



# University of HUDDERSFIELD

## University of Huddersfield Repository

Ng, Magdalene Y.L.

Innocence and Guilt Detection in High-Stakes Television Appeals

### Original Citation

Ng, Magdalene Y.L. (2016) Innocence and Guilt Detection in High-Stakes Television Appeals. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/32621/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

# **Innocence and Guilt Detection in High-Stakes Television Appeals**

---

**MAGDALENE Y. L. NG**

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

SEPTEMBER 2016

## COPYRIGHT STATEMENT

---

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “copyright”) and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian.

This page must form part of any such copies.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property right and/or Reproductions.

**PILATE SAITH UNTO HIM, “WHAT IS TRUTH?”**

*John 18:38, King James Version*

## ABSTRACT

---

The present thesis explored the cognitive and affective mechanisms underlying the cues used to make innocence-guilt decisions in the high-stakes situation of television appeals in which people appeal publicly for the return of a loved one.

Two aspects of the processes involved in making judgements of veracity were studied. Part 2 examined the interaction between explicit and implicit judgments. The studies in Part 2 experimentally manipulated aspects of 12 appeals to test hypotheses about the heuristics judges used to determine veracity. The hypotheses in these studies were initially based on the central assumption that in the absence of unambiguous information judges will draw on heuristics to make veracity judgments. These cognitive shortcuts were hypothesised to lead to biases in innocence-guilt judgments. The innocence bias was introduced as a potential predisposition that stood to be tested. Results did not reveal a consistent innocence or guilt bias. Rather, while all four experiments in Part 2 indicated the presence of different underlying cognitive processes across all experimental conditions, the results from these studies would appear to challenge the existence of any intrinsic tendency towards biases.

In Part 3, the context of the appeals was taken as the basis for the assumption that truthful people would give clearer indications of grief than ones who were lying. Multivariate cues were analysed simultaneously using 39 appeals, with a theoretical basis drawn from the grief literature. Eight previously unidentified aspects consisting of verbal cues drawn from grief literature are found to distinguish honest and deceptive appeals with high accuracy and reliability.

The work thus contributed to the initial understanding of the interaction between explicit and implicit decisions in making innocence-guilt judgments. Standing models of cognitive processing and their implications for the present thesis were also discussed. Contingent upon further clarification of cognitive processes involved during innocence and guilt verdict decision-making, the findings are particularly germane to the area of televised press conferences and have implications for police and practice.

## DEDICATION

---

For two special people, both who passed on during the writing of this thesis:

To my Pa who left in the first year of my doctorate journey – I am forever grateful to you for believing in me at a time when no others did and always leading me to always try be the best of the best. When I first told you I wanted to pursue a Masters in Forensic Psychology in New York, you raised an eyebrow but never stopped me. When I then told you I wanted to pursue a PhD in the UK you raised *both* eyebrows, but made it possible. When I shared with you what my research was going to be about, I will never forget the mixture of contentment and joy (you tried your best to hide them of course, but leakages ensured I saw them) a father wears on his face for his daughter. This thesis is now finished; another one to add to your library of book collection (if you deem it fit).

To my now late aunt who left in the last year of my doctorate journey – Nachatiram, you were and will always be a star in my universe when no other light shone, true to the meaning of your name. You were so happy to know that I was so close to the finish line before you went. This thesis is also for you. You can read it in turns with Pa.

I love you both.

## ACKNOWLEDGEMENTS

---

I would like to express my highest gratitude to my supervisors David and Donna. Both of you have been such great mentors to me, directly and indirectly. I would like to thank the both of you for your selfless time and care in encouraging my research, guiding the writing of my thesis and allowing me to grow as a researcher. It was by chance that I came to be under your supervision.

I was a Research Assistant for a PhD student in John Jay College of Criminal Justice when I confided in them as to the possibilities of doing a doctorate. It was through her that I had heard about the pair of you. I did my homework, and found there was an intake for a doctorate program in October 2013. A few months after submitting my proposal for which I was accepted upon (not in lie detection, but in eyewitness testimony), I'd moved from New York City (that never sleeps) to the (sleepy) town of Huddersfield and here my doctorate journey began with the both of you as my supervisors. On the very first day of my doctorate commencing, I was asked *'There's this opportunity to work with television appeals... do you want to research on lie detection instead?'*

Thank you, for asking me that question.

I thank my fellow Ph.D. colleagues for all the stimulating discussions (and guidance from senior colleagues) you have all provided me. I deeply value all your company, your jokes, your entertaining stories, your uplifting words and the motivation that you have in some way or other afforded me throughout the past few years. Without all of you, the journey would have been *that* much more tedious.

