference = -.06) was non-significant, $u = .29, p = .15, d = .18$, one-tailed. A between-group comparison also suggests that unreliable cues were not rated significantly different by mock jurors who received advice by oral direction (M difference = .08) compared to those who received no advice (M difference = -.08), $u = 1.15, p = .13, d = .50$, one-tailed.

**Expert and judge combined:** Compared against no advice, the change in mean scores on reliable cues from Q1 to Q2 (M difference = .26) for jurors exposed to combined expert/judge advice was statistically different from mock jurors who received no advice (M difference = -.06), $u = 2.78, p = .01, d = .89$, one-tailed. A Mann-Whitney test found that the mean change score on unreliable cues did not differ significantly between mock jurors who received expert and judicial advice (M difference = -.16) from mock jurors who received no advice (M difference = -.08), $u = -1.19, p = .12, d = .26$, one-tailed.
Table 6

*Between-Group Comparisons of Mean Differences in Mock Juror Perceptions by Cue-Type and Source of Advice*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>Reliable Cues</th>
<th>Unreliable Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M diff</td>
</tr>
<tr>
<td>Expert Witness</td>
<td>17</td>
<td>.36</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td>29</td>
<td>.01</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td>29</td>
<td>.01</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td>21</td>
<td>.26</td>
</tr>
<tr>
<td>Control (no advice)</td>
<td>59</td>
<td>-.06</td>
</tr>
</tbody>
</table>

p^ calculated using Mann-Whitney two-sample tests; **Significance at .01; ***Significance at .001  
Note: The mean difference (M diff) reflects the change score from pre-to post advice (Q2 - Q1) for the reliable and unreliable subscales and was calculated to run between-group comparisons against the control.
6.3 What Combination of Nonverbal Cues are Considered Most Important?”

The following analyses are presented as frequencies and percentages to best reflect the cues ranked most important by participants. The top five cues rated most important by mock jurors are reported first followed by the five cues most commonly endorsed. The deception cues considered most important before and after advice are described and compared. For the cues most commonly preferred, total percentages exceed 100% with multiple rankings possible for each item.

6.3.1 Top Five Most Important Cues

Out of a possible 31 items, 19 cues were identified among the most important in determining truthfulness (see Figure 6 below). Prior to advice, the five cues ranked most important by mock jurors in determining truthfulness was: confidence (22.9%); responses that contain a lot of detail (16.2%); implausible responses (10.6%); spontaneous corrections (10.5%), and gaze aversion (7.6%). In total, these cues accounted for more than two thirds (67.6%, \( n = 71 \)) of all items rated.

Almost half of the cues ranked in the top five (46.6%, \( n = 49 \)) were unreliable items of deception. Both the first (confidence) and second most important cues (responses that contain a lot of detail) accounted for the majority of the unreliable items rated in the top five: 39%, \( n = 41 \) of the 49 responses. Two reliable cues (spontaneous corrections and implausible responses) were ranked in the five most important cues accounting for 21% of all responses.

After exposure to advice, mock jurors maintained that confidence (21.9%) and responses that contain a lot of detail (18.1%) were the two most important cues in detecting deception. These two unreliable cues accounted for 40% of all responses after advice was received—almost an identical proportion of responses made prior to advice (39.1%). The remaining three cues rated most important by mock jurors included reliable items: high pitched voice (10.5%); implausible responses (7.6%); and responses that contain very few details (7.6%).
Percentage change analysis revealed that the differences from pre-to post advice for the five most important cues (i.e. confidence, responses that are implausible, responses that contain a lot of detail, spontaneous corrections, and gaze aversion) were among the smallest. For these items, percentage changes ranged from: -4% (n = 1) for confidence to -75% (n = 6) for gaze aversion. The greatest percentage change was observed for high pitch voice that increased 450% (n = 9) followed by: responses that contain very few details that increased 300% (n = 6); body/hand/finger/leg/foot movements that increased 300% (n = 3, respectively); pauses that increased 150% (n = 3); and short responses and delayed responses that both increased 100% (n = 1, respectively). The percentage increase in responses that contain very few details saw the cue shift into the top five most important cues post advice. Figure 7 shows the percentage change rankings for all items ranked most important by mock jurors.
Figure 6. Nonverbal Cues Ranked Most Important Before and After Advice. The grey dash line delineates the top five most important cues from all cues ranked most important. Presentation order reflects cues ranked in order of importance at baseline.
Figure 7. Percentage Change in Nonverbal Cues Ranked Most Important After Advice. The percentage change is delineated for the two measures: reliable cues are depicted in blue and unreliable cues are charted in orange. Items are presented in order of percentage change from largest through smallest and positive through negative.
6.3.2 Most Commonly Preferred Cues

The five most commonly endorsed items of deception rated important by mock jurors were, in order of preference: confidence, (51.4%, \( n = 54 \)); responses that contain a lot of detail (50.5%, \( n = 53 \)); spontaneous corrections (30.5%, \( n = 32 \)); responses that lack logical structure (26.7%, \( n = 28 \)); and repetition of certain words or phrases (25.7%, \( n = 27 \)). The top three most preferred items were unreliable cues. The fourth and fifth most preferred items were reliable indicia of deception, however, the number of times these items were ranked was approximately half that of the first and second most preferred (unreliable) cues.

Despite advice, mock jurors maintained that ‘responses that contain a lot of detail’ (48%, \( n = 50 \)) and confidence (47%, \( n = 49 \)) were the two most preferred cues. Albeit, the order of preference swapped (see Figure 8). Illustrators (36%, \( n = 38 \)); responses that contain very few details (30%, \( n = 32 \)); and implausible responses (30%, \( n = 31 \)) comprised the three remaining cues most frequently endorsed – all of which were newly cited, reliable cues.

After exposure to advice, the largest percentage change was found for: high pitch voice that increased 222% (\( n = 20 \)); responses that contain very few details that increased 113% (\( n = 17 \)); and illustrators that increased 81% (\( n = 17 \)). All three items were reliable cues of deception. Of all the unreliable cues, only self-grooming (75%, \( n = 3 \)) demonstrated a percentage increase post advice.

Of all cues, the largest percentage decrease following receipt of advice was found for the unreliable cue, gaze aversion (-80%, \( n = -12 \)). A percentage decrease was found across three reliable cues: repetition of certain words decreased by 44% (\( n = -3 \)); ambivalent responses dropped 38% (\( n = -16 \)); and spontaneous corrections decreased 6% (\( n = -2 \)). Percentage changes for all items most frequently endorsed by mock jurors are shown in Figure 9.
Figure 8. Most Commonly Endorsed Nonverbal Cues Before and After Advice. Items are presented in order of most preferred to least preferred measured from baseline.
Figure 9. Percentage Change in Nonverbal Cues Most Commonly Endorsed After Advice. The percentage change is delineated for the two measures: reliable cues depicted in blue and unreliable cues charted in orange. Items are presented in order of percentage change from largest through smallest, and positive through negative.
6.4 Does Exposure to Advice Influence Juror Confidence in Detecting Deception?

Perceived confidence was measured at baseline and again following exposure to advice. No directional hypotheses were predicated and so two tailed tests were adopted. An equivalent control was not obtained for confidence so within-group analyses reflect shifts from baseline following exposure to advice. The sample used to test the effect of advice on perceptions of nonverbal cues (at 6.1.1 and 6.1.2) was adopted to examine the effect of advice on jurors’ perceived confidence. Preliminary data cleaning indicated normally distributed data, and so, parametric paired-samples t tests were conducted.

**Expert witness:** Mock jurors’ level of confidence shifted from \( M = 3.25 \) (SD = 1.06) at baseline to \( M = 3.19 \) (SD = .98) following exposure advice by expert witness. The mean decrease in confidence (M difference = -.06, SE = .17) for those who received expert testimony was non-significant, \( t (15) = -.368, p = .72, d = .09 \), two-tailed.

**Written judicial direction:** Confidence mean scores for mock jurors who received written judicial instructions shifted from \( M = 3.62 \) (SD = .64) to \( M = 3.58 \) (SD = .70) following a delay but, the mean decrease was minimal: M difference = -.04, SE = .10; \( t (25) = -.372, p = .71, d = .08 \), two-tailed.

**Oral judicial direction:** For those who received verbal directions, confidence increased by a mean difference of .08 (SE = .15) from \( M = 3.38 \) (SD = .88) at baseline to \( M = 3.46 \) (SD = .78) post-advice but the increase was non-significant, \( t (23) = .569, p = .58, d = .11 \), two-tailed.

**Expert and judge combined:** Levels of perceived confidence shifted for jurors exposed to advice by expert/judge but the mean increase was not significant (Q1: \( M = 3.18 \), SD = 1.19; Q2: \( M = 3.24 \), SD = .97; M difference = .06; SE = .20), \( t (16) = .29, p = .77, d = .07 \), two-tailed.

In sum, shifts in perceived confidence for all types of advice were non-significant.
Chapter 7. Discussion

The current study aimed to establish whether or not evidentiary instructions on indicia of deception by way of expert evidence, and/or judicial directions, may influence juror perceptions of nonverbal diagnostics of deception. Specifically, the current study sought to replicate previous research to confirm/disconfirm whether instructions enhance juror perceptions of reliable cues of deception when determining the truthfulness/deception of a witness. The second objective expanded upon the replication and investigated whether the provision of advice could correct misconceptions of behavioural indicia of deception by improving jurors’ opinions of unreliable cues.

The format of the discussion mirrors that set out in the preceding chapter of results (chapter six). First, the findings of the within-group analyses (i.e. comparisons from baseline) are discussed. The findings detail the effect of type of advice on perceptions of indicia, independent of one another, to establish whether source of advice is relevant in influencing perceptions. A comparative analysis follows with discrepancies in opinion between reliable and unreliable cues investigated. Second, the impact of instructions on perceptions are compared to no instructions (i.e. the between-group comparisons). Third, a review of the combination of nonverbal cues considered most important by jurors in detecting deception is outlined. The current chapter concludes with a summary of the effect of instructions on juror confidence in detecting deception.

7.1 Does Advice Impact Jurors’ Perceptions of Nonverbal Cues of Deception?

7.1.1 Assessing Shifts from Baseline: Analyses within-groups.

The within-group analyses made comparisons from baseline, meaning that participants functioned as their own controls as was determined by baseline scores taken at time one, prior to instructions. As such, each group was independent of one another and analysed accordingly.
Hypothesis 1: Jurors’ perceptions of reliable cues of deception will improve with advice

The current study provides mixed support for hypothesis one, that predicted jurors’ perceptions of reliable indicia of deception would improve with advice. The current study found that when instructions were presented by an expert witness (moderate effect, $d = .58$), or when instructions were delivered via an expert witness together with a judicial direction (expert/judge; moderate effect, $d = .60$), jurors’ perceptions of reliable cues significantly improved. The increase from baseline for advice by expert witness and expert/judge collectively, suggests that providing information on behavioural indicators of deception to jurors may enhance opinions of cues that reliably predict deception with jurors perceiving reliable cues to be more indicative of deception following instruction. Accordingly, these provisional findings run counter to some elements of DDT (Porter & ten Brinke, 2009), that predicts initial perceptions of trustworthiness are so pervasive that they do not change when faced with conflicting evidence. While it is acknowledged that this thesis did not ask mock jurors to make a judgement regarding a witness’ testimony nor judge the credibility of a witness, the findings of this thesis indicate, at least preliminarily, that perceptions of cues perceived indicative of trustworthiness can be manipulated when jurors are confronted with conflicting evidence.

The current study also investigated whether or not the effectiveness of advice on shifting perceptions differed according to the type of instruction given. That is, the study examined whether instructions introduced as expert testimony or by an expert/judge were more effective in improving opinion. The findings revealed that the impact of advice did not differ significantly by type of instruction: neither the expert witness nor the expert/judge was more or less effective in enhancing jurors’ opinions of reliable indicia of deception. Although perceptions of reliable cues may be influenced with instruction, this finding suggests that the source of advice did not materially influence the degree of change in perceptions achieved.
The remaining sources of advice (written judicial directions, verbal judicial directions, and no advice), did not effect jurors’ perceptions of reliable cues. For both judicial directives, opinions remained stable despite advice. In fact, the shifts in perceptions of reliable indicia for both written and oral directions were negligible ($d = .02$, respectively). This finding indicates that not only were evidentiary judicial instructions ineffective in shifting juror opinion, but also, presentation modality (i.e. written versus oral) appeared to have no influence on shifting opinions; both written and oral instructions demonstrated the same mean change and dispersion of opinion from pre-to post advice.

In addition, variability in the baselines between advice types deserves attention. By way of example, the mean increase in perceptions of reliable cues achieved by expert witness (+.36) represents the same difference between the baseline scores between expert witness (3.18) and oral judicial direction (3.54; where 3.54 - 3.18 = .36). Also notable, is that the post-test mean score for expert witness (3.54) was the same as the starting point for judicial direction (3.54). In this example, expert witness impacted perceptions significantly and oral directions did not. Yet, the starting point for oral directions (3.54) is where expert witness shifted to with advice. The point made, albeit crudely, is that while ceiling effects may be at issue, the variability in baselines between advice types may have confounded the effect of advice. This issue is critical to the interpretations made herein. Namely, the differing baseline perceptions may have undermine the opportunity for advice to impact perceptions equally, if much at all. The degree of variability observed across conditions prior to advice, and the degree of change observed with advice, cannot be ignored and requires that the findings of this thesis be interpreted with caution. As such, it is noted that the findings achieved with advice, as reported throughout, remain preliminary. Without further replication, extrapolating the findings beyond this thesis would be ill-advised.

An unexpected finding of the current study was the pattern found for the control. From time one to time two, jurors who received no instruction perceived reliable cues to be less indicative of deception than at baseline. Although non-significant, a change in perception and the direction
of the change was not anticipated. Because the shift in perceptions from time one to time two was non-significant, no conclusion can be made as to cautioning against omitting advice when witness credibility is at issue. However, what can be said, is that the directional nature of the change in opinion raises some concern and warrants future exploration.

Hypothesis one addressed the replication of Coyle and Thomson (2014), and to reiterate predicted, that jurors’ perceptions of reliable nonverbal deception cues would improve with advice. As noted, partial support was demonstrated, with mixed results found for the two advice conditions replicated (expert witness and written judicial direction). On the one hand, the effect of expert evidence on improving perceptions was replicated but the impact of written judicial directions was not. Unfortunately, corresponding effect sizes were not reported by Coyle and Thomson (2014), and so, the magnitude of change between their study and those found herein could not be evaluated. Needless to say, that although a pattern of improvement (significant or not) was not observed when jurors received written evidentiary directions, the introduction of the expert/judge condition provided additional support as to the possible benefits of educating jurors with instructions on detecting deception.

Notably, the baseline scores of reliable cues for jurors exposed to written instructions differed considerably between studies: Coyle and Thomson’s (2014) pre-test mean score was 3.03 whereas the current study was 3.35. It is noted that the discrepancies in baselines (i.e. +.32) may account for some of the difference found as to why there was a smaller mean increase achieved in the current study than the original one. In fact, the inconsistencies in the baseline scores for written instructions between both studies is larger than then change score found herein (i.e. +.32 difference in baseline scores between studies compared to +.01). Had the baseline score for written instructions corresponded with Coyle and Thomson’s (2014) study, then it is possible that the opportunity for perceptions to shift would have been greater and may have facilitated a different outcome. If we consider the differences in starting points between the two studies (+.32) it must
be noted that the baseline differences are almost as large as the biggest difference from pre-to post-advice found herein for expert witness (i.e. reliable cues mean increase was +.36). Similarly, pre-test scores suggest that jurors’ baseline knowledge of deception indicia were already at mid-point. Consequently, it is unknown whether or not potential floor and/or ceiling effects have restricted the degree to which opinions were, or were not, shifted within the current study.

Given the stark contrast in the effect of written judicial instructions, it is of relevance that Coyle and Thomson (2014) proposed that their unexpected finding for the effectiveness of written instructions may have been down to participants cribbing answers from the transcript of advice as this was available to them while they completed the follow-up questionnaire. Given that the opportunity to crib was minimised in the current study, this may explain some of the negligible difference found for written instructions herein. Alternatively, the differences in the impact of advice on juror perceptions demonstrated between studies may be accounted for by differences in methodology.

Although the replication utilised similar procedure and identical materials to those adopted by Coyle and Thomson (2014), key elements of the experimental settings differed: (1) the replication was conducted in an online environment compared to a in-situ, formal mock court setting; (2) participants had the opportunity to complete the study at home or work where interferences could not be controlled for; and (3) the current researchers were not on hand to qualify questions or provide helpful directives to participants in the online environment. Regarding design, the sample representativeness and cell sizes were not replicated, and the smaller samples obtained may have undermined power in the current study. Most notably, the impact that a formal, mock court setting had on juror engagement and attention cannot be underestimated. Equally, the potential that a written direction was afforded more weight and/or credibility in a mock court setting compared to an online study cannot be overlooked. It is likely, that the novel setting and presence of researchers encouraged participation. Recent studies that have looked at training in detecting
deception and that have attempted to modify judicial instructions to encourage comprehension have illustrated the importance of motivation, effort and participating in novel experiences as factors influencing training outcomes (Dunbar et al., 2018; Baguley et al., 2017). The mere social interaction between participants, researchers and materials afforded in a simulated jury setting may have, unwittingly, attributed more importance and seriousness to the advice presented, but also, to the task at hand compared to those who participated in the online experiment. These factors may account for some of the differences in the success of education provided to jurors observed between studies.

**Hypothesis 2: Jurors’ perceptions of unreliable cues of deception will improve with advice**

Partial support was found for hypothesis two that, predicted jurors’ misconceptions as to nonverbal cues of deception could be corrected with advice. Specifically, the current study found that jurors’ false beliefs regarding unreliable deception cues may be corrected with instructions when evidence is adduced by an expert witness (large effect, $d = .85$), or delivered as a written judicial direction (approached a medium effect, $d = .45$). This finding reveals that following instructions, jurors perceive unreliable cues to be less indicative of deception, suggesting that misconceptions of nonverbal cues not indicative of deception might be remedied with advice.

The current study also examined whether the effectiveness of advice on correcting misconceptions differed by type of instruction. The findings indicate that the impact of advice did not differ significantly between source of advice suggesting that jurors may benefit equally from the testimony of an expert or a written direction from a judge when it comes to correcting erroneous perceptions.

Advising potential jurors via expert/judge combined was marginally significant ($p = .06$). As indicated by post-hoc power analyses, it is likely that the mean decrease achieved by expert/judge ($M$ difference = -.16, $SD = .47$) did not reach significance due to a lack of power ($n = 21$). A lack of power is further substantiated by way of comparisons to: 1) the approximate mean
decrease and variability in responses found for written directions (M difference = -.17, SD = .40) that achieved significance with a larger sample (n = 29); and 2) significance obtained by the control for a notably smaller mean decrease (M difference = -.08, and SD = .22, p = .01) but much larger sample (n = 53).

The impact of oral judicial directions however, was less effective with verbal directives demonstrating no effect on shifting opinions of unreliable cues. In fact, although the shift in perceptions was non-significant, the direction of the shift is of note. That is, jurors exposed to oral judicial instructions perceived non-diagnostic cues to be more indicative of deception than before they received advice. While it is re-emphasised that the shift in perceptions was non-significant, the directional pattern warrants comment: jurors’ opinions trended in the opposite direction than desired and perhaps provides some cautionary suggestion that misconceptions may get worse not better. That said, the increase in false beliefs following judicial directions occurred in isolation – all other conditions of advice showed a pattern of improvement with misconceptions decreasing with advice.

From time one to time two, jurors exposed to no instructions perceived false, unreliable cues to be less indicative of deception than at baseline. This finding was unexpected. Although the shift was significant and the size of the effect was moderate, the mean decrease was the smallest of all sources of advice (excluding judicial directions that increased with advice). The level of significance obtained most likely reflects a combination of increased power due to a larger sample (n = 53) together with compacted scores of responses that did not disperse far from the mean (SD = .22).

Aside from measuring shifts from baseline, an additional analysis comparing jurors’ perceptions of reliable and unreliable cues after receipt of instruction was performed for each type of advice. The comparison was a within-participant assessment and so, is detailed here with the within-group analyses.
Baseline differences between perceptions of reliable and unreliable cues did not differ across the five levels of advice. That is, reliable cues and unreliable cues were perceived by jurors as equally indicative of deception prior to instruction. This finding is in and of itself an interesting one. Not only because it has not been explored despite the amass of studies on deception detection, but because it provides preliminary evidence to suggest that although people hold false beliefs about what deceptive behaviour looks like (i.e. Pinocchio’s nose), false beliefs appear to be no more dominant than valid stereotypes. Nor do they appear mutually exclusive. The equivalence found between reliable and unreliable opinions as to nonverbal cues perceived indicative of deception suggests that the focus of current deception literature that perpetuates that stereotypes of nonverbal behaviour are overwhelming wrong (Akehurst et al., 1996; Bogaard & Meijer, 2018; Bogaard et al., 2016) may be misdirected.

Post-test comparative analysis revealed that following instruction, jurors perceived reliable cues to be significantly more indicative of deception than unreliable cues for three types of instruction: when advice was presented by expert witness; written judicial direction; and expert/judge combined. A large effect was achieved when the advice of the expert and judge was combined ($d = 1.46$), the expert approached a large effect ($d = .75$), and the discrepancy in opinion between reliable and unreliable cues achieved with written instructions was moderate ($d = .52$). Only oral judicial directions had a negligible effect ($d = .04$) on differentiating opinion between reliable and unreliable diagnostics of deception. The effect on perceptions of indicia when no instructions were given were small ($d = .21$). Albeit, the difference in perceptions represented little practical significance. The finding highlights that while instructions might not uniformly effect shifts across perceptions of reliable and unreliable indicia when cue-types are evaluated independently, presenting advice to jurors may be effective in splitting opinion between reliable and unreliable cues in a desired direction.
Summary of Within-Group Findings

The current study demonstrated that shifts in perception from baseline might be enhanced with instruction. Overall, only the testimony of the expert witness simultaneously benefited jurors’ opinions of reliable and unreliable cues of deception. Aside from the expert, instructions by expert/judge appeared to enhance perceptions of reliable indicia, and written instructions appeared to positively impact perceptions of unreliable cues. Only instructions delivered as an oral judicial direction had no bearing on perceptions of reliable or unreliable indicators of deception. The ineffectiveness of oral directions was highlighted when jurors’ perceptions between indicia were compared: perceptions of reliable and unreliable indicia remained all but stable despite advice. For the three remaining types of advice, opinions between cue-types were successfully split. Moderate to large effects were achieved with reliable indicia perceived by jurors to be significantly more indicative of deception than unreliable cues following instructions. This demonstrated a considerable shift from baseline where jurors perceived reliable and unreliable cues to be equally indicative of deception.

The pattern of results for the control were unexpected: perceptions for both types of cues decreased from time one to time two, despite no intervention. The findings from the control may indicate a retest effect where the simple repetition of the questionnaire alone influences a change from baseline on the reliable sale, and unreliable scale more so. While the control demonstrated a negative trend, the trend observed for jurors who received advice diverged. Namely, depending on type of advice, perceptions of reliable cues increased whereas perceptions of unreliable cues decreased. Given that the direction of perceptions of unreliable cues were consistent for advice and no advice, the findings for perceptions of unreliable cues need to be interpreted with care.

In sum, the findings indicate that the shifts in perceptions observed from baseline where each participant acted as their own control may be due to the type of advice received, differences in baselines, or both. Similarly, the findings suggest that an inherent degree of change in
perceptions occur as a consequence of no instructions. Needless to say, that the effect of advice (as evidenced by improvements in opinion from baseline) do not tell the whole story. The subsequent section focuses on the findings pertaining to the effect of instructions compared to no instructions, controlling for potential confounding questionnaire retest effects that may have, unwittingly, impacted the findings outlined above.

7.1.2 Assessing Differences Between Advice and No Advice: Analyses between-groups.

Beyond the replication, the addition of a dedicated control group not only advanced the research design of this thesis but, allowed for alternative between-group comparisons. The following section discusses the findings of the between-group analyses relative to the hypotheses, comparing jurors’ perceptions between those who received instructions and those who did not.

Hypothesis 1 Continued: Jurors’ perceptions of reliable cues of deception will improve with advice

When the control was included in the analyses (i.e. mediating for any effects such as repetition), the effect of advice held: perceptions of reliable deception indicia appeared to be influenced by advice. Jurors’ perceptions of reliable indicia benefited from advice when it was presented by an expert witness and when advice was presented by an expert/judge collectively. The impact of written and oral judicial instructions had no effect on shifting perceptions when compared to those who received no instruction. Analysing the effectiveness of advice type revealed that neither the expert nor expert/judge was any more or less successful than the other in enhancing perceptions of reliable cues with both types of advice attaining large effects (expert witness: $d = 1.02$; expert/judge $d = .89$).

The findings for jurors’ perceptions of reliable cues are consistent with those found in the within-group analyses noted above because, the pattern of mean scores for perceptions diverged from the control. The improvement in opinions of reliable cues following advice can be interpreted with some promise because not only does the pattern of results follow the directions predicted (and
as previously demonstrated by Coyle and Thomson, (2014)), but they are contrariwise to the control. Despite these encouraging findings, the variability in baselines between advice types cannot be underestimated. It remains unclear to what extent the differing starting points impeded the effect of advice on influencing opinion.

_Hypothesis 2 Continued:_ Jurors’ perceptions of unreliable cues of deception will improve with advice

When the effect of instructions was compared to no instruction, shifts in jurors’ perceptions of unreliable cues moderated. Only one type of advice improved opinion of reliable indicia and that was when evidence was adduced thought expert testimony. The effect the expert witness in shifting jurors’ opinion regarding reliable cues remained large ($d = .96$).

The correction to false beliefs achieved via written judicial directions found in the within-group analysis dissipated when the effect of written instructions was compared to no instruction. Although the pattern of results suggests an improvement in misconceptions (i.e. a mean decrease), the difference effected by written instructions was not big enough from the effect of no advice alone (mean difference = .09). The same can be said for the pattern of improvement in false beliefs demonstrated with instructions by expert/judge (mean difference = .08). Accordingly, these findings indicate that when the effects of advice are isolated from possible re-test effects, the improvements in perceptions of unreliable cues tends to dissipate when judicial instructions are provided in writing.

Differences in opinion between cue-types achieved with instruction compared to no advice revealed a slightly different picture. The effect of instructions in dispersing opinion between reliable and unreliable cues was maintained for expert evidence and expert/judge combined as found in the within-group comparisons. Only differences in perceptions between reliable and unreliable cues contracted when the effect of written instructions were compared with no instructions. The effects of advice by expert evidence and expert/judicial direction on separating
opinion were large ($d = 1.02$ and $1.10$, respectively) with both types of advice influencing perceptions equally. Given that the two types of advice involve the evidence of an expert, one could argue that there are no foreseeable advantages of instructing jurors by both types of advice over that of the expert alone.

**Summary of Between-Group Findings**

The present study provides preliminary indications that jurors might benefit from instructions on diagnostics of deception over no instruction. The effect of expert evidence was upheld when the influence of advice was compared to no advice, with jurors’ perceptions of reliable indicia enhanced, and misconceptions corrected. The only other type of advice that successfully enhanced perceptions of reliable cues was when the evidence of an expert was combined with a judge’s direction. Aside from the expert, no other type of advice corrected jurors’ misconceptions of indicia of deception following immediate advice.

A broader effect of advice was highlighted when differences in opinion between cue-types were compared. The findings revealed that although advice by expert/judge did not modify perceptions of unreliable cues, the instruction influenced jurors’ opinion of reliable cues so much so that it was significant enough to split opinion with jurors perceiving reliable cues to be significantly more important than unreliable cues when determining the truthfulness/deception of a witness. Given that the advice of the expert was effective in improving perceptions of both reliable and unreliable cues, the impact of the instruction resulted in jurors perceiving reliable cues to be significantly more indicative of deception and unreliable cues to be significantly less indicative of deception. This was confirmed by the significant split in opinion found between type of cue.
7.1.3 Overall Conclusions

The Effect of No Instruction on Perceptions of Deception

The variability found within the control saw decreases in cues perceived indicative of deception (i.e. reliable cues) and decreases in cues perceived not indicative of deception (i.e. unreliable cues). The shift in opinions was consistent for both types of cues indicating that reliable and unreliable indicia were considered equal by jurors not exposed to advice.

The similar mean decreases across cue-type indicate an inherent level of variability in response from time one to time two: the mere function of re-administering the questionnaire appears to have influenced opinions. It is suggested that the degree of variability reflects participants’ inability to accurately remember the ratings of all 31 cues from a possible six-point scale (6 x 31 = 186 possible responses) rather than a true change in perceptions. That said, the differences found within reliable and unreliable ratings indicates a standard level of variability inherent in completing the questionnaire and/ or participating in the study that cannot be discounted.

The degree of change observed within the controls suggests that, at least for unreliable cues, caution is warranted when interpreting what type of advice successfully corrected misconceptions. That is, misconceptions appear to be corrected by a small degree on their own, over time. Therefore, to determine which type(s) of advice impacted misconceptions, it is necessary to consider shifts in opinion with respect to no advice. To not do so may result in mistaken attributions of success. By example, although opinions of unreliable cues decreased following written instructions and marginally following the advice of expert/judge combined, the reductions were no greater (statistically) than was found for jurors who received no advice.

Taken together, the consistent mean decreases in perceptions of reliable cues and unreliable cues indicates that a retest effect was most likely. It is possible that at time two, participants either
doubted themselves or could not remember their responses. It is noted that the changes in mean scores for the control were small and while statistically significant the “significance” in real terms may be limited. The order of differences observed were in terms of hundredths, with degrees of change down to two decimal places. Nonetheless, the omission of a placebo within the control such as an intervening task cannot be overlooked as it may have inadvertently effected the outcome of the results. Future research would be well placed to test an equivalent control to see what impact, if any, it has on the directionality of opinion when no advice is given.

The Effect of Instruction on Perceptions of Deception

Jurors’ perceptions of nonverbal cues of deception appear to be enhanced with advice. Instructions appear effective in shifting perceptions regarding both reliable and unreliable indicia. This was established when shifts were measured from baseline (within-group), as well as when opinions were compared with those who received advice and those who did not (between-group). Notwithstanding these findings, as noted previously, the variability in pre-existing perceptions across conditions require that the differences found with advice be interpreted with caution.

Only one type of instruction consistently benefited jurors and that was the testimony of the expert witness. Improvements were gained with expert evidence across reliable cues perceived indicative of deception as well as corrections made to misconceptions regarding cues non-diagnostic of deception. Aside from the expert witness that successfully influenced jurors’ opinions of both reliable and unreliable cues, perceptions of reliable cues (on their own) were also enhanced when advice was presented via expert/judge and this carried across both analyses (within and between-group). The effect of the two sources of advice were maintained because the pattern of results diverged from the control.

When shifts from baseline were assessed, written judicial instructions were successful in correcting mock jurors’ misconceptions regarding non-diagnostic cues. However, when the effect of the instruction was isolated from any potential confounding retest effects, the influence that the
written instruction had on improving opinions dissipated. This may be attributed to the pattern found for the control that demonstrated a consistent decreasing trend with that of written instructions (i.e. they both decreased from time one to time two), and expert/judge combined.

The control group of no advice demonstrated a decreasing trend on both reliable cues and unreliable cues from time one (Q1) to time two (Q2). For reliable cues, the overall pattern observed across the levels of advice opposed the direction of the control. However, for unreliable cues the pattern was consistent with the direction of the control indicating that the findings for unreliable cues needs to be interpreted with caution when comparing shifts from baseline for unreliable cues.

Although large effects were found (expert witness $d = 1.02$; expert/judge $d = .89$) both were equally effective in promoting opinions of reliable cues. However, the advantage of the expert witness was with respect to its broader application in enhancing not only reliable opinions, but also in correcting false opinions of unreliable diagnostics of deception.

When jurors’ opinions between types of cue were compared, the findings revealed that looking at the effect of instructions on reliable and unreliable cues irrespective of one another may oversimplify the effect of advice in successfully separating opinion between reliable and unreliable cues. The findings suggest that while some types of advice (expert/judge) might only effect change on one cue-type, the effect is big enough to split opinion in a desired direction with reliable cues perceived to be significantly more important in detecting deception even when false cues remain stable.

Regarding the replication, the impact of the expert witness on shifting opinions regarding reliable indicia was consistent with what Coyle and Thomson (2014) found. However, the non-effect observed for written judicial directions differed. Possible explanations include: differences in methodology (i.e. online study versus in situ experiment) that may account for the discrepancies in findings. Furthermore, the written direction may have lacked perceived credibility and/ or authority in the current research compared to the original study because it was not considered by
participants in a moot court setting. While the written direction did impact perceptions of unreliable cues it is possible that the effect would have been bigger if it had been conducted in a moot court. Hence, the importance of context regarding the power of the social interaction and formal setting cannot be discounted when it comes to interpreting the effect, or lack thereof for written judicial directions.

In light of the control, the following observations are made. First, that the simple effect of question repetition cannot be overlooked as the influence of advice by written directions diminished when this was controlled for. So, although written instructions shifted opinions of unreliable cues from baseline these shifts did not hold when the effect of advice was isolated from any effects associated with repeating the survey. To the contrary, what can be said is that expert witness replicated and that it held when retest effects were controlled for. However, the same cannot be said for the effect of advice by written directions: the current study suggests that the results found for written directions in the Coyle and Thomson study should be interpreted with caution. What remains unknown is whether or not the shifts in perception found in the previous research were due to type of advice, the design of the questionnaire, both, or some other extraneous variable(s) such as cribbing.

Given the negligible effect that oral directions had on shifting perceptions, it is probable that the improvement achieved with advice by expert/judge was largely contributed by the effect of the expert testimony alone. Some possible influencing factors include:

1) The presentation order of instructions where the testimony of the expert witness was delivered first followed by the judicial direction. It is possible that finishing the instruction with the less effective judicial direction had a weakening effect and may have contributed to a recency effect. While the combined expert/judge instruction was significant, it is possible that the smaller mean increase observed for expert/judge was a consequence of jurors finishing with judicial direction. Further, although perceptions significantly
improved with advice by expert/judge, the smaller standard deviation indicates less variability in responses post-advice suggesting that either less jurors changed their opinions or shifts in opinion were more consistent when advice was presented by expert/judge than expert alone. Further research would be valuable to investigate whether the order of presentation had any bearing on the efficacy of instruction when more than one type of advice was considered.

2) The duration of the instruction given by the expert in the combined expert/judge condition may have impacted the results. That is, the evidence adduced by the expert was longer than the instruction delivered by the judge. In fact, the instruction given by the expert was 10 times longer than that of the judge and so it would be reasonable to suspect that it carried more weight than the judicial direction.

3) The credentials of the expert witness may have had influence over that of the judge. It is possible that the expert had higher face validity (and therefore ecological validity), given that they were a real expert. Conversely, the judge was not an appointed member of the judiciary but rather was portrayed by a person known to the researcher. Consequently, jurors may have been swayed by the valid credentials of the expert who was presented first (i.e. a primacy effect), enhancing the overall credibility of the combined condition more so than when judicial advice was presented on its own.

4) The findings for expert/judge combined compared to judicial direction alone may indicate that the presentation of two independent forms of advice had a repetition effect (Bertelson, 1963; Kovera et al., 1997; Smith, 1968). The mechanisms of frequency (i.e. advice was duplicated) and exposure duration (i.e. the length of combined advice was longer than that of others; Hintzman, 1976), may provide some explanation as to why the effect of combined advice was greater than found for judicial directions. Furthermore, the repetition and similarity of material presented in the combined expert/judge condition is consistent
with the property of proportionality (Hintzman, Curran & Oppy, 1992), and may have had the unintended consequence of strengthening short-term memory retention and recall.

For the within-group analyses it is likely that change in perceptions of unreliable cues found for judicial instructions (oral and written) was a consequence of retest effects rather than advice alone. This conclusion is substantiated by the fact that the oral and written instructions were identical in content, only the presentation modalities differed. So, unless reading the transcript was more advantageous than hearing a direction (i.e. quick readers could re-read and rehearse the material), the differences are unlikely to be due to the content of the advice. More so, if the modality of written advice was superior to oral instructions one might expect opinions of reliable cues to have differed between oral and written instructions as did perceptions of unreliable cues but, they did not; they were consistent and neither improved with advice.

The provision of judicial instructions, whether verbal or in writing, has been extensively debated. The lack of effect found for judicial directions (written and oral) is consistent with extant research that concludes jurors’ comprehension of traditional judicial instructions are typically poor (Blankenship et al., 1997; Baguley, et.al., 2017; English & Sales, 1997; Luginbuhl, 1992; Ogloff, et al., 2011; Ogloff & Rose, 2005; Rose & Ogloff, 2001; Severance & Loftus, 1982; Spivak, et al., 2018). That said, the judicial directions tested herein were not formulated as traditional instructions. Rather, they were evidentiary-based directives underpinned by psychological research (Ribbers & Henneberg, 2018), and had been simplified with scientific and legal jargon removed (Bagley et al., 2017). Nevertheless, the effectiveness of evidentiary-based instructions reported in similarly conceptualised studies remains mixed. Research that has explored the efficacy of evidentiary-based instructions on correcting jurors misconceptions regarding child sexual assault (Goodman-Delahunty et al. 2011) and warning jurors as to the limitations associated with some forensic sciences (Ribbers & Henneberg, 2018) have demonstrated positive effects for judicial instructions. However other studies that have investigated the benefit of providing judicial aids
(such as visual props, the use of mnemonics and written instructions) and providing directions that are tailed to the facts of a case have been ineffective in educating and assist juror-decision making (Bagley et al., 2017; Jones et al., 2017).

Similarly, discrepancies in the effect of written instructions have been found. The ineffectiveness of written directions compared to no instruction found herein is in keeping with the results of Baguley and colleagues (2017) who found written instructions were of little assistance to triers of fact with the applicability of instructions undermined by low comprehension. Conversely, Thomas (2010) found written directions were twice as effective as oral directions in shaping juror-decision making. However, Thomas’ (2010) finding may be explained by having presented jurors with a written transcript along with an oral direction. The combined effect of an oral and written directions was not part of the current design and would need to be investigated to establish whether the same effects could be achieved when judicial directions are duplicated.

Notwithstanding these observations, before the effect of judicial directions are discounted altogether, a simpler explanation for the lack of effect ought to be contemplated. The judicial directions as an intervention may have been too short not giving jurors enough time to 1) absorb the material and/or 2) to have an impact on beliefs. That is, that the duration of the intervention was comparatively short. In the oral condition the intervention lasted one minute and fifty-one seconds compared to the expert witness’s evidence that went for ten minutes and fifty-three seconds, and the combined expert/judge condition that took twelve minutes and forty-four seconds. In the written condition, a period of three minutes was allowed to read the brief instruction. Accordingly, the short duration of the intervention (both written and oral) presented by the judge may not have afforded participants enough time to focus on and attend to the information presented thereby inhibiting the overall effect of judicial direction.

Additionally, for the oral direction, although the judicial officer was robed in formal legal regalia consistent with that of a Supreme Court Justice it was not delivered by an official judge and
may as a consequence lacked credibility, undermining the effectiveness of and weight attributed to the instruction on the whole. Equally, the written instruction had little context. Participants read the transcript in an informal setting (i.e. home) with no situational reference to the seriousness, importance or authority that would ordinarily accompany such a direction that might, at least, have been achieved in a moot court. Additionally, for the written direction there was no memorable reference point or anchor that may have helped prompt recall.

7.2 What Combination of Nonverbal Cues do Jurors Considered Most Important?

Based on the findings herein, a handful of cues dominated juror opinion as to which cues were ranked important in detecting deception. After exposure to advice, a small number of misconceptions prevailed with jurors maintaining that ‘confidence’ and ‘responses that contain a lot of detail’ were the two most important cues diagnostic of deception. Seemingly, this finding implies support for Dangerous Decision Theory (Porter & Brinke, 2009), in that some misconceptions about deceptive behaviour are so entrenched (i.e. confidence and responses that contain a lot of detail) that they don’t change despite receiving evidence to the contrary. However, post-hoc content analysis revealed that the two cues ranked most important by mock jurors – confidence and responses that lack a lot of detail – were not in fact referred to the advice proffered. Consequently, if jurors did not hear evidence to discredit the unreliable cues then it cannot, as a matter of logic, be expected that the misconceptions be corrected with advice. Moreover, for cues apportioned such importance, it would be counterintuitive to expect jurors’ opinions to change without evidence to the contrary. As such, the impact of advice on these two cues cannot be determined.

As such, to conclude that misconceptions remain well-entrenched somewhat distorts the picture. Although unreliable cues remain ranked among the most important by jurors, it is less misleading to look at the percentage changes to see if advice directly impacted those cues attended to. The results appear promising. By way of example, all sources of advice made specific
reference to the following indicia being indicative of lying: high-pitched voice; body movements; responses containing fewer details; and pauses. Notably, each one of these cues directly correlated with the largest percentage increases in importance following advice: high-pitched voice increased 450%; body movements and responses containing fewer details both increased 300%; and pauses increased 150%. This outcome reflects an observable relationship between the cues specified in the advice given and those cues that increase in importance with advice. At the individual level, these findings suggest that advice can effect considerable change on reliable cues. Although impressive, reliable interpretation of cues such as high pitch and body movements requires some baseline knowledge of the individual. Baseline knowledge is beneficial so that deviations from typical behaviours can be assessed against those exhibited when lying in order for reliable deductions to be made as to the credibility of a witness. The eventuality that the witness will be known to a member of the jury, let alone the entire jury pool, is unlikely to transpire. Consequently, the value in improving jurors’ knowledge base of cues such as these is of little practical import.

Improvements in juror opinion at the individual factor level can be directly linked to those specified by advice. Then, advice that concentrates attention on the combinations of cues perceived by jurors as most important may have a greater effect in shifting juror opinion than advice that is less specific and does not target the handful of cues identified by jurors. The combination of cues that remain stable over time as well as those that shifted with direct evidence present wider implications for the effect of advice in general. A potential consequence of evaluating indicia not cited in advice given to jurors is that the collective analysis of behavioural indicia may have, unwittingly, diluted the overall effect of advice.

7.3 Does Advice Influence Juror Confidence in Detecting Deception?

Exploratory analyses demonstrated that perceived confidence in detecting deception was not influenced by advice and that source of advice appeared irrelevant. This finding indicates that while advice may influence jurors’ opinions of reliable and unreliable diagnostics, these
enhancements do not inflate or degrade one’s confidence. This finding indicates that confidence in detecting deception may not be associated with receipt of advice and/or shifts in opinion.

Nonetheless, it is recognised that no context was provided in which jurors could apply or solidify the advice received, and this lack of opportunity to contextualise the knowledge acquired may have inadvertently moderated confidence. It may be that jurors’ confidence remained stable because the opportunity to apply the information received was not tested. However, because a control for confidence was not obtained, it cannot be determined whether or not advice, or lack of context, moderated confidence. That is, it remains unknown whether perceived confidence would have significantly differed between jurors who received no instruction. It would be of interest to examine a control for juror confidence against the impact of expert testimony to better confirm/disconfirm whether a relationship exists between the provision of advice and confidence in detecting deception.

Based on DePaulo and colleagues (1997) “overconfidence effect” that concludes increased confidence is a poor predictor of accuracy in detecting deception, the indication that advice may not impact confidence might offer some reassurance. If increased knowledge as to behavioural indicators of deception is not associated with increased confidence, then, according to the “overconfidence” principle (DePaulo et al., 1997), jurors’ accuracy in detecting deception should not be compromised. Rather, increased knowledge as to cues indicative of deception with tempered confidence may countervail with increases in accuracy achieved. That said, the relationship between confidence and detecting deception asserted by DePaulo and colleagues (1997), and that has been widely accepted for decades, may well be outmoded. Emerging research by Smith & Leach (in press) purports that confidence can reliably discriminate between accurate and inaccurate assessments of deception. Smith and Leach (in press) argue, unequivocally, that in line with Signal Detections Theory confidence and accuracy are positively related. They conclude that assessments of truthfulness/deception are more accurate, more often, when decisions are high in confidence
compared to decisions low in confidence. While accuracy was not examined in this thesis, should
the emerging evidence declared by Smith and Leach (in press) hold, the ability for advice to
enhance knowledge while also improving confidence may be advantageous other than just
enhancing perceptions alone. However, as it stands, the impact of expert and/or judicial advice on
accuracy in detecting deception remains unknown and the interaction of advice and confidence
found herein is no more than preliminary and cannot be generalised without replication. As such,
these relationships would be worthy of future investigation.