I sincerely thank all the MSc in Investigative Psychology students involved in the data collection portion of this thesis, in collaboration as part of a module I directed in 2014-2015. Thank you for also providing me with an experience in teaching a course and grading your final essays. Without all of your hard work in collecting data, my data pool would have been significantly smaller. In no particular order, thank you to Pheobe, Helene, David, Sharon, Sofia, Aneka, Anita, Karine, Samantha, Maria, Loren, Nicole, Costas, Myrwan, Bryan, Dara, Eirini, Kelly, Andreas, Elena, Megan, Christos, Holly, Dale, Geoff, Kirsty K., Kirsty L., Pakistan and Chiara. A huge thank you for being involved in the data collection for Studies 1 to 4.

I also express my gratitude towards Leanne ten Brinke for generously disclosing to me her list of television appeals so I did not need to start procuring them from scratch. Steve Penrod, thanks for clueing me in to the fact that Leanne had this list to begin with, during a random chat at my very first conference back in 2013.

To my dearest and closest friends both far and near in all five continents (alas, I know of no one in Antarctica), you know who you are – I am infinitely grateful for all your emotional support, belief in my capabilities and tireless encouragement along the way, both in person and through virtual means. Who else can make me laugh that way but the bunch of you, virtual or not?

I thank my band members, who in the past year have offered me an artistic outlet (indispensable during the writing of a thesis!) and an experience I will truly never forget.

To my part-time work colleagues, thanks for providing me with mindless banter. A refreshing welcome and all I can handle some nights after long days thinking about the 'hard stuff'.

Thank you, Ma, for cheering me on from the side-lines during the writing of this thesis; offering me water when I was thirsty during the marathon and picking me up when I was running low on energy to finish the race. Thank you, Godma, for doing the same.

To Pa, I really would not have been able to even start this PhD if it were not for you. You were the best father I could have ever asked for; you adopting and raising me as your own unquestionably changed my life's trajectory. I am eternally indebted.

Alasdair Sykes – thank you, for being there from when this thesis was in its conception to the very last word written in it. My utmost gratitude goes to you.

# TABLE OF CONTENTS

---

## PART 1: OVERVIEW

<b>Chapter 1 Detecting veracity: Issues unexplored</b>	<b>pg</b>
1.1 Introduction	21
1.2 Assessments of truths and lies	22
1.3 Assessments of innocence and guilt	25
1.4 Relationship between cues of lies and truths	25
<b>Chapter 2 Influences in veracity judgments</b>	
2.1 Introduction	33
2.2 Source (A)	33
<i>2.2.1 Source characteristics</i>	33
<i>2.2.2 Source presentation</i>	34
<i>2.2.3 Source content</i>	37
2.3 Judges (B)	38
<i>2.3.1 Truth and lie bias</i>	39
<i>2.3.2 Innocence default and bias</i>	39
<i>2.3.3 Meta-emotion bias</i>	42
2.4 Judgments (C)	
<i>2.4.1 Implicit veracity assessments</i>	43
<i>2.4.2 Conscious vs. unconscious information processing</i>	45
<b>Chapter 3 Relationship between Cues of Lies and Truths</b>	
3.1 Current content analysis methods	48
3.2 Conclusion	50
<b>Chapter 4 Implications for the present thesis and outline for chapters</b>	
4.1 Innocence-guilt judgments using high-stakes scenarios	52
4.2 Considering norms	52
4.3 Addressing biases and heuristics	53
4.4 Explicit and implicit decision measurements	54
<i>4.4.1 A measure of meta-emotion</i>	54

4.5 Relationship between variables	55
4.6 Thesis aims	56
4.7 Organisation of studies	57
4.8 Outline of Studies 1 to 4	58
4.8.1 <i>Study 1: The perception of emotions in appeals</i>	58
4.8.2 <i>Study 2: The perception of Source presentation in appeals</i>	59
4.8.3 <i>Study 3: The perception of story sense in appeals</i>	59
4.8.4 <i>Study 4: The perception of paired versus solo appealers</i>	60
4.9 Study 5: Objective Verbal Indicators of Truthful and Deceptive Television Appeals	61
4.10 Ethical approval	61

## **Chapter 5 Introducing the stimulus material: Televised press conferences**

5.1 Why are television appeals chosen over other stimulus material?	65
5.2 Existing research using television appeals	68
5.3 Ground truth	70
5.4 Length of stimulus material	72