7.4 Chapter Summary

The current study evidenced partial support for the effect of advice on influencing juror
perceptions of indicators of deception. The enhancements gained with instructions suggest that it
may be possible to change initial perceptions of trustworthiness/deception with advice. Although
the current study partially replicated the two conditions of advice investigated by Coyle and
Thomson (2014) (expert witness replicated but written judicial direction did not), a consistent
pattern to Coyle and Thomson (2014) was observed along with the addition of advice by
expert/judge.

Beyond the replication, the current study demonstrated strong evidence for the need to
incorporate an independent control with the influence of advice varying when the effect of
instruction was controlled for. What is more, the current research revealed that jurors’ false beliefs
of which nonverbal cues are indicative of deception may be corrected with advice. Important
findings included uncovering that jurors appear to perceive reliable and unreliable cues to be
equally indicative of deception prior to advice. With instructions, opinions significantly departed
with reliable indicia perceived more indicative of deception than unreliable indicia. Moreover,
when the effect of advice was compared with no instruction, only the testimony of the expert
witness consistently improved jurors’ opinion of reliable cues while also correcting for false beliefs
of nonverbal cues of deception.
Exploratory analyses revealed that jurors afforded more weight to some cues than others. Although juror opinion maintained that a handful of unreliable cues were important in detecting deception following advice, these cues were not stipulated in the advice given. To the contrary, the reliable cues that were specified by advice, demonstrated marked increases which were not found on those unreliable cues initially preferred. The shifts in perception observed with advice suggests, albeit preliminarily, some evidence counter to the theory of Dangerous Decisions (Porter & ten Brinke, 2009) providing some encouragement that jurors may benefit from expert and/or judicial advice that targets the constellation of cues considered most important when detecting deception. Should expert evidence or judicial directives focus on these combinations of cues, then the return on investment may be greater.

Notwithstanding replication of these broader findings, the current study provides some preliminary empirical support for the potential benefit of adducing expert testimony to furnish jurors’ knowledge on diagnostics of deception so as to better equip their decision making when assessing witness credibility. Before triers of fact are directed to consider demeanour evidence, the current findings suggest that perceptions of reliable and unreliable indicia are weighted equally. However, preliminary evidence demonstrated herein indicates that when jurors are invited to consider nonverbal indicators of deception to assess the credibility of a witness, the absence of instructions that sensitise jurors to the limitations of nonverbal indicia may have negative, unintended consequences. Notwithstanding the need for replication, the evidence herein suggests that omitting or denying expert advice on detecting deception does not safeguard juror-decision making but rather may contribute to hazardous, erroneous decision-making.
Chapter 8. Study Two

8.1 Introduction

8.1.1 Delayed Presentation of Evidence: Only a matter of time?

Given the contingencies of a trial, it is not uncommon for members of a jury to consider evidence adduced by an expert at a later stage in proceedings, requiring that triers of fact retain the testimony proffered. This aspect introduces decay and interference effects which can undermine the retention of information (Anderson, 1990; Sternberg, 2003), although these have not been investigated in this context. However, the broader issue of delay has been raised with respect to training in detecting deception. A common shortfall cited in deception research is that the lasting impact of education on gains in knowledge or skill acquisition in detecting deception remains unknown (Blanch-Hartigan, Andrzejewski, & Hill, 2012; Frank & Feeley, 2003; Hauch et al., 2016; Levine et al., 2005; Shaw, Porter & ten Brinke, 2013).

Although short-term effects can be demonstrated (as evidenced in Study one), the question remains whether or not observed improvements are short-lived or persist over time (Blanch-Hartigan et al., 2012; Levine, 2018). To verify that knowledge gains are genuine (i.e. post-test exceed those at pre-test), Frank and Feeley (2003) suggest that trained recipients should outperform a control group as well as their own baseline performance at a week, a month, or one year later (also see, Levine et al., 2005). Similarly, Hauch and her colleagues (2016) suggest that post-test follow-ups should investigate different delay intervals to measure training effects over time. Indeed, they recommend multiple exposure sessions for professionals such as “judges to ensure that the training content will be retained and refreshed” (p. 35). Without such comparison it cannot be concluded that the efficacy of education on deception indicia is maintained over time, or that newly developed skills in detecting deception translate into practice (Shaw et al., 2013).
Research on forensic investigative interviewing has reported similar issues with the long-term impact of training on behaviour change. Powell (2002) notes that despite the provision of guidelines and intensive training programs on how to conduct effective investigative interviews, the long-term impacts on how interviews are carried out are negligible. Powell (2002) offers the conclusion that “while training programs may be effective in teaching interviewers what they should do when interviewing, the knowledge trainers acquire is having little impact on their practice” (p. 47).

In addition to the effects of decay and interference, perceptions about deceptive behaviour may undermine the persistence of education on indicia of deception over time. Frank and Feely (2003) advise “stereotypes about deceptive behaviour may be so strong that they re-emerge with little provocation” (p.67). This proposition mirrors that of DDT that predicts jurors’ initial impressions of trustworthiness are so long lasting that they are unlikely to change over time or in the face of conflicting evidence (Porter & ten Brinke, 2009).

In light of the findings from the replication of Coyle and Thomson’s (2014) findings (Study 1), advice on behavioural indicia may positively influence what triers of fact consider when judging the truthfulness/deception of a witness. However, the long-term effect of advising jurors about reliable/unreliable nonverbal indicia of deception remains unclear. The interval between the testimony of an expert and/or judicial direction and when jurors apply that evidence (e.g. determining the credibility of a witness,) may be a critical factor as to what influence that advice has on enhancing perceptions. Both the effluxion of time and subsequent events have the potential to undermine the retention and retrieval of jurors’ memory of the expert’s testimony or judge’s direction. However, the effect of delay between the testimony of the expert and jurors’ evaluation of that evidence is unknown. Simply put, the extent to which the provision of evidence improves the behavioural factors relied upon by jurors, over time, has not been previously investigated. This knowledge gap is the focus of Study two.
Chapter 9. Aims and Hypotheses

The second study seeks to investigate the effect of a delay between the presentation of advice and jurors’ consideration of that advice on perceptions of reliable and unreliable indicators of deception. This delayed consideration of advice could be expected to occur in a trial where earlier evidence presented by an expert and/or judge has to be relied upon later by jurors and applied to other evidence (i.e. determining the credibility of a factual or opinion witness).

9.1 Aim

Based on the findings from the Study one, Study two aims to provide a novel expansion to the research by exploring the influence of time on deception indicia considered important by jurors when determining the truthfulness/deception of a witness in court. As such Study two aims to:

1) Investigate whether the effect of expert evidence and/or judicial direction on improving jurors’ perceptions persist over time.

9.2 Hypothesis

Study two examines one directional hypothesis. It is predicted that:

2) The delayed consideration of advice will not improve jurors’ perceptions of nonverbal cues of deception. To test this hypothesis no change on the reliable and unreliable subscales are predicted from pre-delay (Q1) to post-delay (Q2).

Exploratory analyses will also examine the modality of presentation between written and verbal judicial directions and will investigate whether or not juror perceived confidence is influenced by advice following a one-week delay. As in Study one, no a-priori prediction is made for the effect of delayed advice on juror perceived confidence.
Chapter 10. Results

Study two investigated whether the delayed consideration of advice impacted cues of deception considered important by mock jurors in determining the truthfulness/deception of a witness in court. To answer this, within-group (Wilcoxon tests) and between-group (Mann-Whitney tests) analyses were conducted to establish whether changes in means scores of perceptions of cue-type held: (1) over time when advice was considered after a delay, and (2) whether changes in perceptions of cue-type differed between those exposed to advice with a delay from those exposed to delay without advice (control).

As in Study one, reliable and unreliable cues are reported independently with each condition of advice outlined separately to establish whether type of advice is relevant in shifting perceptions over time. The analyses of the within-group comparisons are reported first followed by the between-group comparisons. The results conclude with the exploratory analyses that examined what combination of indicia mock jurors perceived most important in detecting deception, and whether or not advice influenced juror perceived confidence with the passage of time.

Normality: Assumption testing including skewness and kurtosis scores, Shapiro-Wilks tests, and inspection of histograms indicated that the distribution of mean perceptions scores were non-normal (Field, 2013; Hubbard, 1978; Sprent & Smeeton, 1989; Tabachnick & Fidell, 2007). Visual inspection of boxplots revealed outliers at all levels of advice. Extreme cases, determined by the IQR x 1.5 rule, were removed from analysis. A total of 12 extreme values were excluded from the reliable subscale (expert witness n = 5; written direction n = 1, oral direction cues n = 3, expert/judge n = 1 and control n = 2). For the unreliable subscale, nine extreme cases were removed (expert witness n = 1; written direction n = 1, oral direction n = 3, expert/judge n = 2 and control n = 2). Consequently the cell counts for advice differ between cue-type.
Internal Consistency: Using Cronbach’s (1951) Alpha, the internal consistency of the reliable scale was good (12 items, \( \alpha = .80 \)) and the internal consistency of the unreliable scale excellent (19 items, \( \alpha = .91 \)).

10.1 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor? Assessing Shifts from Baseline: Analyses within-groups.

Wilcoxon signed-rank tests were undertaken to establish whether the mean score of importance for reliable and unreliable cues of deception diverged after delayed consideration of advice. Four types of advice were examined: expert witness; written judicial direction; oral judicial direction; and expert/judge combined. The mean scores of importance for reliable cues are reported followed by unreliable cues with each condition of advice described separately. Directional differences were predicted, and so one-tailed tests were adopted.

10.1.1 Reliable Deception Cues

Expert witness: The mean score of summed items indicative of deception on Q1 was \( M = 3.46 \) and on Q2 was \( M = 3.69 \). A Wilcoxon test demonstrates that the mean increase (M difference = .22) on reliable cues was statistically significant after seven days when advice was presented by expert witness, \( z = 2.45, p = .01, d = .90 \), one-tailed.

Judicial written direction: The mean of summed items for reliable cues on Q1 was \( M = 3.21 \) and on Q2 was \( M = 3.29 \). The shift in reliable cues (M difference = .08) after a one week delay for advice presented by written judicial direction was non-significant, \( z = .78, p = .22, d = .16 \), one-tailed.

Judicial oral direction: The mean score for reliable cues was \( M = 3.28 \) on Q1 and \( M = 3.30 \) on Q2 respectively. The shift on reliable cues (M difference = .02) following delayed advice by way of oral judicial direction was non-significant, \( z = .10, p = .46, d = .05 \), one-tailed.
**Expert and judge combined:** The mean importance of reliable cues was $M = 3.45$ on Q1 and $M = 3.47$ on Q2. The mean difference increase from baseline (M difference = .02), when mock jurors considered expert/judge advice one week delayed was non-significant, $z = .39, p = .34, d = .03$, one-tailed.

**No advice:** The control for delay without advice reveals that the mean importance of reliable cues was $M = 3.44$ on Q1 and $M = 3.61$ on Q2, representing a non-significant mean increase from baseline (M difference = .17), $z = 1.36, p = .09, d = .46$, one-tailed.

Baseline comparisons were undertaken to determine whether perceptions of reliable cues differed between levels of advice prior to instructions. A Kruskal-Wallis independent samples test reports that baseline perceptions of reliable cues did not differ significantly between groups prior to advice, $\chi^2 (4) = 3.11 p = .54$ (two-tailed), with a mean rank of reliable cues of 49.58 for expert witness, 46.30 for written direction, 52.15 for oral direction, 58.68 for expert/judge and 59.61 for no advice.

The pattern of mean scores for each level of advice is illustrated in Figure 10 where Q1 reflects mock jurors’ baseline score and Q2 represents mock jurors’ score after a one-week delay. The statistical parameters (e.g. cell size, means, mean differences, $p$ values and effect sizes) are displayed in Table 7 for each condition of advice and can be found after the following section, Unreliable Deception Cues, 10.1.2.
Figure 10. Mean Scores for the Reliable Subscale by Source of Advice at Baseline (Q1) and After a One-Week Delay (Q2)
10.1.2 Unreliable Deception Cues

Perceptions of unreliable cues between advice groups did not differ at baseline as confirmed by an independent samples Kruskal-Wallis test, \( \chi^2 (4) = 1.17 \) \( p = .88 \) (two-tailed), with a mean rank of unreliable cues of 56.50 for expert witness, 50.67 for written direction, 53.81 for oral direction, 54.00 for expert/judge and 60.33 for no advice.

**Expert witness:** The mean score rating for unreliable cues was \( M = 3.35 \) on Q1 and \( M = 3.24 \) on Q2. The decrease from baseline (M difference = -.11) was non-significant when expert witness advice was considered one-week delayed, \( z = -.73, p = .24, d = .19 \), one-tailed.

**Written judicial direction:** The mean score of importance for unreliable cues on Q1 and Q2 was \( M = 3.11 \) and \( M = 3.13 \) respectively. The increase on unreliable cues from baseline (M difference = .02) was non-significant for delayed consideration of written judicial advice, \( z = .30, p = .38, d = .04 \), one-tailed.

**Oral judicial direction:** On Q1 and Q2 the mean score of importance for unreliable cues was \( M = 3.14 \) and \( M = 3.10 \) respectively. The decrease in importance after delay when advice was presented as an oral direction (M difference = -.04; SE = 39.33) was non-significant, \( z = -.46, p = .32, d = .08 \), one-tailed.

**Expert and judge combined:** The unreliable cue mean score on Q1 was \( M = 3.20 \) and \( M = 3.07 \) on Q2. The decrease in importance after delay for combined expert and judicial advice (M difference = -.13) was non-significant, \( z = -.106, p = .14, d = .25 \), one-tailed.

**No advice:** Unreliable cues on Q1 had a mean of \( M = 3.26 \) and a mean of \( M = 3.42 \) on Q2. The mean increase for unreliable cues (M difference = .17) after a one-week delay was marginally significant, \( z = 1.59, p = .06, d = .44 \), one-tailed.
The two types of instructions that significantly improved perceptions of unreliable cues (expert evidence and expert/judge) were compared to establish whether one form of instruction was more effective than the other. The results indicate that both types of advice were equally effective in correcting misconceptions of unreliable cues of deception with shifts in opinion achieved between expert evidence (M difference = -.11) and expert/judge (-.13) not significantly different from one another, $u = .00, p = 1.0 d = .04$.

Figure 11 displays the mean scores for unreliable cues across each advice condition before and after delay. Q1 represents the baseline assessment and Q2 represents the one-week time to assessment delay. The summary statistics arising from the between-group comparisons for unreliable cues are reported together with the statistics for reliable cues in Table 7.
Figure 11. Mean Scores for the Unreliable Subscale by Source of Advice at Baseline (Q1) and After a One-Week Delay (Q2).
Table 7

*Shifts in Mean Scores of Mock Juror Perceptions From Baseline (Q1) to After a One-Week Delay (Q2) by Cue-Type and Source of Advice*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>Reliable Cues</th>
<th></th>
<th>Unreliable Cues</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Expert Witness</td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>No Advice (control)</td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

<sup>p<sup>→</sup> calculated using Wilcoxon-matched-pair signed-rank tests (one-tailed); **Significance at .01
10.1.3 Reliable versus Unreliable Cues When Delay is a Factor

Baseline differences were compared between reliable and unreliable cues to determine whether potential jurors’ perceived one cue-type to be any more or less indicative of deception than the other. Similarly, differences in opinions between cue-type were examined post-advice to identify the nature and extent to which cues deviated with advice. As per Study one, although no directional hypothesis were predicted between opinions of reliable and unreliable cues at baseline or post-advice, one-tailed tests were employed given the directional nature of the existing hypotheses that predicted advice would enhance opinions of deception indicia.

Expert witness: Differences in mean scores at baseline (M difference = .08) between reliable cues (M = 3.44) and unreliable cues (M = 3.35) were non-significant within the expert witness condition, z = 1.47, p = .07, d = .27, one-tailed. After mock jurors received expert evidence, reliable cues (M = 3.64) were rated significantly higher than unreliable cues (M = 3.21; M difference = .43), z = 2.27, p = .01, d = .35, one-tailed.

Written judicial direction: The mean difference at baseline (M difference = .10) between reliable (M = 3.21) and unreliable cues (M = 3.11) was significant for participants assigned to written judicial instructions, z = 1.66, p = .05, d = .37, one-tailed. Following written instructions reliable cues (M = 3.29) were rated significantly higher than unreliable cues (M = 3.13; M difference = .16), z = 2.71, p = .05, d = .60, one-tailed.

Oral judicial direction: Prior to advice, the mean difference (M difference = .14) between reliable cues (M = 3.27) and unreliable cues (M = 3.13) was significant for judicial direction, z = -1.66, p = .05, d = .37, one-tailed. After advice reliable cues (M = 3.34) remained significantly greater than unreliable cues by mock jurors (M = 3.09), z = 3.71, p > .001, d = .95, one-tailed.
**Expert and judge combined:** At baseline, mock jurors rated reliable cues (M = 3.37) higher than unreliable cues (M = 3.19), however the mean increase was non-significant: M difference = .18, z = .88, p = .19, d = .40, one-tailed. Post advice, reliable cues were rated significantly higher (M = 3.48) than unreliable cues (M = 3.07; M difference = .41), z = 2.07, p = .02, d = .55, one-tailed. The pattern for reliable and unreliable cues for mock jurors exposed to a one-week delay for each level of advice is illustrated in Figure 12.

![Figure 12. Mock Juror Perceptions at Baseline (Q1) and After a One-Week Delay (Q2) by Cue-Type and Source of Advice.](image)

**No advice:** Comparative analysis reveals that participant’s perceptions of reliable (M = 3.44) and unreliable cues indicative of deception (M = 3.27) differed significantly at baseline (M difference = .18), z = 2.29, p = .01, d = .50, one-tailed. Despite the one-week delay, the mean difference between reliable (M = 3.61) and unreliable cues (M = 3.42) remained constant over time (M difference = .19) however the difference post-test was non-significant, z = 1.46 p = .07, d = .45,
one-tailed. The pattern of perceptions of reliable and unreliable cues for mock jurors exposed to delay without advice is shown in Figure 13.

![Figure 13](image)

Figure 13. Total Mean Scores for the Reliable and Unreliable Subscales at Baseline (Q1) and After a One-Week Delay (Q2) for Mock Jurors Who Received No Advice (Control).

A summary of the findings for the shifts between opinion of reliable and unreliable indicia follow in Table 8.
Table 8

*Post-Test Mean Scores of Mock Juror Perceptions of Reliable and Unreliable Cues by Source of Advice*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>N</th>
<th>M</th>
<th>M</th>
<th>M diff</th>
<th>SD</th>
<th>p&lt;</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Witness</td>
<td>13</td>
<td>3.64</td>
<td>3.21</td>
<td>.43</td>
<td>.62</td>
<td>.01**</td>
<td>.70</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td>30</td>
<td>3.29</td>
<td>3.13</td>
<td>.16</td>
<td>.46</td>
<td>.05*</td>
<td>.35</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td>25</td>
<td>3.34</td>
<td>3.09</td>
<td>.25</td>
<td>.27</td>
<td>&lt; .001***</td>
<td>.95</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td>16</td>
<td>3.48</td>
<td>3.07</td>
<td>.41</td>
<td>.74</td>
<td>.01**</td>
<td>.55</td>
</tr>
<tr>
<td>No Advice (control)</td>
<td>18</td>
<td>3.61</td>
<td>3.42</td>
<td>.19</td>
<td>.45</td>
<td>.07</td>
<td>.43</td>
</tr>
</tbody>
</table>

p< calculated using Mann-Whitney two-sample tests; * Significance at .05; **Significance at .01; ***Significance at .001.

Note: The mean difference (M diff) reflects the difference between the reliable and unreliable subscales post-advice.
10.2 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor?  Assessing Differences Between Advice and No Advice: Analyses between-groups.

Changes in opinion of reliable and unreliable cues after a one-weeks’ delay were compared between mock jurors exposed to advice and mock jurors not exposed to advice. Comparisons with the control were conducted to identify any potential effects associated with the passage of time other than the impact of advice that could not be discerned from the within-group analyses. The analyses aimed to establish whether or not advice impacted shifts in opinion over and above the effect of time alone. The results for reliable and unreliable cues are reported concurrently and the summary statistics for the between-group analyses can be found in Table 9.

**Expert witness:** When compared to the control, a Mann-Whitney test reveals that the mean difference on reliable cues after a delay did not differ significantly between mock jurors exposed to expert advice (M difference = .22) compared to those exposed to a delay but no advice (M difference = .17), $u = .53$, $p = .31$, $d = .15$, one-tailed. The mean change for unreliable cues following a delay differed significantly between mock jurors who received expert evidence (M difference = -.11) compared to those exposed to delay but who did not receive advice (M difference = .16), $u = -1.63$, $p = .05$, $d = .56$, one-tailed.

**Written judicial direction:** A between-group analysis found a non-significant difference between reliable cues rated by mock jurors presented with written directions (M difference = .08) compared to those who did not: M difference = .17; $u = -.71$, $p = .24$, $d = .19$, one-tailed. For unreliable cues, a Mann-Whitney test also indicates that after a delay, unreliable cues were not rated significantly different between mock jurors who considered written instructions (M difference = .02) from those exposed to delay but no advice (M difference = .16), $u = -.85$, $p = .20$, $d = .29$, one-tailed.
**Oral judicial direction:** A Mann-Whitney test confirms that after one-weeks’ delay, the mean difference for reliable cues between mock jurors exposed to judicial direction (M difference = .02), and those not exposed to advice (M difference = .17) was non-significant, $u = -.84$, $p = .20$, $d = .29$, one-tailed. For unreliable cues, the mean difference between jurors who considered judicial directions following a delay (M difference = -.04) from those not exposed to advice was non-significant: M difference = .16; $u = -1.42$, $p = .08$, $d = .42$, one-tailed.

**Expert and judge combined:** The mean difference for reliable cues between mock jurors exposed to delayed consideration of expert/judge advice (M difference = .02) from those exposed to delay without advice (M difference = .17) was non-significant, $u = -.23$, $p = .28$, $d = .24$, one-tailed. A Mann-Whitney test confirms however, that the mean difference for unreliable cues differs significantly between jurors exposed to delayed expert/judge advice (M difference = -.13) from those exposed to a delay but no advice: M difference = .16, $u = -1.75$, $p = .04$, $d = .65$, one-tailed.
Table 9

*Between-Group Comparisons of Mean Differences in Mock Juror Perceptions by Cue-Type and Source of Advice When Delay is a Factor*

<table>
<thead>
<tr>
<th>Source of Advice</th>
<th>Reliable Cues</th>
<th>Unreliable Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N)</td>
<td>M diff</td>
</tr>
<tr>
<td>Expert Witness</td>
<td>14</td>
<td>.22</td>
</tr>
<tr>
<td>Judicial Written Direction</td>
<td>30</td>
<td>.08</td>
</tr>
<tr>
<td>Judicial Oral Direction</td>
<td>26</td>
<td>.02</td>
</tr>
<tr>
<td>Expert &amp; Judicial Direction</td>
<td>17</td>
<td>.02</td>
</tr>
<tr>
<td>Control (no advice)</td>
<td>18</td>
<td>.17</td>
</tr>
</tbody>
</table>

\(p^\) calculated using Mann-Whitney two-sample tests (one-tailed); *Significance at .05

Note: The mean difference (M diff) reflects the change score from pre-to post advice (Q2 - Q1) for the reliable and unreliable subscales and was calculated to run between-group comparisons against the control.
10.3 What Combination of Nonverbal Deception Cues are Considered Most Important Over Time?

10.3.1 Top Five Most Important Cues

Prior to advice, 26 indicators of deception were ranked the most important by mock jurors in determining the truthfulness of a witness in court. The five cues considered most important were: confidence (23.7%); spontaneous corrections (15.5%); implausible responses (13.4%); responses that contain a lot of detail (12.4%), and responses that lack logical structure (9.3%). The top five cues accounted for 74.2% of all cues ranked number one. The proportion of reliable and unreliable cues that were deemed most important were relatively equal: 38.1% (n = 37) and 36.1% (n = 35) respectively. Figure 14 shows the rank order of all cues at pre- and post-delay.

After seven days delay, the number of indicators ranked most important by mock jurors reduced by eight items from 26 cues (pre-delay) to 18 cues (post delay). Over time, confidence (28.9%), implausible responses (14.4%), responses that contain a lot of detail (14.4%), spontaneous corrections (7.2%) and responses that contain very few details (5.2%) were ranked top five—four of which were also rated in the top five prior to delayed advice. Confidence remained the most important cue over time where the least important cue – responses that logical structure (4.1%) – dropped one rank to 6th most important cue replaced by responses that contain very few details (5.2%). The top five cues accounted for 70.1% (n = 68) of all items rated. Unreliable cues accounted for the largest proportion of cues ranked in the top five (43.3%) compared to reliable cues (26.8%) with confidence accounting for the largest share of ratings (28.9%). Although three of the five cues ranked most important (implausible responses, spontaneous corrections and responses that contain very few details) were reliable cues, the frequency of importance assigned to them (and therefore order) fell at the lower end (26.8%; order of rankings: 3, 4 and 5).
In terms of percentage change, high pitch voice (300%, \( n = 3 \)), body/hand/finger/leg/foot movements (200%, \( n = 2 \)), and responses that contain very few details (150% \( n = 3 \)) showed the largest increase after consideration of delayed advice (see Figure 15). All three items were reliable cues of deception. Unreliable cues that increased following delayed advice included self-manipulators (50%, \( n = 1 \)), confidence (22%, \( n = 5 \)), and responses that contain a lot of detail (17%, \( n = 2 \)).
**Figure 14.** Nonverbal Cues Ranked Most Important Before and After Advice When Delay Was a Factor. The grey dash line delineates the top five most important cues. Presentation order reflects cues ranked most important through least important at baseline.
Figure 15. Percentage Change in Nonverbal Cues Ranked Most Important After Advice When Delay Was a Factor. The percentage change is differentiated for the two measures: reliable cues represented in blue and unreliable cues shown in orange. Items are presented in order of percentage change from largest through smallest.
10.3.2 Most Commonly Preferred Cues

The cues most commonly endorsed by mock jurors were the same as the top five cues ranked most important: confidence (56.7%), responses that contain a lot of detail (50.5%), spontaneous corrections (37.1%), responses that lack logical structure (30.9%) and implausible responses (29.9%). Aside from confidence, only the order of preference shifted.

After delayed consideration of advice, the deception cues most commonly rated were: confidence (59.8%), responses that contain a lot of detail (52.6%), spontaneous corrections (37.1%), illustrators (37.1%); and implausible responses (35.1%). When asked to rank cues by importance, illustrators fell at the lower end for first and second order preferences (2.1%, n = 2; 4.1%, n = 4). However, when responses were considered overall, illustrators were rated equal third/fourth with spontaneous corrections as a preferred cue in determining truthfulness/deception. As a result, responses that lack logical structure dropped to sixth most preferred cue after delayed advice despite the proportion of responses remaining the same (30.9%). Furthermore, the two unreliable cues were preferred by more participants after exposure to delayed advice than before. While spontaneous corrections remained stable over time, implausible responses increased from 28.9% at baseline to 37.1% post delay. Figure 16 charts the cues most commonly endorse before and after delayed advice.

Self-grooming (333%, n = 10), high pitch voice (229%, n = 16), and blinking (117%, n = 8) revealed the largest percentage increases after delayed advice (see Figure 17). The largest percentage decreases were for ambivalent responses (-38%, n = -6), pauses (-28%, n = -7), speech disturbances and covering the mouth (-23%, n = -3, respectively) of which the split between number of reliable and unreliable cues was equal.
Figure 16. Most Commonly Endorsed Nonverbal Cues Before and After Advice When Delay Was a Factor. Cues from left to right are in order of most preferred to least preferred from baseline.
**Figure 17.** Percentage Change in Nonverbal Cues Most Commonly Endorsed After Advice When Delay Was a Factor. The percentage change is delineated for the two measures: reliable cues depicted in blue and unreliable cues charted in orange. Items are presented in order of percentage change from largest through smallest.
10.4 Does Exposure to Advice Influence Jurors’ Confidence Over Time?

Perceived confidence in detecting deception was measured at baseline and again over time to determine whether delayed consideration of advice influenced jurors’ perceived ability to detect deception. Investigating the effect of delay on advice was exploratory and so no hypotheses were formulated. Shifts from baseline (i.e. within-group) were evaluated because an equivalent control was not obtained for confidence. The aim of the current analyses was to determine whether the effect of instruction influenced perceived confidence. Therefore, the sample that was used to investigate the impact of instructions on perceptions was also investigated here. Data cleaning supported parametric analyses with paired-sample t-tests undertaken to compare shifts in confidence reported over time.

**Expert witness:** Mock jurors’ level of confidence shifted from $M = 3.38$ (SD = .77) at baseline to $M = 3.69$ (SD = .85) following exposure advice by expert witness. The mean increase in confidence ($M$ difference = .31, SE = .17) for those who received expert testimony was non-significant, $t (12) = 1.76$, $p = .10$, $d = .48$, two-tailed.

**Written judicial direction:** Confidence mean scores for mock jurors who received written judicial instructions shifted from $M=3.50$ (SD = .73) to $M = 3.47$ (SD = .94) following delay but the mean decrease was negligible: $M$ difference = .03, SE = .12; $t (29) = -.273$, $p = .79$, $d = .05$, two-tailed.

**Oral judicial direction:** For those who received judicial directions, confidence increased by a mean difference of .09 (SE = .11) from $M = 3.50$ (SD = .91) at baseline to $M = 3.59$ (SD = .73) post-advice, but the increase was non-significant: $t (21) = .70$, $p = .43$, $d = .17$, two-tailed.

**Expert and judge combined:** Confidence mean scores remained stable over time for mock jurors who received both expert and judicial advice combined ($M = 3.25$, SD = .85, 1.00, respectively) with no change observed from baseline.
Chapter 11. Discussion

The current study aimed to test whether delayed consideration of advice would impact jurors’ perceptions of nonverbal cues of deception. The current study extended on from Study one to establish whether changes in perceptions held over time. The objective was to establish whether the effects of decay and interference would influence the efficacy of advice provided to jurors when delay was a factor. In doing so, the current study tested the basic tenet of Dangerous Decision Theory (Porter & ten Brinke, 2009), that argues opinions of trustworthiness/ deception are so ingrained that despite evidence to the contrary (i.e. instructional advice) they don’t change.

The structure of the discussion is as follows: the findings of the within-group analyses (e.g. changes from baseline) are discussed first followed the between-group comparisons (e.g. the effect of instruction compared to no instruction). The discussion concludes with an analysis of the exploratory findings regarding the constellation of cues considered most important by jurors and whether or not juror perceived confidence in detecting deception shifts over time as a consequence of instructions.

11.1 Does Advice Impact Jurors’ Perceptions of Cues When Delay is a Factor?

11.1.1 Assessing Shifts from Baseline: Analyses within-groups.

Initially, the effect of instruction on jurors’ perceptions of nonverbal indicia were compared from baseline opinion where each respondents’ initial perception served as their own control. The findings are discussed below: first for perceptions of reliable indicia followed by unreliable indicia.

Hypothesis 1: Jurors’ perceptions of deception cues will not improve with advice over time.

Partial support was found for hypothesis three that, predicted jurors’ perceptions of nonverbal deception indicia would not shift with advice when delay was a factor. Mixed results
emerged: the current study found that, over time, instructions impacted opinion of reliable indicia but not unreliable indicia.

Despite delay, jurors perceived reliable cues to be more indicative of deception after they received instructions from an expert witness on diagnostics of deception. The finding indicates that not only do perceptions shift when advice is provided by an expert, but more so, that shifts in perception may hold over time. The effect of expert evidence was large \((d = .90)\) and the shift in opinion remained centred around the mean compared to the other sources of advice. The reduced spread of responses suggests that over time, the change in opinion that occurred with expert evidence was more similar among jurors than were shifts in opinion observed with other types of advice.

Aside from the testimony of the expert, no other form of instruction significantly impacted jurors’ perceptions of reliable indicia. While the pattern of results reflected general improvements over time (i.e. increases in mean differences), with the remaining types of advice, the increases were trivial as evidenced by negligible effect sizes (expert/judge: \(d = .03\); oral judicial direction: \(d = .05\); and written judicial direction: \(d = .16\)). Equally, while the mean increase for written judicial directions was three-fold that found for oral directions and expert/judge, the difference was not meaningful.

Variability between baseline perceptions were observed. Although non-significant, the difference between the largest and smallest baseline perceptions of reliable cues (i.e. expert witness: 3.46 and written instruction: 3.21) were similar to that found for the effect of advice by expert witness (i.e. .25 versus .22, respectively). Despite the baseline irregularities, ceiling effects are unlikely to have influenced the impact of advice for judicial directions (both written and verbal) or expert/judge combined as the scores obtained for expert witness and no instruction (the control) demonstrated larger increases in perception from relatively consistent baselines (i.e. expert: 3.46; expert/judge: 3.45; no advice: 3.44).
Regarding the control, jurors exposed to no instruction perceived reliable cues to be more indicative of deception over time than at baseline. Although the increase was not statistically significant, time alone appeared to have a moderate effect ($d = .46$) on enhancing perceptions. That is, the effect of measuring perceptions when delay is a factor indicates that opinions of reliable indicia improve on their own to some extent, without advice.

Non-significant shifts in perceptions of unreliable indicia from baseline were found for all types of instruction indicating that when jurors considered advice after a one-week delay, instructions were unsuccessful in correcting misconceptions. Aside from the anomaly found with written judicial directions, the remaining sources of advice demonstrated mean decreases in perceptions of unreliable indicia over time.

The impact of written and oral judicial directions was negligible ($d = .04$ and .08, respectively). While the pattern found for written directions was at odds with the other types of advice (the mean increase suggests unreliable indicia were perceived to be more indicative of deception following written instructions), the increase lacked any practical meaning – not only was the shift non-significant but more particularly, the increase was .02. The negligible changes observed for written and oral directions may have been undermined by their respective baseline scores. Both conditions had the lowest baseline perceptions of unreliable cues which may have impeded the opportunity for instructions to reduce perceptions further. What remains unclear is whether or not there are maximum shifts (i.e. reductions) that can be realised when measuring changes in perception. In the case of written and oral judicial directions, it is possible that perceptions could not be contracted more from already, relatively low baselines. For example, initial perceptions for judicial directions (3.11 and 3.14, respectively) only differed in the order of .04 and .07 from the lowest mean perception (3.07) achieved following expert/judge advice.

When jurors received no instruction, perceptions of unreliable cues increased with delay. This finding suggests that jurors perceived false cues to be more diagnostic of deception with the
passage of time when advice was omitted. The difference from baseline perceptions was marginally significant and the effect of time appeared to have a moderate impact \((d = .44)\) on shifting perceptions. The level of significance and size of effect reflect the largest mean difference obtained from baseline notably, an increase, together with the smallest dispersion of scores around the mean. Unexpectedly, the pattern for the control for perceptions of unreliable cues diverged from the pattern found for type of advice. Specifically, the directional trends were opposing indicating that that over time, perceptions of unreliable indicia begin to contract with instructions, but when instructions were not provided, jurors attributed more importance to unreliable indicia perceiving false cues to be more indicative of deception over time.

Baseline comparisons between perceptions of reliable and unreliable indicia were assessed as equivalent across two of the four types of advice suggesting that jurors may perceive reliable and unreliable cues to be of relative importance in detecting deception prior to receiving instructions. Although written and verbal judicial instructions differed at baseline as determined by one-tailed tests \((p = .05)\), the differences in opinion magnified with advice. Following instruction analysis suggest that jurors perceived reliable cues to be significantly more important in detecting deception than unreliable diagnostics with the effect of instruction found for all types of advice.

The largest split in opinion between perceptions of reliable and unreliable indicia was observed when jurors received advice by expert witness (+.43) and by expert/judge combined (+.41) with medium effects achieved for both \((d = .70\) and .55, respectively). After jurors received advice by written instruction the difference in perceptions was significant \((p = .05)\) however, the effect detected (a small effect: \(d = .35\)) was likely powered by the larger sample size: not only was the difference in perceptions post-advice the smallest of all instructions (including no advice), the spread of scores was equivalent to the control (that was non-significant), and the size of sample
was the largest. Despite the statistical significance, the practical significance for written instructions was limited: a larger difference was found for jurors who received no instruction.

**Summary of Within-Group Findings**

The findings for the effect of instructions on opinions of reliable and unreliable cues were mixed demonstrating partial support for hypothesis three that, predicted advice would not impact jurors’ perceptions over time. On the one hand, perceptions of reliable indicia were positively impacted by instructions despite the passage of time, whereas the provision of advice did not significantly correct misconceptions of unreliable indicia.

Only one source of advice was effective in shifting jurors’ perceptions when instructions were considered over time: only through the provision of an expert’s testimony did jurors perceptions of reliable indicia improve. This finding suggests that opinions of reliable indicia may be enhanced when jurors are educated on diagnostics of deception but whether or not the provision of advice has long-lasting effects appears unlikely with enhancements achieved with advice dissipating over time.

The within-group analyses revealed that in the absence of advice, jurors’ perceptions of deception indicia improve with time alone. This was the case for perceptions of both reliable and unreliable factors of deception. Therefore, the effect of instructions on jurors’ opinions of reliable cues needs to be interpreted with caution: the improvements observed with advice – significant or not – may be due to time, type of instruction (i.e. expert witness), an interaction of both, or other extraneous factor(s). Furthermore, while the pattern of improvement for perceptions of unreliable cues observed with instructions counters that found for no instruction, the difference was not significant and cannot therefore, be explained by the effect of advice alone. What is more, the pattern observed for no instruction warrants some concern: jurors’ false beliefs of nonverbal cues appear to increase over time when instructions are not provided.
11.1.2 Assessing Differences Between Advice and No Advice: Analyses between-groups.

Attempts to mitigate the limitations associated with comparing shifts in opinion from baseline were the focus of the between-groups analyses. The findings for the between-group comparisons follow.

_Hypothesis 1 Continued: Jurors' perceptions of deception cues will not improve with advice over time._

When shifts in perceptions of reliable cues were compared between jurors who received instructions and those who did not, no form of advice was successful in influencing opinion beyond that achieved with no instruction. The order of difference between the effect of expert evidence (that was significant when compared from baseline) and no advice in improving opinions of reliable cues was negligible ($d = .15$) as was evidenced by a minimal mean difference in opinion between conditions. Specifically, the mean increase in opinions of reliable cues following expert evidence (.22) and no advice (.17) were similar, so much so, that the shift demonstrated by expert witness may be largely accounted for by time over and above any effect of instruction. If in fact expert advice did have any influence on reliable perceptions – it was no more effective than no advice.

The small effects found for judicial directives (written instructions $d = .19$ and oral instructions $d = .29$), and expert/judge combined ($d = .24$) appear to be due to greater discrepancies in means between no advice and type of advice. However, these small effects represent differences of limited practical significance (.02 to .08). Based on the pattern found for reliable cues when jurors received no instruction, the smaller mean increases obtained for these types of advice suggest, that instructions may have moderated perceptions of reliable indicia – especially so for judicial direction and expert/judge combined where the mean increases were as minute as .02.

The moderated perceptions found for reliable indicia when jurors were exposed to judicial directions (written and verbal) and expert/judge instructions may reflect a suspiciousness of advice
given. That is, the nature of the instructions may have warned jurors as to the difficulties faced when evaluating demeanour evidence. It is possible that the instruction tempered jurors’ expectations and sensitised them to the challenges associated with reliably interpreting behavioural cues. Thus, resulting in jurors perceiving reliable cues to be less indicative of deception when advice is considered over time. Other studies have found similar outcomes when jurors receive instructions by way of judicial directions or expert evidence. Coined the “scepticism effect” (Cutler et al., 1990), research has found that when advice cautions jurors to the limitations of scientific knowledge instructions can result in scepticism of the evidence impacting its effectiveness on decision-making (Jones et al., 2017; Ribbers & Henneberg, 2018). Herein, a general sense of scepticism may have been a consequence of giving jurors time to reflect on the advice without providing a requisite understanding/knowledge of the factors that influence the reliability of nonverbal indicia when detecting deception. Compared to the evidence presented by the expert who offered more context and explanation as to factors that influence accuracy (i.e. having a baseline knowledge of a witness’s behaviour) jurors exposed to judicial directions did not receive this. Given that the impact of the expert witness differed when it was combined with a judicial direction, it is hypothesised that when advice was collapsed the factors referred to by the expert were lost with the passage of time with a recency effect resulting in the judicial direction having a lasting impact overriding jurors’ recollection of the evidence presented by the expert.

Although both the written and oral judicial directions were short interventions, the duration of the evidence presented unlikely accounts for the total lack of effect found. The combined advice by way of expert and judicial direction was the longest of all forms of instruction presented and yet it had no impact. As noted, it is possible that the influence of the expert in the expert/judge condition was lost over time due to a recency effect: perhaps the negative effect of the direction stayed with people over time, countering any influence of the expert.
Regarding jurors’ opinions of false cues, when shifts in perception were compared between those who received advice and those who did not, some instructions influenced perceptions of unreliable indicia. Jurors’ misconceptions of false cues were corrected over time when advice was delivered by an expert witness or when instructions were presented collectively by an expert/judge. Medium effects were found for both (expert witness $d=.56$ and $d=.65$ for expert/judge) with neither type of advice more or less effective in modifying false beliefs of nonverbal cues of deception.

Consistent with shifts from baseline identified in the within-group analyses, written and oral judicial directions did not influence jurors’ misconceptions of false cues over and above that achieved by time alone. Similarly, due to the degree of variability observed within the control, instructions appeared to have little impact on discriminating jurors’ opinion between reliable and unreliable cues. Although there were differences in opinions over time of reliable and unreliable indicia, the discrepancies in opinion were no more or less than those found for no advice.

**Summary of Between-Group Findings**

Mixed findings resulted in partial support for hypothesis three that, predicted shifts in jurors’ perceptions of nonverbal indicia would not hold over time despite receipt of advice. On the one hand, providing jurors with information on nonverbal diagnostics of deception did not enhance opinions of factors reliably associated with deception more so that omitting information. Namely, aside from written directions, the pattern of improvement found with instructions mirrored the effect of no instruction: with the passage of time, jurors perceived reliable cues to be more indicative of deception irrespective of advice. This finding suggest caution when interpreting the effect of advice when delay is a factor. That is, gains observed (albeit non-significant) in opinion of reliable cues may be due to time alone. Conversely, when jurors were cautioned of which cues are unreliable indicators of deception, the effect of advice was more successful than no instruction;
jurors’ false beliefs corrected – with equal success – by expert evidence and when combined with a judge’s verbal direction.

11.2 Overall Conclusions

11.2.1 Perceptions of Deception Over Time When No Instruction is Given

The pattern of mean increases observed among jurors who received no advice was consistent across cue-type: the differences from time one to time two were marginally significant and the magnitudes of difference almost reached medium effects ($d = .46$ and $.44$ for reliable and unreliable cues respectively). Notably, the difference in perceptions of reliable and unreliable cues remained constant over time suggesting that the effect of time alone may have influenced perceptions and that this effect was equal across both types of cues.