## **Chapter 6 Construction of questionnaire items**

6.1 Defining the ‘implicit’ nature of question items	74
6.2 Questionnaire measures and items	75
6.2.1 <i>Credibility</i>	76
6.2.2 <i>Cognition of Source content</i>	76
6.2.2.1 <i>Story sense</i>	76
6.2.2.2 <i>Story plausibility</i>	77
6.2.2.3 <i>Well-thought-out</i>	77
6.2.2.4 <i>Organised</i>	78
6.2.2.5 <i>Wording</i>	78
6.2.2.6 <i>Practice</i>	79
6.2.3 Emotional perception of Source	79
6.2.4 Attractiveness/appearance perception of Source	80

6.2.5 Meta-emotion	80
6.2.6 Supplementary questions	82
6.3 General methodology across Studies 1 to 4	83

## PART 2: EXPERIMENTAL STUDIES OF SUBJECTIVE CUE PERCEPTION

### **Chapter 7 Perception of emotions in television appeals (Study 1)**

7.1 Perception of emotional displays	89
7.2 Meta-emotion bias	91
7.3 Attractiveness/appearance bias	92
7.4 Innocence bias	92
7.5 Cognition of Source bias	93
7.6 Key issues to explore	93
7.7 Methods	
7.7.1 <i>Participants</i>	94
7.7.2 <i>Material and Procedure</i>	94
7.7.3 <i>Ground truth</i>	98
7.8 Results	
7.8.1 <i>Participants</i>	99
7.8.2 <i>Explicit veracity judgments</i>	99
7.8.3 <i>Relationship between level of appeal emotionality and explicit accuracy rates</i>	90
7.8.4 <i>Implicit veracity judgments</i>	100
7.8.5 <i>Logistic regressions analyses</i>	101
7.8.6 <i>Not-sure answers frequencies</i>	102
7.9 Discussion	105

### **Chapter 8 Perception of Source presentation in television appeals (Study 2)**

8.1 Presentation format	110
8.2 Cognitive loading	112
8.3 Attractiveness and appearance bias	113

	11
8.4 Innocence bias	114
8.5 Meta-emotion bias	114
8.6 Cognition of Source content bias	115
8.7 Key issues to explore	115
8.8 Methods	
<i>8.8.1 Participants</i>	116
<i>8.8.2 Material and Procedure</i>	116
<i>8.8.3 Ground truth</i>	118
8.9 Results	
<i>8.9.1 Participants</i>	119
<i>8.9.2 Explicit veracity judgments</i>	119
<i>8.9.3 Relationship between source presentation and explicit accuracy rates</i>	119
<i>8.9.4 Innocence bias</i>	119
<i>8.9.5 Implicit veracity judgments</i>	122
<i>8.9.5.1 Honest versus false appeals</i>	119
<i>8.9.6 Logistic regression analyses</i>	125
<i>8.9.7 Not-sure answers frequencies</i>	130
8.10 Discussion	130
<b>Chapter 9 Perception of story sense in television appeals (Study 3)</b>	
9.1 Definitions of coherence	136
9.2 How does Coherence correlate with Judgements of Plausibility and Story sense	136
9.3 Judges' Cognitive perceptions of Story Content	138
<i>9.3.1 Rehearsal and Structure</i>	140
9.4 Sender detectability	141
9.5 Other implicit cues	141
9.6 Key issues to explore	141
9.7 Methods	
<i>9.7.1 Participants</i>	142
<i>9.7.2 Material and Procedure</i>	142

	12
9.7.3 <i>Ground truth</i>	144
9.8 Results	
9.8.1 <i>Participants</i>	145
9.8.2 <i>Explicit veracity judgments</i>	145
9.8.3 <i>Relationship between level of story sense and explicit accuracy rates</i>	145
9.8.4 <i>Innocence bias</i>	145
9.8.5 <i>Implicit veracity judgments</i>	149
9.8.5.1 <i>Honest versus false appeals</i>	149
9.8.6 <i>Logistic regressions analyses</i>	152
9.8.7 <i>Not-sure answers frequencies</i>	156
9.9 Discussion	156
<b>Chapter 10 Perception of paired versus solo appealers in television appeals (Study 4)</b>	
10.1 The need for studying perception of paired liars	160
10.2 Lack of studies in perception of paired liars	162
10.3 Cognitive loading	166
10.4 Pair/contagion bias	167
10.5 Other biases in perception of paired liars	167
10.6 Key issues to explore	167
10.7 Methods	
10.7.1 <i>Participants</i>	168
10.7.2 <i>Material and Procedure</i>	169
10.7.3 <i>Ground truth</i>	169
10.8 Results	
10.8.1 <i>Participants</i>	170
10.8.2 <i>Explicit veracity judgments</i>	170
10.8.3 <i>Relationship between explicit rates and source presentation (Paired or Solo)</i>	171
10.8.4 <i>Innocence bias</i>	171
10.8.5 <i>Implicit veracity judgments</i>	171

10.8.6 <i>Logistic regression analyses</i>	173
10.8.7 <i>Not-sure answers frequencies</i>	178
10.9 Discussion	178

## **Chapter 11 Summary of results from experimental studies**

11.1 Observers' explicit judgments	183
11.2 Explicit veracity judgments across Studies 1 to 4	184
11.3 Guilt bias across Studies 1 to 4	187
11.4 Implicit judgments across all studies	188
11.4.1 <i>Overall credibility</i>	189
11.4.2 <i>Emotional bias errors</i>	189
11.4.3 <i>Appearance bias error</i>	190
11.4.4 <i>Cognition in coherent speech bias error</i>	191
11.4.5 <i>Pair or contagion bias</i>	191
11.5 Effect sizes across Studies 1 to 4	192
11.6 Validity across Studies 1 to 4	
11.6.1 <i>Internal validity</i>	194
11.7 Discussion	194
11.7.1 <i>Suggestions for improvements in questionnaire</i>	199
11.7.2 <i>Predictive value of questionnaire</i>	199
11.7.3 <i>Suggestions for future analyses</i>	201
11.7.4 <i>Advantages of the current questionnaire</i>	203
11.8 Conclusion	205

## **PART 3: OBJECTIVE VERBAL CUE ANALYSIS**

### **Chapter 12 Objective Verbal Indicators of Truthful and Deceptive Television Appeals (Study 5)**

12.1 Verbal cues to deception in high stakes conditions	212
12.2 Language aspects of grief and related emotional states	212
12.2.1 <i>Language in assessing psychosocial and emotional changes post-loss</i>	213

12.3 Language areas in assessing changes post-loss	214
<i>12.3.1 Changes in psychological intimacy</i>	215
<i>12.3.2 Changes in psychological distancing</i>	215
<i>12.3.3 Changes in emotional release or reserve</i>	215
<i>12.3.4 Social changes</i>	217
<i>12.3.5 Movement-related changes in speech</i>	218
<i>12.3.6 Tense changes</i>	218
<i>12.3.7 Grammatical changes</i>	218
<i>12.3.8 Chronological changes</i>	218
12.4 The current study	220
12.5 Method	
<i>12.5.1 Developing Content Variables</i>	220
<i>12.5.2 Video selection and ground truth</i>	220
12.6 A Partial Order of Genuine and False Appealers with Base Coordinates (POSAC)	226
12.7 Results	
<i>12.7.1 Differences between genuine and false appeals</i>	231
<i>12.7.2 Results of POSAC</i>	231
<i>12.7.2.1 The X-axis: Movement</i>	232
<i>12.7.2.2 The Y-axis: Emotional</i>	237
<i>12.7.2.3 The J axis: Appeal veracity</i>	238
12.8 Discussion	240
12.9 Conclusions and implications	241
<b>Chapter 13 General discussion and conclusions</b>	
13.1 Overarching theoretical argument and development	245
13.2 Grief theories as a context in studying deception	250
13.3 Perceptions and expectations of grievors	251
13.4 Implicit and explicit decisions	253
13.5 Verbal analysis	253
13.6 Methodological limitations	256
13.7 Suggestions for future research	260

13.8 Practical applications and implications

271

13.9 Coda

273

## TABLES

<i>Title</i>	<i>Page</i>
1 Judgment rates for appeals for High and Low Emotionality conditions	99
2 Wilcoxon's Rank implicit assessments between High and Low Emotionality conditions	103
3 Cronbach's alpha coefficients for Study 1	101
4 Bootstrapped binary logistic regression models between High and Low Emotionality conditions	104
5 Judgment rates for True and False appeals for Audio-only and Audio-visual conditions	120
6 Cronbach's alpha coefficients for Study 2	120
7 Mann Whitney U implicit assessment for Audio-only and Audio-visual conditions	121
8 Mann Whitney U item scores for false appeals between Audio-only and Audio-visual conditions	123
9 Mann Whitney U item scores for true appeals between Audio-only and Audio-visual conditions	124
10 Bootstrapped binary logistic regression models for explicit veracity judgments between Audio-only and Audio-visual conditions	128
11 Bootstrapped binary logistic regression models for accurate explicit veracity judgments between Audio-only and Audio-visual conditions	129
12 Judgment rates for true and false appeals for High and Low Sense conditions	146
13 Mann Whitney U item scores between High and Low Sense conditions	148
14 Cronbach's alpha coefficients for Study 3	149
15 Mann Whitney U item scores for true appeals between both conditions	150
16 Mann Whitney U item scores for false appeals between both conditions	151
17 Bootstrapped binary logistic regression models for explicit judgments in High and Low Sense conditions	154
18 Bootstrapped binary logistic regression models for accuracy of explicit judgments in High and Low Sense conditions	155