No predictions were specified for the control. However, the findings herein infer methodologically that delay, as a factor, may have had an inadvertent effect on cues perceived indicative of deception. Differences, albeit marginal, at time one and time two between reliable and unreliable cues considered indicative of deception were not expected. To the best of the author’s knowledge, baseline perceptions between reliable and unreliable cues have not been explored previously: that is, we don’t know whether initial perceptions of reliable/unreliable indicators are weighted differently. Although existing research has established that false beliefs are commonplace (DePaulo et al., 1997; 2003; Porter & ten Brinke, 2009), the focus of the literature has been on what the stereotypes are and their prevalence (Bond & DePaulo, 2006) especially as they relate to undermining veracity judgements in detecting deception. However, deception researchers have not presented the data in such a way that determinations can be made as to whether or not the prevalence of invalid and valid beliefs are equitable. Notwithstanding the differences found between cue-type at baseline, it is unknown whether a larger sample may have seen the pattern converge more closely at pre-intervention like that found in Study one.
Regarding the direction of the control, it is hypothesised that in the intervening time (that was the delay) participants exercised the opportunity to reflect on the question asked as to what factors indicate deception. On the one hand, this may have afforded participants the opportunity to rehearse the cues nominated at time one. However, rehearsal would have been difficult: there were 31 cues assessed of which 6 ratings were possible resulting in 186 potential combinations. Alternatively, the increase in opinions over time may be due to participants reflecting on the cues they thought were indicative of deception, rightly or wrongly, and these perceptions solidified in the absence of advice and became more important and embedded over the course of the week. Prior to the study, it is unlikely that participants had given much thought, if any, to what nonverbal cues are/are not indicative of deception, and that the research prompted them to reflect on and give consideration to those cues that they had initially identified. Some confidence is gleaned from the fact that not one participant in the control (or experimental groups for that matter) had received training in detecting deception.

The control highlights an important issue for consideration: it is contended that the results of the control alone indicate that when jurors are encouraged to consider indicators of deception, as routinely happens in court when they are directed to consider demeanour evidence, the omission of advice as to reliable/unreliable indictors of deception may in fact have a deleterious, inverse effect on jurors’ perception of nonverbal cues of deception. That is, without evidence to counter false beliefs, jurors’ misconceptions may not only prevail but magnify over the course of a trial, presenting egregious ramifications for jury decision-making.

The current study highlights the importance of including a control. Drawing conclusions based solely on in-built controls (as demonstrated by shifts from baseline) could be misleading. The findings derived from the control show that a degree of change in opinion occurs irrespective of advice. What remains unknown however, is to what degree perceptions shift as a result of other confounding factors not accounted for by the questionnaire, or time. To better understand the shifts
in opinion associated with no instruction requires further investigation. Future research could consider examining the impact of a control that employs a distraction, placebo task to enhance ecological validity while also ameliorating any confounding factors associated with participants predicting the purpose of the study.

11.2.2 The Effect of Instruction on Perceptions of Deception Over Time

When the influence of advice was measured in shifts from baseline the effect of expert evidence on enhancing perceptions of reliable cues appeared promising. However, when the effect of instructions was isolated, and potential retest effects were controlled, the influence of the expert’s advice on enhancing jurors’ perceptions of reliable indicia dissolved. Although expert evidence demonstrated improvements as indicated by the pattern of results, the shifts in opinion did not differ significantly from one another with the shift observed for expert witness no more than that achieved with time alone.

An inverse relationship was observed for jurors’ perceptions of false, unreliable cues. When shifts in opinion were measured from baseline, no source of advice impacted opinions. While the pattern demonstrated improvements by way of mean decreases with instruction, the shifts were non-significant and the order of change was, at best, small (e.g. expert witness). Conversely, when the effect of instruction was isolated from any potential consequences associated with the quick succession of the questionnaire, shifts in perception between those who received advice and those who did not, emerged. Providing jurors with advice in the form of an expert witness or expert/judge appeared to correct misconceptions regarding unreliable indicia of deception with medium effect sizes obtained. Interestingly, neither the expert alone nor combined as expert/judge was any better or worse at correcting misconceptions.

The current study suggests that jurors’ perceptions of reliable indicia do not appear to change over time, but their false beliefs may. The influence of expert advice on perceptions of reliable cues did not carry over when considered against jurors who received no advice, whereas,
the influence of expert evidence on perceptions of unreliable cues strengthened along with expert/judge advice combined. The variability in the effect of instruction on opinions is shaped by the pattern of results found for the control. Opinions of reliable indicia were not impacted by advice because they improved on their own account with time. Similarly, opinions of unreliable indicia were impacted by advice because the decrease in opinions found over time (i.e. improvements) diverged from the effect of time. The variability within the control extended to a lack of difference between opinions of reliable and unreliable indicia – while there was inherent discrepancy between cues perceived as reliable and unreliable indicators of deception found for advice, the differences could not be differentiated from the effect of time alone.

The results herein suggest that the effect of advice on changing jurors’ perceptions over time, was somewhat limited. On the one hand, it is possible that jurors were cautioned by advice (Bromby et al., 2007; Ribbers & Henneberg, 2018). Meaning, the stability in perceptions of reliable cues together with the corrections made to perceptions of unreliable indicia may indicate that, over time, advice attenuated jurors to the problems associated with assessing nonverbal cues. Indeed, it is possible that with time, the effect of advice moderated perceptions to the extent that reliable indicia were not perceived to be any more reliable than originally believed (which was already above mid-point), but also, that advice sensitised jurors to the difficulties associated with interpreting demeanour. This proposition goes some way to explaining why opinions of reliable cues did not improve but perceptions of unreliable cues did. It is hypothesised therefore, that in the intervening week, advice tempered perceptions of reliable indicia along with correcting false beliefs.

An alternative explanation however, is that pre-existing perceptions of nonverbal indicia over-ride advice with time. That is, over the course of a one-week period, the influence of advice on reformed perceptions depreciates, and what were once initial impressions re-materialise. This explanation is consistent with the extant literature that attests beliefs about deceptive/truthful
behaviour are so deeply embedded that they resurface without provocation and irrespective of evidence that conflicts with opinion (Anderson et al., 1998; DePaulo, 1994; Feeley & Young, 2000; Frank & Feeley, 2003; Porter & ten Brinke, 2009, 2010; Porter et al., 2013; Zuckerman, Koestner, & Colella, 1985).

Notwithstanding replication, the absence of advice appears to be more damaging to perceptions than introducing the evidence of an expert witness or expert/judge combined. While the testimony of an expert and collective instruction of an expert/judge may only moderate perceptions of reliable cues they too correct misconceptions of false cues providing perhaps, a helpful adjunct to safeguarding juror-decision making. Disconcertingly, the very absence of instructions may pose harmful consequences when witness credibility is at issue. However, this supposition is just that, and cannot be extrapolated without further investigation.

11.3 What Combination of Nonverbal Deception Cues do Jurors Considered Most Important?

The current study found little evidence to suggest advice was effective at shifting individual cues perceived most important in detecting deception when delay was a factor. Specifically, the findings revealed little change in the constellation of cues perceived most important. Over time, mock jurors maintained that confidence, spontaneous corrections, implausible responses and responses that contain a lot of detail were the most important cues in detecting deception and that only the ranking order of these cues adjusted. That said, the most important cue perceived indicative of deception (confidence) prior to advice was also maintained over time with its overall level of importance increasing with advice.

While the split of reliable and unreliable cues ranked the most important did not change over time, no consistent pattern emerged for reliable and unreliable cues with both increasing and decreasing in response to advice and time. What did emerge was that at the factor level, the results provide support for existing research that suggests stereotypes about deceptive behaviour may be so entrenched that they remerge without temptation (Frank & Feeley, 2003). The stability
demonstrated among the top five most important cues goes one step further and provides support for Dangerous Decisions Theory (Porter & ten Brinke, 2009) that predicts perceptions not only prevail over time but do so in the face of conflicting evidence.

At the factor level, a more detailed picture as to the effect of advice can be seen when the percentage change is analysed. Most notably, with delayed advice, although the combination of top five cues ranked most important does not shift, direct improvements can be seen on high-pitch voice that increased three-fold, body/hand/finger/leg/foot movements doubled and responses that contain very few details increased 150%. That is, of all percentage changes both positive and negative, these three reliable factors represented on the whole the biggest percentage changes found.

Given the consistency in combination of cues perceived most important as well in the percentage changes achieved with advice it might be reasonable to infer that the findings may reflect response bias as a consequence of the questionnaire design. However, the order of factors listed in the survey do not explain the combination of results found and so response bias is unlikely. The only corresponding pattern that could be inferred was that the first item listed on the question that asked participants to rank cues from most important to least important was confidence. This may account for why confidence was the predominant cue of choice at both baseline and post-test. Furthermore, its order of ranking in the factors listed may have unintentionally had a primacy effect attracting more attention compared to those factors in the middle of the survey. This eventuality could be controlled for by randomly ordering the factors tested should future research wish to investigate the combination of cues perceived most indicative of deception. For the three reliable cues identified across both studies, it is possible the items themselves conveyed high face validity and where chosen accordingly.

Effect of advice aside, the nonverbal cues identified reveal important insight as to the combination of cues considered most important by mock jurors in detecting deception. These being
spontaneous corrections, implausible responses, responses that lack logical structure or responses that contain very few details (after advice). Distinct from the cues that demonstrated the largest percentage increases (high-pitch voice and body/hand/finger/leg/foot movements), the combination of cues perceived most important provide jurors with a greater opportunity to reliably evaluate the cues identified because they do not require a reference point or baseline knowledge from which the indicia needs to be compared. By way of example a witness’ statement either contains spontaneous corrections or it does not, it appears implausible or does it not, and/or a statement lacks logical structure, or does not. Although high-pitch and body movements are among the more reliable indicia of deception, accurately assessing these nonverbal cues are largely dependent on having a baseline knowledge of the witness’ behaviour from which deviations can be evaluated.

The combination of reliable cues identified as most important herein offer promise for future research. Should the effects of advice be investigated further it would be of interest to test the impact of advice on jurors’ ability to accurately detect deception if advice directly targeted the combination of cues perceived most important. As found by Hartwig and Bond (2014), accuracy in detecting deception is enhanced when a constellation of cues are evaluated. That said, a small number of studies have found large discrepancies between the cues people identify as important in detecting deception and the cues they actually rely on when trying to detect deception (Hartwig & Granhag, 2014; Bogaard & Meijer, 2018). Should advice concentrate on the handful of cues identified in current research, the ability to enhance reliable cues already considered important by mock juror may be more meaningful in the context of interpreting demeanour, assessing witness truthfulness and therefore overall credibility.

11.4 Does Advice Influence Jurors’ Confidence Over Time?

The effect of advice did not appear to influence perceived confidence when delay was a factor, with no significant shifts from baseline observed by source of advice. This finding is inconsistent with Willis & Todorov (2006) who found confidence in baseline perceptions of
trustworthiness increased with time. Although the shifts found herein were non-significant, it is noted that the largest mean increase (M difference = .31) in perceived confidence, despite delay, was obtained for expert evidence; the one condition of advice that successfully impacted jurors’ perceptions of reliable cues when assessed from baseline (as per the within-group analyses). The magnitude of change in confidence observed for expert evidence, although non-significant, represented a medium effect (rounded, $d = .48$) with the increase more than three-fold that of the next largest, oral judicial directions (M difference = .09), and ten times that of written judicial directions (M difference = -.03). Some of the difference between shifts in confidence may be accounted for by the lower baseline perceptions reported by jurors assigned to the expert witness condition (M = 3.38) compared to those who received oral or written directions (M = 3.50, respectively). That said, expert/judge advice had the lowest baseline confidence (M = 3.25) of all sources of advice, and yet, confidence remained stable over time, despite advice. By contrast, it is possible that the size of the effects of advice on shifting perceptions were not large enough to translate into shifts in perceived confidence. Namely, the largest shift in perceived confidence obtained by expert evidence also achieved a large effect ($d = .90$) on shifting perceptions. Whereas, the marginal shifts in perceived confidence correspond with either less than small effects on shifting perceptions (written directions: $d = .16$), or negligible effects: verbal directions, $d = .05$; expert/judge, $d = .03$. Of course, the simplest and most likely of explanations is that the provision of advice is not related to perceived confidence in detecting deception, and therefore, was not prejudiced by advice.

11.5 Chapter Summary

The current study investigated whether jurors’ consideration of advice over time impacted perceptions of nonverbal cues considered important in determining the truthfulness of a witness in court. Partial support was found for hypothesis three, that predicated jurors’ opinions of deception indicia would not shift over time, despite advice. The evidence was mixed depending on cue-type
but more so, was dependent on whether or not the effect of instructions was compared with no instruction. The current design highlighted the importance of including an independent control to negate, where possible, potential confounding factors such as re-test effects and the influence of time alone.

The control suggested that delay was a factor in its own right. The pattern of mean increases observed within the control for both cue-types from pre-to post-delay present an interesting finding. Not only did jurors’ perceptions of reliable cues increase with time, but so too did their perceptions of unreliable cues. Although the increase in opinion of reliable cues reflects an improvement over time, the increase in importance attributed to unreliable cues implied a worsening effect – with false, misleading cues possibly perceived to be more important in detecting deception over time.

Added to this, two observations were made. First, that the order of difference between reliable and unreliable cues remained constant over time with reliable cues remaining more important in detecting deception than unreliable cues, despite a weeks’ delay. Second, that both cue-types increased over time reflecting that jurors perceived both reliable and unreliable cues to be more indicative of deception after a one week period. This was despite no advice.

The current study revealed limited support for the benefit of presenting advice to jurors in an effort of improve perceptions of reliable cues over time. The effect of instructions revealed that when jurors’ perceptions of nonverbal indicia were compared between those who received advice and those who did not, opinions of reliable cues did not differentiate: the pattern of improvement observed with instructions was no more influential than the effect of no advice. This conclusion is largely due to the consistent pattern of improvement found in perceptions of reliable cues achieved with instruction and with time alone. Although the largest improvement in perceptions of reliable cues was achieved by expert witness it was no more significant than the effect of no instruction. This finding indicates that any education redressing misconceptions among jurors – be it in the form of expert testimony, judicial directives or a combination of both, may be dissolved by the
effects of decay and interference. As such, the significant improvement in jurors’ opinions of reliable cues achieved from baseline with advice by an expert witness remains uncertain: the shift in perceptions cannot be solely accounted for by advice.

Counter to the findings for perceptions of reliable cues, the current study demonstrated partial support for the effect of advice on correcting misconceptions when delay was a factor. While misconceptions of false cues amplified over time when no instruction was given, they appeared corrected when jurors considered advice over time. Only expert evidence and expert/judge systematically impacted false beliefs of deception when delay was a factor with both types of advice equally effective in shifting perceptions. That is, the effects of decay and interference did not appear to undermine the persistence of education when information was proffered as expert evidence or presented collectively with a judicial direction.

Whether or not ceiling effects inhibit the degree to which perceptions of reliable cues changes and/or floor effects impeded the extent to which perceptions of unreliable cues improved remains uncertain. On the one hand the findings herein might suggest that misconceptions are less resistant to change when faced with conflicting evidence than are opinions that are consistent with and supported by the evidence. However, this is counter to what we know about confirmation bias. Rather, the findings of the current study may indicate that, with the passage of time, instructions sensitised jurors to the limitations inherent in relying on nonverbal cues; perceptions of reliable cues didn’t significantly enhance whereas perceptions of unreliable cues contracted with advice because the instruction served as a caution as to the difficulties faced in reliably assessing the demeanour of a witness. Certainly, whether or not the effect of advice on perceptions was counteracted by the effect of time alone and/ or undermined by the persistence of perceptions cannot be resolved entirely. What can be said, however, is that given the limited effect of instructions on shifting perceptions, it would appear more likely that initial beliefs reinstate when time is a factor, thereby offsetting any benefits achieved with instructions.
As a consequence, the findings herein demonstrated mixed support for the theory of Dangerous Decision (Porter & ten Brinke, 2009). While it is not discounted that perceptions were impacted with advice (at least, false beliefs were corrected by way of expert evidence and expert/judge instructions), the changes were in the minority, and were not uniform across cue-type or advice-type. Consistent with the prevailing literature, the current findings suggest that per-existing perceptions may be enduring (Blanch-Hartigan et al., 2012; Frank & Feely, 2003; Hauch et al, 2016; Shaw et al., 2013), and that while advice might shift some perceptions, its influence is limited with perceptions predominantly reverting back to baseline.

Moreover, the results of the control highlight that in the absence of conflicting evidence, opinions as to deceptive behaviour not only prevailed over time but in fact became stronger with time when advice was not present. That is, jurors in the control were asked to consider their opinions as to cues indicative of deception, rightly or wrongly. When these participants did not hear any evidence or instruction to countervail their initial perceptions these opinions solidified and were perceived to be more indicative over time. The increase in opinions of unreliable cues over time without advice highlights some cause for concern – if jurors are directed to consider demeanour but no advice is forthcoming, jurors may in fact consider false cues to be more indicative of deception over time relying on these erroneously when determining the credibility of a witness. As the control suggests – without knowledge to counter ill-informed opinions, jurors are not just at risk of maintaining misconceptions, but rather, these misleading beliefs may exacerbate over time when advice is not omitted. Notwithstanding replication, this finding infers serious implications for legal practice. If just outcomes are to be protected the current study presents some precursory evidence to suggest that educating jurors about nonverbal cues of deception through expert testimony and expert/judge advice may be a useful adjunct to safeguard decision-making.
Chapter 12. General Discussion

This thesis aimed to test components of Dangerous Decisions Theory (DDT; Porter & ten Brinke, 2009) that posits initial judgements of nonverbal behaviour relied upon in determining the trustworthiness/credibility of a witness not only form rapidly but are so deeply ingrained and pervasive that they are resistant to change even when presented with contradicting evidence. To test the fundamental premise of DDT three broad aims were examined through two studies.

The first study aimed to establish whether the provision of advice in the form of expert evidence and/or judicial directives, improved the nonverbal cues perceived indicative of deception by jurors when determining the truthfulness/deception of a witness in court. The primary objective of Study one was to determine whether or not the findings of previous research could be replicated with the aim to confirm/disconfirm whether jurors’ perceptions of reliable indicia could be enhanced through the provision of advice. The second aim of this thesis sought to establish whether or not juror misconceptions of unreliable indicia of deception could be corrected with advice.

It has been well established in the literature that stereotypes of deceptive behaviour are wrong and that these false beliefs undermine the accuracy of assessing deception (Bond & DePaulo, 2006; Levine & McCornack, 2014; Vrij et al., 2019). Prior to this thesis, the effects of advice on shifting opinions of false, unreliable cues (i.e. misconceptions) had not been investigated. While Coyle and Thomson (2014) concluded that the provision of expert and/or judicial advice “can correct jurors’ well-entrenched misconceptions of which behavioural indicia are indicative of deception” (p.10), they demonstrated that the impact of advice improved perceptions of reliable indicia of deception. Whether they addressed misconceptions or not, remains unknown. The current thesis begun to explore this gap in knowledge, examining to what extent false indicia are believed by jurors to indicate deception and tested the effect of advice on correcting misconceptions.
The second study extended on from Study one by investigating the effects of decay and interference on the provision of advice proffered to jurors. The unique objective of examining delay was to establish whether improvements in juror opinion achieved with instructions, as demonstrated by Study one, held over time. The delay between the introduction of evidence (i.e. the expert witness) and the subsequent consideration of that evidence are a reality of legal proceedings. It is not uncommon for the admission of evidence to take several days, or for evidence to be admitted earlier in a trial and for members of a jury to consider that evidence at some later stage. Equally, juries are often directed by judges at different periods during proceedings such as instructing juries on how they should weigh up evidence or when delivering their summations. Accordingly, the third aim of this thesis was to test whether the delayed consideration of instructions by way of expert evidence and/or judicial directions would impact jurors’ opinions of nonverbal cues of deception. The purpose was to establish whether or not the effects of advice as found in Study one, were transient with pre-existing perceptions over-riding the effects of advice or, whether changes in opinion were more enduring.

The current design included a dedicated control group over and above that of using participants as their own control. This allowed for between-group analyses in addition to within-group analyses to be undertaken whereby shifts in perceptions could be compared between jurors who received advice and those who did not. This facilitated a means by which the repetition of questions alone and the effect of time could be isolated from the effect of advice to see what effect, if any it had on perceptions at follow-up. Shifts in in opinion from baseline where participants were their own control were discussed in the proceeding discussions. However, the comparisons from baseline are not the focus of this chapter as the effect of advice could not be differentiated from potential retest effects or the effect of time alone. Rather, the findings as they relate to the effect of instruction compared to no instruction are discussed.
The structure of the current chapter follows herein: the findings of Study one and Study two are briefly summarised with respect to the three aims. Each aim is presented in chronological order. To reiterate, aims one and two were explored in Study one and the third aim was investigated in Study two. A review of each aim concludes with the findings of both studies considered together. The exploratory analyses are discussed in turn and the implications, strengths, and limitations of the current thesis are examined. The chapter concludes with considerations for future research.

12.1. **Aim One: Does the Provision of Advice Influence Jurors’ Opinions of Reliable Indicia of Deception?**

Study one confirmed that advice may have the potential to successfully improve jurors’ opinions of which nonverbal cues reliably indicate deception. The study demonstrated that through the provision of expert evidence and the collective advice of an expert and a judge’s direction, jurors might benefit equally from both instructions taking onboard more reliable cues to better inform their interpretations of witness credibility. The findings partially supported previous research with the effect of expert evidence replicated but written judicial directions not (Coyle and Thomson, 2014). Differences in research design including experimental settings most likely account for the differences in effects obtained for written instructions. What remains unknow is whether or not the simple re-administration of the questionnaire had a retest effect on the improvement found by Coyle and Thomson (2014; Scharfen et al., 2018), albeit unintentionally, and if so, to what extent. Nevertheless, the finding herein is consistent with a wider body of research that shows written directions appear ineffective in improving juror decision-making (Baguley et al., 2017; Essex & Goodman-Delahunt, 2014). What is more, the difference in findings between the original study and this thesis is not dissimilar to the disagreement found in the literature (Cicchini & White, 2016; Ribbers & Henneberg, 2018; Spivak, Clough & Ogloff, 2018).
The negligible impact found for the influence of oral judicial directions was consistent with existing research that shows little benefit in improving juror comprehension with judicial directions (Blankenship et al., 1997; Baguley, et.al., 2017; English & Sales, 1997; Luginbuhl, 1992; Ogloff, et al., 2011; Ogloff & Rose, 2005; Rose & Ogloff, 2001; Severance & Loftus, 1982; Spivak, et al., 2018). An unexpected finding was the effect of no instruction. Although the shift in opinion was non-significant, the variability and negative relationship was of interest: jurors perceived reliable cues to be less indicative of deception from time one to time two when advice was withheld. Whether this variability reflects a decline in memory, doubt, or change in opinion could not be established. It is possible that the shifts were so small that they merely represent a level of inherent discrepancy that comes with asking jurors to consider an extensive list of nonverbal factors.

12.2. **Aim Two: Does the Provision of Advice Correct Jurors’ Misconceptions of Nonverbal Indicia of Deception?**

Extending the replication to examine unreliable cues independent of reliable cues was a unique contribution of this thesis. Study one demonstrated that instructions may successfully correct jurors’ misconceptions regarding false cues of deception. Specifically, Study one established that perceptions of unreliable cues can be improved when evidence is adduced by an expert witness with jurors placing less importance on unreliable indicia when determining the truthfulness/deception of a witness after receipt of advice. Aside from the testimony of the expert, no other type of advice influenced perceptions of unreliable cues more than could be achieved by no instruction.

Although written judicial directions and the collective advice of the expert/judge were no more effective in influencing opinion than the effect of no advice, the trend of results are noteworthy. That is, a pattern of improvement was observed (i.e. decreases from baseline) for jurors who received written directions or the combined advice of the expert/judge with perceptions of unreliable cues declining with advice albeit, not significant statistically. By way of comparisons,
only the advice of the oral judicial direction presented an anomaly: again, although non-significant, the shift opposed the direction anticipated with jurors perceiving unreliable indicia to be more indicative of deception with advice. In light of these findings, it is possible that the lack of effect found for expert/judge on correcting misconceptions may have been unduly influenced by the negative contribution of the judicial direction. While the effect of expert witness imparted similar shifts in opinion on both reliable and unreliable cues (i.e. +.36 and -.34, respectively), the shift in opinion on unreliable cues evidenced by expert/judge (-.16) may have achieved a similar impact to that found for reliable cues (+.26) had judicial direction not increased by .08 (i.e. -.16 - .08 = -.24). Nonetheless, the finding that only the instruction from an expert witness effectively corrected jurors’ misconceptions presents an original, albeit provisional finding. Similarly, the lack of influence on jurors’ perceptions of unreliable indicia observed for judicial directions (both written and oral), or when combined with an expert (expert/judge) present new, preliminary findings. Despite these findings, and as noted throughout, the variability in baseline perceptions across advice groups warrants caution: the conclusions drawn in this thesis can provide no more than preliminary indications of the effect of instruction on enhancing perceptions of detecting deception.

Nevertheless, an unexpected discovery was the significant impact found when jurors were asked to assess nonverbal cues of deception in the absence of evidence. While the control from Study one cannot be generalised outside replication, the results implied, somewhat concerningly, that at follow-up jurors’ perceived false cues to be more indicative of deception when advice was not received.

In sum, when the findings of aim one and two are taken together, the following conclusions can be made. First, only the advice of the expert witness appears to successfully influence jurors’ perceptions with improvements in opinions achieved across both reliable and unreliable cues. Second, the collective effect of expert/judge on enhancing perceptions of reliable indicia did not transfer to correcting perceptions of unreliable indicia. Third, inviting jurors to assess nonverbal
cues in the absence of advice on reliable/unreliable diagnostics of deception may, notwithstanding replication, contribute to erroneous determinations of witness credibility with jurors’ false beliefs amplifying in the absence of evidentiary safeguards. These findings substantiate earlier research and suggest that jurors’ perceptions of reliable diagnostics of deception may improve with instruction and provides some contrasting evidence (albeit, cautionary), that may counter the core principle of DDT. As demonstrated, advice may effect changes in perceptions of nonverbal cues of deception in the immediacy. The question as to whether or not the changes to pre-existing perceptions are transient, or more enduring, was the focus of aim three and was investigated in Study two.

12.3. **Aim three: To Determine Whether the Effect of Advice on Improving Jurors’ Perceptions Persist Over Time.**

Study two expanded on from Study one that indicated perceptions of nonverbal deception cues may shift with advice and investigated whether these changes in opinion held over time. Study two revealed mixed findings. Regarding reliable cues, jurors’ perceptions did not improve over time, despite advice. The findings indicated that when delay was a factor, not one type of advice was effective in enhancing opinions of reliable cues more than that achieved by time alone. Namely, over time, jurors’ perceptions of reliable indicia improved on their own, without advice. Although the shift within the control was non-significant, the changes in opinion observed across instructions did not differ enough from those found when jurors received no advice. Indeed, the “positive” shifts observed with advice by way of oral direction or combined with an expert were trivial whereas the largest mean shift was for jurors exposed to expert evidence. Despite the lack of statistical significance, the differences in magnitude of change between the mean increases were notable. What remains unresolved is whether or not these differences were due to the efficacy of advice, the design of the control, or an interaction of both. Additionally, whether or not some types of advice result in a scepticism effect of the information presented, or were more or less memorable
than others remains to be seen. All that can be said is that the effect of advice on enhancing perceptions of reliable indicia over time cannot be differentiated from the effect of time alone suggesting that for perceptions of reliable cues, at least, the effects of decay and interference impacted the efficacy of advice provided to jurors when delay was a factor.

Regarding unreliable cues, the effect of instructions on jurors’ opinions differed to that found for unreliable cues. When perceptions were compared between those who received advice and those who did not, half of the instructions corrected jurors’ false beliefs of which indicia are, and are not indicative of deception. Jurors’ misconceptions regarding deceptive behaviour were corrected over time with expert evidence and when instructions of an expert witness and member of the judiciary were presented together. Both types of advice achieved medium effects and were deemed equally effective in modifying false beliefs of nonverbal cues of deception.

Interestingly, neither type of judicial direction – be it written or verbal – impacted jurors’ false beliefs with perceptions remaining relatively unchanged over time irrespective of advice. The consistently negligible impact of judicial directions suggests that not only might evidentiary judicial instructions be ineffective in correcting opinion over time, but moreover, the presentation modality of the instruction appears immaterial. Indeed, if the lack of effect observed for judicial directions is not due to the content of the advice given, or the source from which it comes, then the most likely alternative is that the duration of the intervention was too short to have an discernible effect on perceptions.

The findings of Study two were unexpected. Based on the literature that (1) people’s beliefs of which nonverbal cues indicate deception are fundamentally flawed (DePaulo et al., 2006), and that (2) these false beliefs are so entrenched they not only persist over time (Frank & Feely, 2003), but are maintained in the face of conflicting evidence (Porter & ten Brinke, 2009), it was not expected that advice might influence jurors’ opinions of deception. Mixed results demonstrated partial support for hypothesis three that, predicted shifts in jurors’ perceptions of nonverbal indicia
of deception would not hold over time irrespective of advice. One the one hand no support was found for advice enhancing opinions of reliable cues over time, yet, some evidence was found for advice influencing opinions of unreliable cues over time.

Presenting jurors with information on the difficulties associated with detecting deception based on nonverbal behaviour may have posed a risk of undue caution (Bromby et al., 2007; Ribbers & Henneberg, 2018). The aim of adducing evidence is to furnish triers of fact with information as to reliable/unreliable indicators of deception to induce their sensitivity to cues that will and will not aid their assessments of credibility. However, drawing attention to the problems and pitfalls associated with nonverbal cues may have had an unintended consequence of inducing juror scepticism. Leverick (2014) defines juror scepticism as an overall mistrust of all evidence, even when caution is not warranted. Consequently, it is possible that with the passage of time that was the delay, jurors had more opportunity to reflect on the content of the advice given whereby the impact of advice resulted in a scepticism effect (McAuliff and Duckworth, 2014; Ribbers & Henneberg, 2018, p. 266) and may account for why the influence of advice not only contracted with time but diminished. Of course, the simplest, obvious alternative proposition can be gleaned from the deception literature (e.g. Blanch-Hartigan et al., 2012; Frank & Feely, 2003; Hauch et al, 2016; Shaw et al., 2013) that pre-existing, initial beliefs are so deeply-embedded and pervasive that they re-instate without effort, over-taking newly acquire knowledge on nonverbal cues of deception sooner or later.

In sum, the effect of providing information to jurors over time was mixed: it did not enhance opinions of factors reliably associated with deception more so that omitting the information, but it appeared to caution triers of fact of which cues are unreliable indicators of deception. The findings of the control and conclusions on the effect of advice extrapolated herein need to be considered cautiously. The gains observed (albeit non-significant) in opinion of reliable cues may be due to time alone, and not instruction. Furthermore, the degree of variability in
opinions observed for jurors who received no advice warrants attention. Whether or not the variability in perceptions is due to the omission of advice, limitations of the research design or another extraneous factor(s) remains unknown. Regarding the research design, it must be emphasised that the length of advice presented to participants is conflated with source of advice. Consequently, it is noted that gains in accuracy achieved with advice may have occurred as a consequence of the duration of the advice presented, the content itself, the source (i.e. people) presenting it, or a variation of the above. The point herein, is that these conditions can be manipulated with a cumulative effect. Nevertheless, until the directional pattern found for no advice is replicated in other studies using an independent, equivalent control, the generalisability of these findings need to be interpreted with care.

12.4. **Overarching Conclusions: Do Instructions to Jurors on Nonverbal Diagnostics of Deception Matter?**

Together, the findings from both studies suggest that jurors may benefit from instructions on reliable/unreliable diagnostics of deception. The current thesis provides preliminary support for the effect of advice on influencing jurors’ perceptions across both reliable and unreliable indicia. While the results were mixed depending on type of advice, type of cue, and time-to assessment (i.e. immediate versus delayed), some broad conclusions can be made. These are as follows.

Jurors’ perceptions of reliable cues may be enhanced with advice in the immediate term when that advice is presented by an expert witness or an expert and judge. On first glance, one might suspect that the accuracies gained in perceptions of reliable and unreliable cues of deception might be attributable to the duration of advice presented with the lengthier conditions of advice (i.e. expert witness and expert/judge) facilitating greater repetition of ideas and breadth of coverage. However, if the duration of advice alone was sufficient to effect change in opinions then jurors’ perceptions of unreliable cues ought to have improved when advice was presented by expert and
judge combined – which it did not. Accordingly, the evidence of this thesis indicates that the duration of advice alone is unlikely responsible for shifts in perceptions of cues of deception.

Further, one might infer that the improvements observed for those advice conditions that were greater in duration (i.e. expert witness and expert and judge combined), may be explained by the bigger delay between the times to assessment (i.e. pre-test to post-test), making jurors less likely to affix to their initial opinion. However, this proposition is undermined by the results demonstrated in Study two that found delay, as a factor alone, does not improve perceptions as a function of length of advice. If it did, then one would expect to have found improvements on reliable cues as well as unreliable cues for both advice by expert witness and expert/judge combined – which it did not. Conversely, the benefit of instructions by expert witness and expert and judge combined, appear to abate with the passage of time. Indeed, no type of advice appears effective in improving jurors’ knowledge of which cues are reliable diagnostics of deception when delay is a factor.

On the contrary, when jurors consider advice over time, the effect of advice on correcting misconceptions might be maintained when advice is proffered by the expert witness. Notably, when jurors considered advice immediately after it was provided, expert/judge instructions had no impact on correcting false beliefs. However, when the instruction was considered over time, the potential benefit of the expert/judge emerged – improving perceptions of unreliable cues of deception. Again, while it might be inferred that the bigger delay between times to assessment encouraged participants to reconsider and change their opinions this only occurred for false beliefs, not both.

A noteworthy finding was the impact, or lack thereof, for both written and verbal judicial directions: the influence of judicial instructions on jurors’ opinion of reliable indicia appeared negligible, irrespective of whether advice was considered immediately after it was adduced, or over time. Interestingly, this finding suggests that the mode of presentation (written versus oral) may be of little consequence. Whether or not the ineffectiveness of judicial directions was due to: (1)
the short duration of the interventions; (2) a perceived lack of credibility and/or importance attributed to or imparted by the advice; (3) a consequence of a general sense of scepticism of the information provided or because the factors were not adequately explained; or (4) simply because the instructions were less memorable, could not be determined.

Generally, when considered together, the impact of advice differs depending on type of advice tendered, type of cue assessed and the time-to-assessment. When looking at the effect of time-to-assessment (given that it was the distinguishing factor between studies) on influencing juror opinion, in the immediate term, instructions appeared to enhance perceptions of reliable indicia yet over time, the benefit of the instruction diminished. While the pattern of improvement for perceptions of reliable indicia improve with the passage of time, the effect of advice appeared to be no more than the effect of time alone. The inverse was true for the effect of advice on jurors’ perceptions of unreliable indicia. Namely, when jurors consider advice immediately after it is presented, only expert evidence successfully corrects misconceptions of cues not indicative of deception. However, with the passage of time, not only was the benefit of the information proffered by the expert witness maintained, but also, instructions delivered by the expert/judge emerge as effective in shifting jurors’ perceptions regarding unreliable indicia.

Together, these findings suggest that the effect of instructions weaken with time for perceptions of reliable indicia but broaden for perceptions of unreliable indicia. This was evidenced by the expert witness and expert/judge having a significant impact on enhancing perceptions of reliable indicia in the immediate term but when the information presented by the expert or expert and judge was considered over time, the merit of the instruction dissipated. This finding is further substantiated by the mean differences across advice types condensing, and reducing in magnitude with time. Whether the benefit of advice, including the expert specifically, diminished over time because the efficacy of the advice weakened, because jurors’ opinions increased due to time alone, or simply because initial impressions over-ride instructions at some point in time could not be
resolved. Equally, whether differences in perceptions were due to the effect of the source of the advice, the content of the advice, the length of the advice or an interaction of the above, cannot be determined. What can be said, is that the mean increases achieved with expert evidence were the largest across type of advice in both the immediate and long-term. The same cannot be said for the benefit of expert/judge advice: not only did the benefit of the instruction diminish over time but the shift in perceptions were negligible when delay was a factor with the mean increase among the smallest of all advice presented.

This thesis highlights the importance of incorporating independent controls rather than drawing conclusions from in-built controls where the latter may overlook possible confounding factors such as retest effects. Regarding the inconsistencies between the effect of advice in the immediate and longer-term on influencing perceptions on reliable and unreliable cues, the discrepancies between the controls cannot be discounted. The opinions of jurors changed when instructions were withheld both in the immediate term and over time but the direction of these changes were contrariwise to one another. By way of explanation, for Study one, perceptions on reliable and unreliable cues decreased from baseline for those who did not received advice. Those who received advice in the immediate term demonstrated improvements on reliable cues (i.e. increases in mean scores) and improvements on unreliable cues (i.e. decreases in means scores). As such, the differences found on unreliable cues between those exposed to advice and those not, were not dissimilar enough to suggest immediate advice had a greater effect. A similar divergence occurred when advice was delayed, albeit the direction was reversed. For Study two, perceptions on reliable and unreliable cues increased from baseline with time for those who received no advice. Equivalent in pattern to those exposed to immediate advice, opinions over time demonstrated improvements on both reliable and unreliable factors with increases and decreases in mean scores achieved respectively. Therefore, the differences found between those exposed to advice and those not for reliable cues were not distinct enough to suggest perceptions were influenced by advice more than time alone.
To summarise, in light of the controls incorporated herein, the findings indicate that the effect of no advice influenced jurors’ opinions differently in the short-term compared to the longer-term. When jurors were asked to consider which nonverbal cues were indicative of deception without instructional advice, in the immediate term, jurors perceived reliable cues to be less indicative of deception. By contrast, when jurors had time to reflect on the task at hand, their opinions strengthened, with reliable indicia believed to be more indicative of deception than at baseline. The same patterns were observed for perceptions of unreliable indicia. In the short-term, jurors perceived unreliable cues to be less indicative of deception in the absence of advice but over time, false beliefs exacerbated in the absence of instructions. Consequently, the effect of omitting instructions effected jurors’ perceptions differently depending on when they were asked to evaluate nonverbal cues. Whether the variability in effects of no advice were due to differences in sample size ($n = 53$ in Study one; $n = 18$ in Study two), due to differences in time-to-assessment, or both cannot be said.

Regarding reliable cues, both studies point to the benefit of instructing jurors on diagnostics of deception in the immediate term – not only do jurors’ perceptions appear to improve from the information proffered but doing so counteracts possible negative consequences that may be associated with neglecting to instruct jurors on nonverbal cues of deception. Maximising the benefit of instructions is not conditional on the type of advice introduced, with both expert evidence and expert/judge equally effective in enhancing jurors’ perceptions in the immediate term. Practically however, there appears to be little advantage in adducing both the evidence of the expert and direction of the judge if the testimony of the expert is sufficient alone to assist juror decision-making.

Regarding jurors’ perceptions of unreliable cues the findings, taken together, indicate that the effect of expert evidence successfully corrects jurors’ misconceptions in both the immediate term and longer-term. Although instructions by expert/judge were also found to improve opinions
of unreliable indicia over time it is argued that it does not add any value to that achieved by expert witness. Firstly, expert/judge is as effective as the expert alone in correcting for misconceptions when delay is a factor; second expert/judge does not correct for misconceptions in the immediate term whereas the expert does, and third, pragmatically presenting jurors with two independent forms of advice is unnecessary when the same impact may be achieved as effectively by one form of advice.

When considered together, the current thesis demonstrates that a number of factors influence jurors’ perceptions of nonverbal diagnostics of deception. These include whether perceptions pertain to reliable or unreliable indicia. Shifts in perception also depend on the source of advice from which instructions are given and whether the instruction are considered immediately or over time. While there is clearly an interaction between the three factors investigated (cue-type, advice, and time-to-assessment) the nature and extent of the interaction cannot be extrapolated without replication and should be treated with caution.

Based on the findings of this thesis, what can be concluded is that perceptions of indicia of deception may shift, in both the immediate and longer term. However, the extent to which opinions are influenced vary between reliable indicators and unreliable indicators and are impacted differently by the method in which advice is delivered. There appears to be some advantage in introducing expert evidence as an adjunct to the court to educate jurors, bettering perceptions of diagnostics of deception. Not only may the testimony of an expert witness enhance opinions of reliable indicia while at the same time reducing false beliefs of cues not indicative of deception, some improvements achieved with expert evidence may preserve over time.

Notwithstanding the possible benefits demonstrated by the effect of expert advice in both the immediate and long-term, the findings of this thesis suggest that perceptions may benefit from instruction, however, the impact of advice on perceptions appears to be transient as demonstrated by changes in the immediate term (Study one) over those achieved in the longer-term (Study two).
Although enhancement in perceptions were observed in the long-term, these changes appear to be less enduring. Consistent with the existing research (Blanch-Hartigan et al., 2012; Frank & Feely, 2003; Hauch et al, 2016; Shaw et al., 2013), the findings of this thesis point to pre-existing perceptions over-riding advice either sooner or later. While the benefit of advice by the expert appears to be broader than other forms of advice, the ability of instructions to effect permanent change to pre-existing beliefs appears unlikely. That said, in the context of juror decision-making, effecting permanent change in juror opinion may not be the necessary threshold. At least for the benefit of the court, jurors’ knowledge of deception indicia need only be enhanced to reliably inform their interpretations of witness credibility while they perform their duty to the court. Beyond their role as triers of fact, should perceptions revert back to baseline, all is not lost. If changes in perception are only achievable in the short-term or at most, over the course of a one-week period, this may suffice in providing some safeguard to otherwise erroneous interpretations based on false, misplaced beliefs of indicia of deception.

12.5. **What Combination of Nonverbal Cues are Considered Most Important in Deception Detection?**

The current thesis explored at factor level the combination of cues identified as the most important by jurors in their determinations of witness truthfulness/deception. The exploratory analyses revealed a number of findings. Before advice was received, jurors across both studies identified the same four factors as among the most important cues in determining the credibility of a witness. These were: confidence; spontaneous corrections; implausible responses; and responses that contain a lot of detail of which two were diagnostic and two were not. Only the rank order of importance differed between studies. Of note, it was the fifth most important cue that differed between studies with responses that lack logical structure and gaze aversion nominated. Interestingly, the number one factor ranked most important was the same across studies. This was the non-diagnostic cue, confidence.
While there is no universal cue akin to Pinocchio’s nose that enables us to detect deception, confidence despite its non-diagnostic quality, is routinely perceived conclusive of witness truthfulness and has been shown to account for the majority of variance in jurors’ determinations of eyewitness believability (McClellan, 2006). More recently, Bogaard and colleagues (2016) found gaze aversion to be the first and second most identified cue to deceive among students and police officers respectively. Aside from gaze aversion none of the cues identified in the current research overlapped with those identified in their study (Bogaard et al., 2016) even though the pool of nonverbal cues investigated were largely the same.

An unexpected finding followed after jurors received advice. Not only did the combination of cues deemed most important remain relatively unaffected with-in studies (i.e. three of five cues were maintained after immediate advice and four of five were retained after delayed advice) but the combination of cues remained largely stable regardless of whether advice was considered immediately or over time. These were confidence, responses that contain a lot of detail, implausible responses, and responses that contain very few details: two diagnostic and two non-diagnostic.

In terms of the influence of advice on shifting importance, both immediate advice and advice over time increased jurors’ levels of importance on a handful of deception cues with the largest percentage changes demonstrated across the same three indicia. Not only did these changes reflect increases in importance on reliable cues but the gains apportioned remained in the same hierarchical order irrespective of time-to assessment (i.e. immediate or delayed advice). The three cues reflecting the largest percentage change with advice were high pitch voice, body movements including hand, finger, leg and foot movement, and responses that contain very few details.

The results summarised for cues ranked most important raise a few scenarios. On the one hand, advice demonstrated little benefit in adjusting hierarchical rankings of importance for cues considered in the top five. That said however, the “top five” parameter set was an artificial one and is unlikely to bear any significance in practical terms when making determinations as to the
credibility of a witness. On the other hand, the results clearly illustrated that overall importance attributed to both reliable and unreliable cues can be shifted with advice. The difficulty posed however is that with advice, some reliable cues decreased in importance and some unreliable cues increased in importance. This finding may indicate a number of different possibilities. First, that there may be a saturation effect for advice with only so much of the evidence presented to jurors retained and considered. A sum of 31 different cues – some similar, others not – may have exceeded the threshold of retention reasonably expected of jurors.