19	Number of appellars across UK, USA, Canada, New Zealand and Australia cross-tabulated with their appeal styles	165
20	Judgment rates of explicit decisions for Paired and Solo conditions	170
21	Cronbach's alpha coefficients for Study 4	172
22	Mann Whitney U implicit assessments for Paired and Solo conditions	172
23	Bootstrapped binary logistic regression models for explicit judgments between Paired and Solo conditions	176
24	Bootstrapped binary logistic regression models for Accuracy between Paired and Solo conditions	177
25	Overall explicit rates by conditions across Studies 1 to 4	184
26	Differences between conditions for Overall Credibility across Studies 1 to 4	189
27	Summary of effect sizes for significant implicit cues across Studies 1 to 4	193
28	Cronbach's alpha coefficients for each scale items	194
29	Significant implicit judgments used by judges across Studies 1 to 4	208
30	Significant accurate implicit judgments used by judges across Studies 1 to 4	209

## FIGURES & APPENDICES

---

	<i>Figure title</i>	<i>Page</i>
1	A framework in identifying influences in veracity judgments	29
2	Relationship between explicit and implicit judgments and the associated underlying mental processes	197
3	Example of a path model with exogenous variables, endogenous variables and disturbance terms	203
4	<i>X</i> axis on POSAC solution	230
5	<i>Y</i> axis on POSAC solution	230
6	<i>J</i> axis on POSAC solution	231
7	Two-dimensional POSAC solution	235
7a	Partitioned item diagram for the aspect of movement with frequencies of profiles	236
7b	Partitioned item diagram for the aspect of emotions with numbered frequencies of profiles	237
7c	Partitioned item diagram for appeal veracity with frequencies of profiles	238

## APPENDICES

Appendix A Content dictionary of variables	313
Appendix B Consent form	315
Appendix C Debrief form	316
Appendix D Question sheet	317
Appendix E Question sheet for audio-only condition in Study 2	318
Appendix F Demographic sheet	319
Appendix G Sample answer sheet	320
Appendix H Frequency of evidence types used to establish ground truth	321
Appendix I Not-sure percentages reported by judges for the High and Low Emotionality conditions	322

Appendix J Not-sure percentages reported by judges for the Audio-only and Audio-visual conditions	323
Appendix K Not-sure percentages reported by judges for the High and Low Sense conditions	324
Appendix L Not-sure percentages reported by judges for the Paired and Solo conditions	325
Appendix M Hypothesised cues for genuine and deceptive appeals by aspects	326
Appendix N Empirical support for Eight Language Areas	328
Appendix O Direct and indirect appeals according to appeal veracity	329
Appendix P Monotonicity coefficients between variables with POSAC map axes <i>X</i> and <i>Y</i>	330

**Total thesis word count (excluding appendices) – 77,246.**

# PART 1

---

## OVERVIEW

# CHAPTER 1

---

## DETECTING VERACITY: ISSUES UNEXPLORED

## 1.1 INTRODUCTION

Longstanding research has repeatedly focused on identifying behavioural cues of truth-tellers and liars, with inconclusive results. Research in lie detection has established that individuals are poor lie detectors (Kraut, 1980; Bond & DePaulo, 2006). Yet, less is known of how people perform when making innocence-guilt judgments. The current thesis focuses on implicit judgments relating to innocence-guilty verdicts and the potential usage of heuristics which may lead to biases and errors, to better understand underlying cognitive processes during this type of decision-making. When asked to make this type of judgment, an innocence or guilt bias and even an innocence default may be possible. Questions such as ‘what encourages possible errors and has this any association with source saliency’ has not yet been studied where a specific context is considered.

The process of systematic information processing and decision-forming, information can be attended to, comprehended, reacted to and accepted. During this process, receivers may resort to heuristics especially when the information source holds uncertain outcomes and with only limited information available to judges (Bond & DePaulo, 2006). This may lead to biases in searching from the judge’s own memory, experiences, stereotypes or external stimuli from the information source itself, either through a systematic and thorough processing of the message from the source, and/or through superficial, automatic judgments. If the amount of information and presence of valid cues in a context-specific stimulus material given to judges can be controlled and manipulated across different conditions, in which of these conditions will judges be more likely to resort to heuristics and which heuristics may these be?

It would make logical sense that the concepts of innocence and truthfulness (together with guilt and deceit) would work in concert, at least in some cases. Kassin and Fong (1999), in their paper on judgments of truth and deception during interrogation, discussed in length the extent to which one can differentiate suspects who are guilty and suspects who are innocent. As did Inbau, Reid and Buckley (1986), the authors in this paper interchangeably discussed truth-deception concepts from general lie detection research with that of innocence and guilt, because if the suspects were guilty they would very likely be lying in their

statements and vice versa. Link between innocence-guilt and truth-lie concepts are evident in Granhag and Hartwig's (2008) paper, where the authors discussed frameworks concerning 'the psychology of guilt' and 'the psychology of innocence', and how understanding these can potentially improve one's ability to detect lies and truth. The authors maintained that someone who is guilty will differ in the information they know about the crime from someone who is innocent and may differ in self-regulating behaviour. While there would appear to be a link, certain researchers are careful to underline their distinction between these two concepts. They cite examples of cases where innocent suspects can lie too (Sukumar, Wade & Hodgson, 2016). Nevertheless, lie detection literature is still relevant to the present thesis. As many of the theoretical underpinnings are also pertinent and applicable to innocence-guilt judgments, they will be discussed.

The present thesis explored the cognitive and affective mechanisms underlying the cues used to make innocence-guilt judgements using the high stakes situation of television appeals in which people appeal publicly for the return of a loved one. The specific context of televised appeals was also taken as a basis for providing an objective continent in relation to theories and models of grief, which will be expounded in a later chapter. This chapter will focus on current findings in lie detection research as well as literature on innocence-guilt judgments.

## **1.2 Assessments of truth and lies**

Time and again, studies of lie detection show that judgements are not much better than the toss of a coin (Bond & DePaulo, 2006). Consistently shown in several studies is that observers often use invalid cues to distinguish lies from truth, which partially explains poor accuracy rates. Vrij, Semin and Bull (1996) and Sporer and Schwandt (2007) underlined the difference between observed 'objective' behaviours of liars and truth tellers, and in contrast the subjective beliefs of these objective behaviours. Using unreliable criteria for making veracity assessments can lead to wrong decisions, such as an inaccurate belief that 'liars always do not tell stories in rich detail'.

The usage of invalid cues by observers seems to be well noted in deception detection literature (Anderson, DePaulo, Ansfield, Tickle & Green, 1999; Vrij, 2008). Zimmerman, Compo and Carlucci (2013) asserted that poor deception detection ability and accuracy may relate to common misconceptions about cues to deception. Hartwig and Bond (2011) discussed that the actual behaviours that liars display are inconsistent with people's beliefs about what these behaviours are. They also argue that self-reports cannot adequately capture the cues that people use in making veracity judgments because people may be unaware of the basis for their decision making. Vrij, Granhag and Porter (2010) introduced what is known as the 'Othello error' where certain nonverbal cues such as nervousness are used wrongly to indicate that a person is lying, when the interviewee could simply be fearful for being wrongly accused of committing a crime he or she did not do, or in fear of not being seen as truthful.

So far it has been put forward that the two main reasons that human judges are poor in detecting deception is because they tend to use undiagnostic cues while doing so, and because there is a lack of diagnostic cues of deception. (Anderson et al., 1999; Vrij, 2008). Conversely, in a recent meta-analysis Hartwig and Bond (2011) found that this is rather due primarily to weaknesses of these behavioural cues itself. Their meta-analyses would suggest that individuals seldom rely on invalid cues, counter to earlier explanations and suggestions for these poor results.

Furthermore, some researchers have also claimed that lie detectability in high-stakes conditions will be different from laboratory simulations (Cohen et al., 2010; Porter et al., 2008; Ekman, O'Sullivan & Frank, 1999; Miller & Stiff, 1993; Granhag & Strömwall, 2004; Porter & ten Brinke, 2008; Frank, Paolantonio, Feeley & Servoss, 2004; Carlucci, Compo & Zimmerman, 2013). Low-stakes scenarios include trivial lies or lies that are not of a critical nature with critical penalties, typically set within a laboratory condition or university setting restricted by ethical concerns (Cohen, Beattie & Shovelton, 2010; Porter et al., 2008; Vrij, Akehurst, Soukara & Bull, 2004; Vrij, Edward, Roberts & Bull, 2000). It was posited that the cues liars exhibit will be more discernible in a high-stakes situation than in a low-stakes one (Ekman, O'Sullivan & Frank,

1999). Miller and Stiff (1993) proffered that the higher the stakes, the higher the chance more reliable deception cues will be elicited where deceit will be more likely to be detected.