An alternative hypothesis as outlined previously may be that the content of advice tendered to jurors lacked specificity as to the individual cues jurors were asked to evaluate. While some indicia mirrored those in the questionnaire, some were discussed in more general terms. As noted in chapter seven, content analysis revealed that neither confidence nor responses that contain a lot of detail were pinpointed in the advice presented. That is, jurors did not hear evidence to discount these indicia as non-diagnostic of deception. Therefore it’s not surprising that the two beliefs held paramount by jurors did not change post-intervention. Moreover, the increase in importance observed for ‘responses that contain a lot of detail’ may have arisen for no reason other than common sense. That is, the advice dealt specifically with the factor ‘responses that contain very few details’ and drew attention to the omission of spatial, temporal or perception details as helpful in evaluating the truthfulness of a witness. As common sense would have it, the opposite ought to be true and therefore ‘responses that contain a lot of detail’ must be indicative of deception. However, this is not the case (DePaulo et al., 2003; Levine, 2018). It is proposed that the label used to identify the cue may have had an inadvertent, prejudicial effect on responses in that it conveyed the opposite of what was presented in the advice and could have therefore reasonably been induced as indicative of deception.

A further possibility that warrants consideration is that although the content of advice was similar across conditions – judicial transcript and judicial direction were identical – the content of
the evidence presented to jurors did not address all 31 cues equivalently with some types of advice deliberating more conscientiously on some factors at the expense of others. As a consequence, including cues that had not been specified by advice, may have diluted the effect of advice. Furthermore, asking jurors to consider 31 cues may have exceeded the number of cues that could be realistically been dealt with by jurors. Consequently, it is possible that the jurors’ assessment of advice for some cues but not others may have, albeit inadvertently, muddied the overall effect of advice found. These considerations raise further questions as whether the provision of advice needs to strike a balance between adducing information on a smaller number of cues, and whether there is a utopian combination of cue types that would assist jurors the most. Determining a practical threshold pertaining to juror comprehension would also be worthy of future research to establish how many cues (reliable and unreliable) would be optimal to instruct jurors on so that advice could achieve the maximum benefit on shaping jurors' beliefs to better inform their assessments of witness credibility.

The increase in importance observed on confidence and responses that contain a lot of detail, found across both Study one and Study two highlight an issue for consideration. Namely, when the most salient of beliefs about deceptive behaviour are wrong, and those beliefs are not counteracted with tendered advice specifically, then oversight of that advice may correspond with that factor amplifying in importance. This finding, as was demonstrated in the current research, lends support to the idea that not only should jurors hear from expert witnesses as to nonverbal diagnostics of deception, but that the content of that evidence should target the combination of cues perceived by jurors as most important. As has been demonstrated herein preliminarily, if a small selection of unreliable cues are overlooked and not directly countervailed jurors may assign more importance to these behaviours in both the immediate term and over time. Determinations as to witness credibility are a function held, unreservedly, by the jury. However, without the aid of expert evidence as a mechanism to correct jurors most salient beliefs pertaining to non-diagnostic cues of deception the potentiality for damaging and unsafe decisions may continue unabated.
12.6. Did Advice Effect Juror Perceived Confidence in Detecting Deception?

The effect of education on juror confidence was explored in this thesis. To date, no research has examined whether or not the provision of advice influences perceived confidence among jurors. This thesis aimed to redress this gap in knowledge. Given the exploratory nature of comparing the effect of instruction on confidence, no directional hypotheses were predicted. The analysis was of interest with respect to the deception literature that suggests over confidence – aptly coined the “over-confidence effect” – has a negative relationship with accurately detecting deception (DePaulo et al., 1997).

While veracity judgements were not the focus of this thesis, investigating whether instructions influenced confidence is of value. In related literature, it has been well established (i.e. eyewitness testimony) that confidence in knowledge is a weak predictor of accurate knowledge (Brewer & Wells, 2006; Sporer, Penrod, Read & Cutler, 1995). This relationship is an important one as it has been shown that jurors rely on perceived confidence when assessing eyewitness evidence (Charman, et al., 2011; Deffenbacher, 1980; Wells, Lindsay, & Ferguson, 1979).

This thesis demonstrated, provisionally, that juror perceived confidence does not appear effected by the provision of advice. Not only did the efficacy of advice not influence juror confidence (i.e. shifts in perceptions of nonverbal cues were not associated with shifts in confidence); but shifts in confidence were not influenced by type of advice (i.e. no type of advice, successful or otherwise, was associated with changes in perceived confidence), and shifts in confidence were not influenced by time-to-assessment (i.e. levels of confidence were not influenced by advice considered immediately or in the long-term). Despite previous research to suggest confidence in initial judgements of trustworthiness increase with time (Willis & Todorov, 2006), the current findings suggest no apparent relationship between the provision of advice, enhancement of perceptions of nonverbal cues of deception/trustworthiness and perceived confidence in detecting deception. Overall, using instructions as a safeguard to educate jury
members on diagnostics of deception does not seemingly warrant caution because negative effects associated with over confidence (DePaulo et al., 1997), were not found when advice on nonverbal cues of deception was tendered.

12.7. Implications

A number of conclusions were derived from Study one and Study two. Although aforementioned, these are considered with respect to implications for existing theory and implications for forensic practice.

12.7.1. Implications for Existing Theory

It has been widely established that beliefs about deceptive behaviour are so embedded that they reemerge without enticement (e.g., Anderson et al., 1998; DePaulo, 1994; Feeley & Young, 2000; Frank & Feeley, 2003; Zuckerman, Koestner, & Colella, 1985). This thesis set out to test the central tenants of DDT that posits jurors’ initial perceptions of trustworthiness are so entrenched that they are unlikely to change over time or in response to conflicting evidence (Porter and ten Brinke, 2009).

At the outset of this thesis, the jury was out as to the legitimacy of DDT and its prediction that initial opinions are so steadfast that they cannot, and do not, change even when confronted with evidence to the contrary. The findings herein provide mixed support for DDT. Primarily, perceptions may benefit from advice. However, the benefit appear to be short-lived with enhancements to perceptions largely obtained in the shorter-term. Conversely, in the longer-term, the indications herein suggest that the education proffered to jurors is undermined by beliefs that are pervasive, to one extent or another, with the majority of juror opinion restoring to baseline. Only expert evidence and expert/judicial advice combined appeared to impact perceptions over time and the efficacy of the advice was limited to correcting misconceptions. As such, this thesis proposes that the attestations made by DDT – that stereotypes about deceptive behaviour are
impervious to change – may be oversimplified and that rather, with advice, it may be possible to modify opinions in the short-term with changes in the longer-term, less so.

Indeed, regarding the specifics of advice, the current thesis demonstrated that information on indicia of deception may benefit decision-makers in reshaping their initial impressions as to determinants of witness credibility. This thesis showed that introducing expert evidence on detecting deception may have utility not only with respect to enhancing jurors’ opinions of reliable nonverbal cues of deception in the short-term but too may correct misconceptions of unreliable indicia in both the immediate and long term. Although instructions by expert witness was the only type of advice that was uniformly successful across cue types, the current research also revealed that instructions by way of expert evidence and/or judicial directions may enhance opinion of reliable cues in the immediate term while correcting false beliefs in the longer-term.

The findings arising from this thesis suggest that perceptions of deceptive behaviour may be more complex than first proposed by DDT. The theory predicts that not only are initial impressions typically unreliable, but they too are so strongly contented that they negatively influence subsequent interpretation and assimilation of new information, especially when that information is contrary to belief (Porter & ten Brinke, 2009, 2010). Based on the findings of this thesis the claim that initial perceptions are characteristically unreliable may be misplaced. An unexpected finding of this research was that although people hold false beliefs regarding which behaviours infer deception they also hold stereotypes that are valid. Further to this, the prevalence of false beliefs do not appear to outweigh reliable ones. Rather, the studies herein revealed that perceptions of reliable and unreliable cues may in fact be perceived equally at baseline and that the weight afforded to cues are not mutually exclusive. This finding provides some preliminary reassurance that perhaps the picture depicted in the extant literature is not as dire as first thought. Further, this finding may shed some light on why veracity judgements of detecting deception are no better than chance: the base rate of detection accuracy may well be accounted for by reliable
and unreliable cues being considered equal. Research suggests that poor detection rates are a consequence of people relying on invalid nonverbal cues (Levine & McCornack, 2014; Vrij et al., 2019) and/or because valid cues are faint (Hartwig & Bond, 2011; Strömwall, Granhag, & Hartwig, 2004; Masip, 2017; Vrij et al., 2010; 2019; Vrij & Turgeon, 2018). While this might be the case, the findings herein at least challenge the premise that people rely on invalid cues and instead suggest that veracity judgements may be more indicative of a case of odds: if in fact jurors perceive valid and invalid cues to be equal and nonverbal cues are expressed equally, then jurors have an even chance of getting it right or wrong. Of course, this proposition remains no more than a hypothesis. If additional support can be gleaned to show that nonverbal cues are weighted equally, then it would provide an alternative – and perhaps superior one – as to why people are poor at detecting deception.

Moreover, the current thesis has been able to show that reliable perceptions may be encouraged with new information through expert testimony, but also, that unreliable perceptions may be corrected. The studies also demonstrated tentative evidence for shifting initial perceptions regarding deceptive behaviour in both the immediate term and over time. That said, however, based on the variability in effects of advice observed, the thesis provides some evidence in keeping with DDT that pre-existing perceptions endure with time, over-riding some forms advice either sooner or later.

12.7.2. Implication for Forensic Practice

In practical terms, a number of implications are identified for forensic practice, legal doctrine and jury trials more generally. Firstly it can be concluded that through the admission of responsible expert testimony, jurors’ beliefs about deceptive behaviour may be positively influenced to better inform their determinations of witness credibility. It can also be said that expert testimony that attends to reliable/unreliable diagnostics of deception appears to achieve its maximum effect in the immediate term. It has also been established that the benefit of instructions
by an expert witness appears to diminish over time, with expert testimony only found to correct for mistaken beliefs about deceptive behaviour in the longer term whereas opinions of reliable indicia of deception do not appear sustainable.

If jurors are to receive maximum benefit from expert testimony to aid in their evaluations of witness credibility then information as to reliable/unreliable diagnostics of deception would be best placed if it was adduced immediately before jurors hear from a witness. Given the intricacies of a trial the ability to introduce expert testimony requires planning and forewarning to ensure all parties are available and in attendance. When advocates for either the defense or prosecution consider witness credibility to be at issue consideration should be given, as much as is practical, to delivering the evidence of an expert within the immediacy of the witness in question. When these machinations are not possible, and members of the jury find themselves considering the advice of an expert witness a week thereafter, then their assessments, as a result, may be weaker. While decisions as to the credibility of a witness may be protected from false stereotypes, the benefit of instructions on jurors’ beliefs regarding reliable cues may have diminished and become less reliable. While instructions may not directly benefit opinions concerning reliable indicia in the longer term, some comfort can be taken from jurors’ opinions of reliable cues improving on their own with time. Further, although enhancements in opinion of reliable cues are unlikely to be maintained, at least and perhaps more importantly, advice may correct misconceptions that are more hazardous and at risk of misleading interpretations of witness credibility.

The findings of this thesis present some potential opportunities for legal reform where jury trials are concerned. This thesis evidences some benefit of adducing expert testimony as to the scientific basis of nonverbal cues as a worthy adjunct to protect against unsafe decisions when witness credibility is at issue.

The current research has demonstrated that the effects of instruction on shifting and maintaining perceptions can be impacted depending on whether instructions are considered
immediately or after a delay. What remains unknown is at what point does the influence of advice wear off. Taken from the findings herein, the benefit of instructions on jurors’ perceptions of reliable cues drops off by one-week whereas the effect appears to maintain to some degree for perceptions of unreliable cues. Therefore, it would be of value to establish whether there was an effective period of time in which to give jurors the necessary opportunity to reflect on advice tendered so that they can consider it while ensuring its effect is maximised.

Although the findings from the effect of no advice (i.e. the controls) need to be considered cautiously, it is noted that the findings herein point to the following interpretations. If jurors continue to be encouraged by the judiciary to consider demeanour evidence over the course of a trial, then such directions ought to be accompanied with advice. That is to say, that the instruction to consider cues of deception without requisite advice may have an inadvertent, negative effect on perceptions of unreliable cues (people perceive unreliable cues to be more indicative of deception over time when they don’t receive advice). If instruction by way of expert evidence cannot accompany jurors requirement to consider the demeanour of a witness then potentially no invitation to consider demeanour should be given. Notwithstanding replication, it could be inferred (preliminarily and with caution) that failing to instruct jurors as to the reliable/unreliable cues of deception may have negative, unwanted consequences on perceptions independent of advice.

All conclusions aside, it cannot be understated that although the findings herein demonstrated shifts in juror’s perceptions of nonverbal diagnostics of deception, what remains to be seen is whether or not shifts in opinion will apply to final assessments of credibility. This is a salient point not only because assessing witness credibility is a fundamental role reserved for the jury but also because if the acquired knowledge cannot be applied, then the benefit of instructions becomes inconsequential. Further to this, if the instructions taken onboard can be applied, what remains unknown is whether or not they can be applied correctly. If the transfer of knowledge from instruction to application result in erroneous assessments of credibility, then adducing
evidence as to reliable/unreliable diagnostics of deception not only invalidates the purpose of introducing the evidence but it would undermine the entire curial process. These issues present logical steps for future consideration.

Exploratory investigations revealed a small combination of cues commonly perceived by potential jurors as highly important in evaluating the truthfulness of a witness (e.g. confidence, implausible responses, responses that contain very few details and spontaneous corrections). While the findings are only precursory, instructions that focus on these cues could achieve greater returns than attained when an exhaustive list of empirically established cues are presented. Jurors are more likely to be able to retain and recall a smaller number of factors than an extensive list of cues. Furthermore, the handful of factors identified as most important by jurors – both reliable and unreliable – may be used to improve credibility assessments because they do not require prior knowledge of a witness’s usual presentation of behaviour to interpret them. Even though baseline knowledge of behaviour is not a prerequisite to accurate assessments of deception (Porter, et. al. 2007; 2010), if jurors can evaluate cues that do not require relevant a-prior knowledge of behavioural mannerisms then it can be gleaned that assessments of credibility may be improved from developing jurors’ knowledge of these cues alone.

12.8. Strengths of Thesis

This thesis attempted to systematically replicate previous research to establish whether providing advice to jurors improved the behavioural cues relied upon when assessing the truthfulness/deception of a witness in court. This body of work has not only contributed to a scarcity of replication studies, but also has attempted to expand current knowledge in what is a vast area of research on deception detection with specific contribution to the knowledge bank on nonverbal, behavioural cues of deception that are plagued by criticism and debate. More broadly, the current body of work aids the field of forensic practice with its finding that the provision of evidence may influence deception cues relied upon by jurors, and so, has the potential to positively
shape jury reform with the evolution of demeanour evidence and role psychological evidence could have in enhancing the ability of triers of fact in their duty to evaluate witnesses credibility.

Further to this, the thesis was the first to explore how delay impacts consideration of advice given to jurors. Additionally, other original contributions of this thesis included the incorporation of independent controls that allowed for confounding factors such as retest effects and time to be extrapolated. What is more, investigating the effect of instructions on perceptions of unreliable cues in addition to perceptions of reliable indicia was also significant. Not only did this extension to the investigation uncover some unexpected findings but it also identified evidence that directly challenges existing assumptions regarding deception detection accuracy.

12.9. Limitations

Setting aside the strengths of the current work, the novel investigations conducted, and findings herein, a number of limitations are likely to have shaped the results and ought to be considered when interpreting the findings. The limitations are outlined with respect to the project’s research design, the materials and stimuli employed, and the statistical analyses performed.

12.9.1. The Research Design.

Deficiencies in the design of the replication and subsequent extension were observed in the process of analysing and interpreting the results. At issue for both studies was the design of the control groups. Although the introduction of controls is a great asset of the thesis, elevating its experimental design compared to that of the original study that relied upon baseline responses as inbuilt controls for each group, the design of the controls was not sophisticated enough to rule out the potential for practice effects or rote learning/memory. That is, the pre/post-test control for both studies was not equivalent; it omitted an interference, distractor task that may have benefited analysis by militating the risk of practice effects. As it stands, the current researchers are unable to ascertain with certainty that the improvements in behavioural cues relied upon by jurors can be
attributed to the effect of advice alone, or alternatively reflect participant’s ability to perform a memory task. As such, the possibility of other unknown confounding variables cannot be eliminated. This perspective, although pessimistic, is also problematic for the study conducted by Coyle and Thomson (2014) whose quasi-experimental design did not include a dedicated, equivalent control. Lastly, controls were not collected for perceived confidence and so the analysis of juror confidence was subject to the same limitations identified in the within-group analyses where baseline measures were adopted as in-built controls.

A-prior power analyses indicated that a sample of 164 participants would be adequate to detect small effect sizes for within-group analyses. A-prior analyses were not conducted for between-group comparisons as this was not identified prior to data collection. However, given that the advice groups were smaller than anticipated (i.e. instead of 30 subjects per cell, some advice conditions were as small as 13 participants), post-hoc power analyses revealed power ranging from as low as .21 to .40 for Study one and .18 to .20 for Study two in order to detect small effects. Accordingly, the reduced power observed for both studies suggests the small cell sizes may have failed to detect true, small relationships for the within-group comparisons. Equally, to detect small effects from the between-group comparisons that examined shifts in perceptions against the controls, post-hoc analyses indicated power ranging from .39 to .42 for Study one and from .34 to .36 for Study two. It is suggested therefore that the likelihood for Type 2 errors render caution when it comes to interpreting the findings of studies one and two for both the within and between-group analyses.

Sample representativeness presents a limitation for both studies. The gender split across both samples was predominantly female and is not therefore reflective of the statistical profile of Victorian jurors obtained from Juries Victoria (2019). That said however, the make-up of juries empaneled in any given trial are unlikely to be split 50/50 due to numerous factors such as challenges for cause resulting in potential jurors being excused from service. Nevertheless, the
impact of gender differences herein could not be investigated because of the limited number of male participants recruited.

Social media platforms have great utility for recruitment because of the potential to reach large and specific audiences, however, the evidence base for the efficacy and validity of online recruitment in terms of recruitment success, sample representativeness and response rates is limited (Bennets et al., 2019). One unintended consequence of using online channels for recruitment has been shown to result in unrepresentative and unique samples relative to traditional approaches (Arigo et al., 2018). Consistent with research into convenience sampling that found Facebook users are significantly more likely to be females (Well & Link, 2014), recruiting participants via online channels may explain the imbalance of gender found in studies one and two. Similarly, the higher proportion of tertiary qualified participants found herein is consistent with a large-scale Australian study that found participants recruited via Facebook were more likely university educated (Bennett et al., 2019) compared to population-based sampling.

The potential for selection bias in the current thesis warrants consideration. The content of advertisements on Facebook has been shown to lead to different rates of recruitment and engagement (Basa-Martinez et al., 2018; Chio et al., 2017; Wozney et al., 2019) but also research has found that the content of promotional material used in recruiting participants has lead to selection bias directly impacting demographics and target measures under investigation (Teo et al., 2018; Chio et al., 2017). Consequently, the messaging content of the research flyer along with the online networks available to the author may have inadvertently impacted the external validity of the studies with the pool of participant recruited affecting the generalizability of findings to population-based samples and potentially the original study replicated.

12.9.2. The Materials and Stimuli.

Pertinent to this thesis was the aim to replicate the findings of Coyle and Thomson (2014). Accordingly, the materials and stimuli used in the original study were required herein, warts and
all. Nevertheless, a number of shortcomings were observed. First and foremost, is the research questionnaire that asked participants to rate each nonverbal cue on a six-point scale. The questionnaire used the verbal anchors “extremely infrequently” for a scale score of 1 and “extremely frequently” for a corresponding score of 6. On the face of it the verbal anchors appear appropriate. However, direct feedback from participants suggests that the verbal anchors adopted in the survey were confusing to interpret and impacted participants’ ability to interpret the task at hand: “rate each factor that enables you to determine whether someone is truthful if they were giving evidence in a court case.” As corroborated by participant feedback, it is the opinion of the author that the verbal anchors may have muddied the instruction given. As a consequence, it is probable that participants found the survey arduous and may have had some bearing on survey comprehension and completion thereby raising the issue of internal validity for the measures assessed. The proposition that participants found rating behavioural cues in terms of frequency confusing and arduous is supported by a drop-out rate of 57% (n = 364) of the total research sample at this item. By comparison, it is unknown whether the same questionnaire implemented in Coyle and Thomson’s study (2014) was accompanied by any verbal instructions or clarification on how to interpret the item. Furthermore, it is unknown whether participants had the opportunity to seek clarity on how to interpret the question. A pilot study testing the validity of the verbal anchors of the questionnaire may be warranted and future research would be well placed to investigate the most appropriate, intuitive verbal anchor for the task at hand.

12.9.3. The Statistical Analyses.

The potential for Type 1 error rates also needs to be considered given the multiple tests performed across the current studies. However, confidence that genuine relationships were found is increased by the fact that the pattern of results are in the directions expected and the low power thresholds achieved with the samples recruited.
The prevalence of non-parametric methods used throughout the thesis could be perceived as a limitation because non-parametric methods are less powerful than parametric procedures (Field, 2013). However, non-parametric tests were dictated by the fact that normality was violated, first and foremost, but also in purists’ terms, consistency commanded that the statistical methods replicate those performed in the original study thereby militating any confounding factors that may have arisen as a result of undertaking different statistical approaches.

An unfortunate, missed opportunity for the thesis was that the two studies met criteria for randomized control trials (RCT) had the projects been registered. Although not a limitation in the traditional sense, it is a shortcoming that has had a bearing on the potential standard of experimental design that could have been achieved as well as limiting the potential contribution to the field had the RCT not been overlooked.