Contrary to these propositions, Hartwig and Bond's (2014) recent meta-analysis found that even in higher stakes scenario truth-lie detection did not fare much better or easier. Using a sample of over 9000 liars and truth-tellers, the authors found little difference in how detectable lies were across different scenarios. These scenarios include statements moderated by level of sender's motivation (i.e., sender is unmotivated or highly motivated) and highly emotional statements. This meta-analysis would show that in the latter scenario mentioned, high-stakes lies were equally detectable to inconsequential lies. Vrij and Granhag (2012) also argued against the belief that cues to deception will be more distinct in high-stakes situations. They critiqued O'Sullivan, Frank, Hurley and Tiwana's (2009) review citing that the selection of 13 'high-stakes' studies and 18 'low-stakes' studies in this work is dubious. Vrij and Granhag (2012) commented that the O'Sullivan (2008) paper included in this selection of work was only a commentary with little to no information given on how the experiment cited was conducted, and it was not peer-reviewed. Furthermore, Vrij and Granhag (2012) stressed that because it is arguable that in a high-stakes situation not only does this affect a liar's behaviour but it could be rationalised that someone telling the truth will act in a comparable way, hence there will be little difference between the cues exhibited by liars and those by truth-tellers.

### **1.3 Assessments of innocence and guilt**

As put forward in the opening paragraphs of this thesis, because decision-making for innocence-guilt judgments are reasoned to be comparable yet distinct from that for truth-lie judgements, literature pertaining to innocence-guilt judgments will be considered separately in this section.

The second paragraph of Article 6 in the European Convention on Human Rights asserts the presumption of innocence until proven guilty by law of anyone with a criminal offence charge (Council of Europe, 1988). Garrett's (2008) review stressed the risks of a false prejudgment of guilt when suspects are in fact innocent

in the judicial system. Kassin, Goldstein and Savitsky (2003) highlighted the threats relating to prejudging a suspect as guilty during police interrogations. When holding a prejudgment of guilt, the authors found that more interrogators felt more compelled to obtain a confession regardless of the actual innocence or guilt of the suspect. Also, interrogators exercised a higher number of interrogation procedures and there was a higher evidence of behavioural confirmation taking place. This in turn affected the suspect's own conduct, as well as the assessments of impartial observers. Kassin et al. (2003) divided this experiment into two parts. In the first part, suspects were either involved in a mock theft paradigm or an innocent act, and were later interviewed by interrogators who were either led to presume a judgment of innocence or one of guilt. This part was audio-taped. In the second part of this experiment, another group of participants listened to these taped interviews and assessed whether the suspects in the mock theft paradigm were innocent or guilty. This group of impartial observers judged 42% of the guilty suspects correctly as guilty, and judged 28% of those who were innocent as guilty. 42% of the interrogators with a prejudgment of guilt judged the suspect as guilty, and only 19% of those with a prejudgment of innocence judged the suspect as guilty. The authors inferred from these results that both interrogators and neutral observers could not distinguish innocence from guilt very well in this experiment.

In general, Kassin et al. (2003) and Inbau et al. (2001) stated that police officers usually decide to instigate an interrogation based on a (potentially inaccurate) a priori belief that a suspect is guilty. Kassin et al. (2003) maintained that once this presumption is formed, police officers may unintentionally pursue, construe or even produce evidence to substantiate this prejudgment, displaying behavioural confirmation bias (Nickerson, 1998; Darley & Fazio, 1980) and a self-fulfilling prophecy (Rosenthal & Jacobson, 1968).

Shoemaker et al. (1973) found that negative and positive facial stereotypes were linked with innocence-guilt judgments in a scenario where evidence given to participants were ambiguous. After being shown a crime vignette, participants were told to choose a photo of a Caucasian white male from an array of 12 photos, who seemed the least probable one to have committed the crime in the vignette. This decision was associated with how participants assigned levels of guilt or

innocence. The authors also found that men tended to rely more on facial stereotypes in assessing guilt or innocence more did women in their sample. This study, however, reveals very little of the ability to correctly judge innocence or guilt when actual veracity of the suspect is considered, what people are thinking when they are making innocence-guilt judgments, and how they arrive at these decisions.

Tamborini, Huang, Mastro and Nabashi-Nakahara (2007) maintained that stereotypes and heuristics are sometimes used by a juror as a cue to help them reach a verdict. Yet, the consequences of using stereotypes, heuristics and biases to determine someone's innocence or guilt are detrimental and improper as a member of a jury in a legal trial. Tamborini et al. (2007) applied the heuristic-analytical model to explore the influence of racial biases and heuristics people have of the judicial system on jury decision-making. They found that people using mental shortcuts relating to race and the legal-system influenced how they assigned innocence-guilt judgments. In particular, Tamborini et al. (2007) also discussed a term they coined the 'presumed innocence', which is defined as a mental shortcut in believing that a defendant is innocent until proven guilty. Their study seemed to substantiate the existence of a presumed-innocent mental shortcut and that it was expected to impact the way one perceives the judicial system and its processes. An example of this is the belief concerning the credibility of witnesses.