12.10. Future Directions

Research into the potential benefits of educating jurors requires further exploration if we are to understand how best to improve jurors’ opinions of nonverbal cues of deception if they are to be aided in their determinations of witness credibility. A number of opportunities for future investigation were identified throughout. A next logical step would be to establish whether the instructions that shifted perceptions can be applied when making decisions as to the truthfulness/deception of a witness. The ability to comprehend and apply instructions as to reliable/unreliable cues of deception engage distinct cognitive processes (Krathwohol, 2002). As highlighted by Baguley and her co-authors (2017), jurors’ comprehension of instructions are not sufficient to ensure application occurs. Consistent with McGuire’s model of persuasion (McGuire, 1968, 1972), the current research went as far as to show that jurors not only comprehended the instructions but they may remember them over-time. What we don’t know, however, is whether or not they can apply the instructions as would be indicated by jurors acting in a way that is consistent with the instructions received. If jurors’ ability to apply the instructions are not
confirmed or denied then the shifts in opinions found herein cannot be concluded to reflect acceptance and/or agreement with the information presented. Establishing juror’s ability to apply the advice also requires that the information be applied correctly. With nonverbal cues widely accepted as faint and difficult to detect (Masip, 2017; Vrij et al., 2010; 2019; Vrij & Turgeon, 2018), jurors’ ability to apply the knowledge accurately may be limited. These considerations warrant further research.

Assessments of witness credibility are fundamental to the role of jurors. A gap in current knowledge is to what effect instructing jurors has on the accuracy of assessing witness truthfulness and further, what impact the provision of advice may have on the outcome of verdicts. If it can be shown that the application of knowledge as to reliable/unreliable cues of deception not only enhances the veracity of jurors’ assessments of witness credibility but also promotes safe verdicts then the benefit of instructing jurors maybe strengthened. To the contrary, if instructions by way of expert or judicial directions cannot be applied or are applied incorrectly then the ability of advice to shift perceptions becomes a moot point.

As identified in the discussions of studies one and two, the impact of source of advice differed not only by type of cue but also over time. An area worthy of further exploration would be to systematically analyse differences in presentation modalities across source of advice but also differences in the combinations of source of advice. Herein the combined effect of expert evidence and judicial directions was investigated. However, the impact of expert evidence followed by a written judicial instruction was not. Neither were oral instructions followed by written instruction nor expert evidence followed by oral and written judicial directives combined. With amendments to the Jury Directions Act (2015, 2017) facilitating both the use of written and oral directions to the jury to aid in assessing the reliability of a witness’s evidence or a witness’s motive to lie, members of the judiciary are delivering more written directions than before (Judicial College of
Victoria, 2019). However, what remains to be seen is what impact oral and written instructions combined have on jurors’ decision-making when demeanour evidence is at issue.

The current thesis raised the issue that the effect of advice may be due to a number of factors. Given the lack of effect evidenced by judicial directions, it is probable that the benefit of instructions are influenced by the duration of the advice given with preliminary indications to suggest advice needs to strike a balance between length of advice given. If too short and the advice does not provide sufficient explanation as to the research basis of the information presented then it is unlikely to have an impact. Further research could consider exploring what duration of advice is optimal in order to maximise the benefit of the advice while balancing the needs to maintain motivation and attention.

It was established that the benefit of instructions through expert testimony are effective in enhancing jurors’ perceptions of reliable cues in the immediate term but that these do not hold over time. What we don’t know is at what point in the intervening week do instructions dissipate in effecting change on opinions of reliable cues of deception. It is after one day, two days and so forth? Equally, we don’t know at what point the benefit of expert testimony drops off altogether having no impact on shifting perceptions of reliable and unreliable cues of deception. These factors are worthy of further exploration. Future research could provide great insights into this unknown with considerable implications as to when expert testimony should be adduced to achieve maximum benefit.

For the provision of advice to be effective jurors must be capable and motivated to follow the instructions given and ultimately, apply that knowledge. Based on the findings herein, the results suggest that the impact of instruction on shifting perceptions is effective when it is presented before jurors are asked to evaluate nonverbal cues. What remains to be seen is what effect, if any, advice has on educating jurors after they have heard from a witness and potentially evaluated demeanour evidence. Therefore, research is yet to established whether or not the timing of
instructions influence the efficacy of advice, and if so, at what point does advice have the greatest effect or conversely whether or not the timing of the advice is of any import.

Furthermore, despite the merits of adducing expert evidence on improving perceptions what has not been determined is what effect expert evidence would have on shifting perceptions when jurors were faced with competing expert testimony. It has been well established that conflicting experts can have negative effects on juror decision-making (Jones et al., 2017; see Ribbers & Henneberg, 2018, p.265) yet the effects of confusion on shifting perceptions when jurors are faced with duelling experts remains unknown.

The findings observed in both Study one and Study two, that showed shifts in perception in the immediate and long term when advice was withheld necessitates further research into what other factors may be driving jurors’ opinions of indicators of deception. Namely, jurors who received no advice evidenced opposing trajectories depending on whether they evaluated deception cues immediately or over time and the current work could not isolate whether these differences were due to time alone or other extraneous factors.

A shortcoming identified in the current thesis was the potentially inadvertent effect that measuring jurors’ opinions across 31 cues may have had in undermining the impact of advice when advice did not address each cue equally, and for some not at all. It is suggested therefore that future research could examine the correlation between the cues specified in the instructions given with improvements in opinions achieved. Following on from this, future investigations could identify whether or not there is a cut-off pertaining to jurors’ comprehension of instructions given and if so, how many cues (reliable and/or unreliable) would optimise jurors’ opinions as to indicators of deception.

In the same vein, Study one and Study two highlighted that a small handful of factors are deemed by jurors to be most important in detecting deception. This finding presents an opportunity for future research to explore what gains could be made in educating jurors as to reliable/unreliable
diagnostics of deception if these specific cues were targeted. Given that the combination of nonverbal cues perceived by jurors as most important in detecting deception before and after advice was first explored herein, the findings are at best preliminary. Although one other study has examined beliefs as to nonverbal cues these related specifically to law enforcement personal (Bogaard et al., 2016) and did not look beyond baseline perceptions.

The value of any replication or ability to generalise findings cannot overlook the benefits of studying jurors who can deliberate on the demeanour of real witnesses in actual trials, or at the very least mock trials. The ability to recruit jurors relinquished from service would likely reveal valuable insights that cannot be achieved through convenient sampling. Further to this, future replication attempts could alleviate the issues faced with limited power by consolidating the number of advice conditions studied if sizeable samples were unlikely. Last but not least, the anomalies observed between the control groups in this research would benefit from further investigation. Future research would be well placed to redress the methodological shortcomings of the controls by ensuring, at least, that a placebo task is incorporated that mimics that of advice but without the value of the information instructed.

12.11. Conclusion

In their role as triers of fact, jurors are expected to rely on common sense when interpreting the demeanour of a witness to assist them in their determinations of credibility. This venerated belief assumes that a commonsense approach is reliable. It is not; relying on demeanour evidence to infer witness credibility ignores the fact that jurors typically perform at chance in accurately discriminating truthfulness from deception. With the veracity of detecting deception undermined by invalid and unreliable perceptions pertaining to nonverbal indicators of deception, verdict deliberations may be egregiously flawed if jurors’ misconceptions are not redressed.

It has been well established that judicial directions have limited efficacy in influencing juror-decision making. Also, it cannot be overlooked that the judiciary are vulnerable to the same
thinking errors and misconceptions faced by jurors. Hence, relying on judicial directions in their current form to guide jurors as to interpreting demeanour evidence is misdirected and may be contributing to and perpetuating erroneous deliberations. Accordingly, educating jurors as to the scientific basis vis-à-vis nonverbal indicators of deception is of great import. Given that introducing specialised knowledge to assist jurors in an actual trial can only be adduced through expert evidence or evidentiary-based instructions, this thesis investigated the impact of advice on benefiting jurors’ knowledge of nonverbal diagnostics of deception.

In doing so, the thesis highlighted that jurors’ perceptions of nonverbal cues of deception may be influenced with advice by way of expert testimony and/ or when combined with oral evidentiary instructions. The current research also demonstrated that advice impacts jurors’ opinions differently, over time. The current research shows how best to assist jurors in interpreting demeanour evidence if assessments of witness credibility and their resultant deliberations are to be safeguarded from risky decisions. Moreover, exploratory analyses revealed preliminary findings that suggest educating jurors through the provision of instructions does not effect perceived confidence in detecting deception. Notwithstanding replication, the finding that advice does not influence confidence is encouraging suggesting that negative effects associated with over-confidence are unlikely to undermine interpretations of nonverbal cues of deception.

This thesis draws attention to the need to advance empirical knowledge on the benefits of educating jurors as to witness demeanour before a verdict can be returned that triers of fact cannot reliably and accurately identify nonverbal cues of deception. To do so would be premature; the jury is still out as to the uniform benefits of advising jurors over the longer-term with the thesis suggesting that pre-existing beliefs may in fact prevail with initial opinions re-emerging with time, irrespective of advice. That said however, this body of work has provided some indications that empirically-based education introduced responsibly through expert evidence may achieve both
immediate and longer-term effects on militating jurors’ biases and misconceptions around flawed stereotypes of deception.

Considered as one body of research, this thesis demonstrates the importance of understanding deception in courtrooms. In summary, the current thesis provides mixed support for DDT (Porter & ten Brinke, 2009). Although the evidence suggests that pre-existing beliefs endure, sooner or later, well-established opinion that stereotypes of deceptive behaviour cannot be changed is not entirely correct – at least in the immediate term. This outcome is a welcome one and provides some reassurance that the foundation on which jury trials operate are not undermined by flawed assessments of credibility based on fallible perceptions resistant to change. The conclusions do however highlight the responsibility imparted on the legal fraternity to improve and safeguard decision-makers determinations of witness credibility by facilitating jurors’ education with empirical knowledge tendered correctly, responsibly and timely through expert testimony.

Deception experts criticise the use of nonverbal behaviours in detecting deception, however, research on deception detection in courtrooms has remained relatively unexplored. It is hoped that this thesis will stimulate research interest and future investigations that will explore how reliable, responsible and timely psychological evidence can protect against jurors relying on misplaced common sense when assessing witness demeanour and thereby removing the need for members of the judiciary to proffer psychology from the bench.
References


*Abalos v. Australia Postal Commission* (1990) 171 CLR 167


Ayturgrul v. The Queen (2012) 15 HCA


Baira v. RHG Mortgage Corporation Limited (2012) 387 NSWCC.


*CSR Ltd v. Della Maddalena* (2006) 224 CLR 1


*Daubert v. Merrell Dow Pharmaceuticals, Inc*, No. 92-102, 509 579 (Supreme Court of the United States 1993).


*Devries v. Australian National Railways Commission (1993)* 177 CLR 472


Evidence Act, (ACT), (2011)

Evidence Act, (NSW), (1995)

Evidence Act, (Tas), (2001)


Farrell v. The Queen, (1998) 194 CLR 286


Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie
detection training. *Journal of Applied Communication Research, 31*(1), 58-75.

Frank, M. G., & Svetieva, E. (2013). Deception. *Nonverbal communication: Science and
applications*, 121-144.

Freckelton, I. (2004). The closing of the coffin on forensic polygraph evidence for
and Law, 11*(2), 359-366.

presented at the The Australian Institute of Criminology, Law, Medicine and
Criminal Justice Conference, Surfers Paradise, Queensland.

*Frye v. United States*, (1923) 293 Federal Court of Appeals, 1013 D.C.. Cir

Geiselman, R. E. (2013). Training novices to detect deception in oral narratives and

Gemma, W., Schertler, E. & Bull, P. (2019). Detecting deception from emotional and

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way
through the publication bottleneck but undermine science. *Perspectives on
Psychological Science, 7*(6), 562-571.

Psychology, 37*(1), 60-74.

Cengage Learning.

evidence and judicial directions to counter misconceptions in child sexual abuse
INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION


INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION


Jones v. Hyde, (1989) 63 ALJR 349


Jury Directions Act, (Vic) (2013)


*Makita (Australia) Pty Ltd v. Sprowles* (2001) NSWCA 52 NSWLR705

*Mallard v The Queen* (2003) WASCA 296

*Mallard v The Queen* (2005) HCA 224 CLR 125


*Mcintyre Nominees P/L v Tradeworks Transport Pty Ltd* (2014) SASC 109


*Nominal Defendant v. McLennan* (2012) NSWCA 14


Paterson v. Paterson, (1954) 89 CLR 212


*R v. Francois* (1994) SCC 827

*R v. Marquard* (1994) SCC 223


*SAS Realty Developments Pty Ltd v Kerr (2013) 56 NSWCA*


*Société d'Avances Commerciales (Société Anonyme Egyptienne) v. Merchants' Marine Insurance Co (The "Palitana"),* (1924) 20 Lloyd's List Law Reports 140


*State Rail Authority (NSW) v. Earthline Constructions Pty Ltd (in liq) (1999) HCA 588*


*Suvaal v. Cessnock City Council* (2003), 77 HCA 41 ALR1449


*Warren v. Combs*, (1979) 142 CLR 531


Appendices

Appendix A: Plain Language Statement

PLAIN LANGUAGE STATEMENT

Full Project Title: Instructing Jurors on Nonverbal Indicia of Detecting Deception: Is the Jury Still Out?

Student Researcher: Jessica Woskett

Principal Researcher: Emeritus Professor Don Thomson

Dear Participant

You are invited to take part in this research project. This Plain Language Statement contains detailed information about the research project. Its purpose is to explain to you as openly and clearly as possible all of the procedures involved in this project so that you can make a fully informed decision about whether or not you would like to participate.

Please read this Plain Language Statement carefully. Feel free to ask questions about any information in the document by emailing the student researcher, Jessica Woskett at jwoskett@deakin.edu.au or by emailing the principal researcher, Professor Don Thomson at donald.thomson@deakin.edu.au.

Consent to collect your data for this study will be assumed once you have read this Plain Language Statement and have commenced the survey; please do not commence the questionnaire unless you understand the information provided regarding the study and you give your consent to participate in the research project. You are encouraged to print a copy of this Plain Language Statement to keep as a record.

In order to take part in this study you must:

- Be over 18 years of age
- Be an Australian resident

Purpose and Background

This is a research project being completed in partial fulfilment for the degree of Doctor of Psychology (Forensic) at Deakin University. The results of the study will be collated into an academic research thesis with the aim to be published in a peer reviewed psychology journal.

The purpose of this study is to investigate mock juror perceptions about cues to deception and to determine whether false stereotypes about lying can be corrected with education. You are being recruited as a member of a mock jury. We seek your perceptions regarding indicators of deception through an online questionnaire. The results of this study will further inform research in the field of detecting deception and legal decision-making.

Funding

This research project is entirely funded by Deakin University.
Procedures
Participation in this project will involve watching or reading a transcript of an expert witness or judicial officer presenting evidence on detecting deception. You will be asked to complete a short questionnaire two times: once before receiving the evidence on detecting deception and again after. There are no right or wrong answers: we only seek your perceptions. You will either be asked to participate in the project in one sitting in which you will complete the second questionnaire immediately after receiving advice on detecting deception. Alternatively, you may be required to complete the second questionnaire a week later. This will be determined randomly. If you are asked to return to complete the study one week later, we will send you an email notification as a reminder. The email notification will have a link to take you back to the online survey.

Participation in the study is estimated to take approximately 15 to 20 minutes. The video or transcript of evidence and questionnaire are in English and each question has a fixed-choice response scale. Please answer each question carefully. The questionnaire will ask you to rate the importance of different behavioural indicators that you would rely upon when determining the truthfulness of a witness in court.

Participation in this research project is entirely voluntary and anonymous. All personal information such as demographic details will be non-identifiable. You may withdraw from the study at any stage prior to submitting the survey. Should you close this window prior to submitting your completed survey, your responses will not be recorded.

Possible Benefits
Participation in this study may increase your knowledge of detecting deception, but there are no direct benefits for participation. Benefits to the community may arise from the results of this study in the form of increased understanding of detecting deception and legal decision making.

Possible Risks
We do not anticipate that there are any serious risks associated with participating in this study.

Privacy, Confidentiality and Disclosure of Information
Any information obtained in connection with this project and that can identify you will remain confidential. The questionnaire you complete will remain entirely anonymous - you will not be asked for any possibly identifying information.

The information obtained from this study will be kept in secure storage at Deakin University in digital form for 5 years after the final publication, after which it will be destroyed.

Results of Project
Results of the project will be reported in a research thesis with the aim to be published in a peer-reviewed journal. It will not be possible to contact you regarding your individual results or regarding publications due to the anonymity of the study. However, please enter your email address at the end of the survey or contact the researchers by email if you are interested in receiving an electronic copy of the final thesis. If this is the case, please take this opportunity to take note of the researchers’ contact details provided as they will not be available after this screen. Alternatively visit the studies website www.howtospotaliar.info for the research team’s contact details.

Ethical Guidelines
This project will be carried out according to the National Statement on Ethical Conduct in Human Research
INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION

(2007) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies. The ethical aspects of this research project have been approved by the Human Research Ethics Committee of Deakin University.

Complaints
If you have any complaints about any aspect of the project, the way it is being conducted or any questions about your rights as a research participant, then you may contact: The Manager, Ethics and Biosafety, Deakin University, 221 Burwood Highway, Burwood Victoria 3125, Telephone: 9251 7129, research-ethics@deakin.edu.au

Please quote project number HEAG-H 152_2015.

Further Information or Queries
If you require further information, wish to withdraw your participation or if you have any problems concerning this project, you can contact the lead researcher via email at donal.thomson@deakin.edu.au, or the student research at jwoskett@deakin.edu.au

The researchers responsible for this project are:

Ms Jessica Woskett, Professor Donald Thomson, Professor Ian Coyle and Professor Andrew Day
Deakin University
School of Psychology
221 Burwood Hwy
Burwood V/C 3125
Email: jwoskett@deakin.edu.au

CONSENT FORM

• I have read and I understand the above Plain Language Statement.
• I freely agree to participate in this project according to the conditions stated.
• I have been the opportunity to print a copy of the Plain Language Statement to keep.
• The researcher has agreed not to reveal my identity and personal details, including where information about this project is published, or presented in any public form.

By choosing to proceed to the survey, you are agreeing that you are 18 years or older, and are an Australian citizen.

(Survey will open in a new window)
### Appendix B: Questionnaire

*Note:* As displayed to participants online.

Q1

Instructions:
For each of the items below, you are asked to rate the following factors that enable you to determine whether someone is truthful if they were giving evidence in a court case. Select the number that best describes how you feel. There are no correct answers; we only seek your opinion.

<table>
<thead>
<tr>
<th></th>
<th>Extremely Infrequently (1)</th>
<th>Very Infrequently (2)</th>
<th>Somewhat Infrequently (3)</th>
<th>Somewhat Frequently (4)</th>
<th>Very Frequently (5)</th>
<th>Extremely Frequently (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pauses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High pitched voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidgeting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restless behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implausible responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postural changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech disturbances/fillers (i.e. “ah’s” and “umm’s”).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxious behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-manipulators (i.e. hand movements to relieve a bodily need like scratching an itch).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mumbling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covering the mouth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body/hand/finger/leg/foot movements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stuttering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrupting the questioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses that lack logical structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressed lips</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repetition of certain words and phrases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses that contain very few details</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q1 Continued

<table>
<thead>
<tr>
<th>Nonverbal Cue</th>
<th>Extremely Infrequently 1</th>
<th>Very Infrequently 2</th>
<th>Somewhat Infrequently 3</th>
<th>Somewhat Frequently 4</th>
<th>Very Frequently 5</th>
<th>Extremely Frequently 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-grooming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nervous behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambivalent responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze aversion (i.e. avoiding eye contact)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crossed arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illustrators (i.e. hand movements that accompany and illustrate the content of the speaker’s responses)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responses that contain a lot of detail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
INSTRUCTING JURORS ON NONVERBAL CUES OF DECEPTION

Q2
From the list below, which five factors are the most important to you in determining whether someone is truthful when giving evidence in a court case? Please rate them in order of importance from 1 (most important) to 5 (least important).

- Confidence
- Pauzes
- High pitched voice
- Fidgeting
- Restless behaviour
- Implausible responses
- Postural changes
- Speech disturbance/fllers (i.e. “ah’s” and “um’s”).
- Anxious behaviour
- Self-manipulators (i.e. hand movements to relieve a tickly need like scratching an itch)
- Short response
- Murmuring
- Curling the mouth
- Body/hand/fingerleg/foot movements
- Stuttering
- Interrupting the questioner
- Responses that lack logical structure
- Pressed lips
- Repetition of certain words and phrases
- Blinking
- Responses that contain very few details
- Tenseness
- Self-grounding
- Nervous behaviour
- Ambivalent responses
- Gaze aversion (i.e. avoiding eye contact)
- Crossed arms
- Illustrators (i.e. hand movements that accompany and illustrate the content of the speaker’s responses)
- Delayed responses
- Stereotyped responses
- Responses that contain a lot of detail

Q3
I am confident in my ability to determine whether someone is lying or not.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

End of Questionnaire.
Appendix C: Experimental Stimuli

1. Judicial Written Direction

Judicial Direction

Your Honour: Members of the jury. You have heard the various evidence in this case. While the weight that you give to the views is, of course, a matter for you I am going to talk to you, in a general sense, about the weight that may be given to the evidence. You should consider the demeanor of the witnesses in addition to what they had to say in response to questions asked of them. That is, you should consider the way and manner in which the witness gave evidence in addition to what they said.

I now isolate and identify for your benefit, the following matters of significance, particularly with regards to demeanor, which the Courts generally recognise:

- Liars tend to speak in a higher pitch relative to their normal speaking voice when they are not telling the truth.
- Liars make fewer movements with their hands, arms and fingers.
- Liars use fewer illustrators (hand and arm movements to demonstrate and illustrate points they are making).
- Liars take longer pauses.
- Liars’ answers are less plausible.
- Liars’ stories contain fewer details.
- Liars give more indirect answers.
- Liars’ answers contain less temporal, perceptual and spatial information.

The presence or absence of some or all of these factors might reasonably, depending of course on your own view, be regarded as undermining the reliability of the evidence of the witnesses.

2. Expert Witness Condition

To view video link click [HERE]

3. Judicial Oral Direction Condition

To view video link click [HERE]

4. Expert Witness and Judicial Direction Condition

To view video link click [HERE]