If there is a likelihood that jurors employ biases and heuristics when making decisions of innocence-guilt of a sender, researchers in this field would do well to first locate areas where errors may stand to occur. This understanding may perhaps even improve innocence-guilt detection rates, should these rates materialises to be poor. Errors may be made during the judgments and interpretations of source characteristics, arising from the source and/or from the judge. A diagram of this framework is illustrated in Figure 1 below (Ng & Youngs, 2016). Figure 1 is an operational and workable framework that has been systematically reviewed and categorised, applicable to both truth-lie and innocence-guilt decision-making. It is still currently largely unknown how people form impressions and process information when asked to make innocence-guilt

judgments. This thesis aims to better understand the underlying cognitive processes during this type of decision-making, and to uncover errors or misconceptions that judges may incidentally rely on when making such decisions. Expectantly, this will help shed light towards the detection rates in making innocence-guilt judgments. As such, the next chapter discusses errors that may be made during the judgments of a source, which may interrelate with the interpretation of source characteristics, presentation and content.

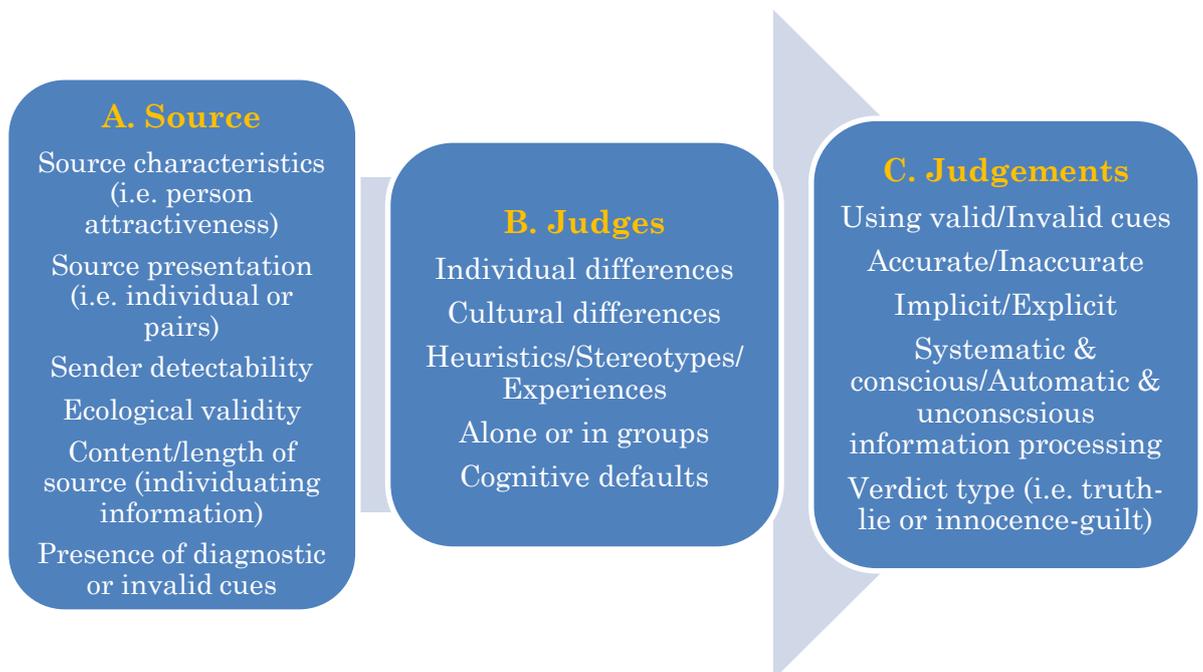
Furthermore, the effects of a prejudgment of guilt or a presumption of innocence is recognised in available literature, but how much of this ‘prejudgment’ is a provisional state of mind and how much of this is perhaps a more stable innate tendency that is difficult to be changed (akin to a Spinozan bias) is largely unknown (Gilbert, Krull & Malone, 1990; Gilbert, 1991; Pennington & Hastie, 1991). This distinguishes the concepts of a presumption applied to the reasoning in having to make a particular veracity decision, from an inherent predisposition in presuming innocence or guilt. Furthermore, more research is needed in determining how neutral observers perform in distinguishing someone who is innocent from someone who is guilty in high-stakes scenarios.

#### **1.4 Relationship between Cues of Lies and Truths**

Another aspect that has not been studied is the relationship between cues of lies and truths in regard to their message content analysis. Whelan, Wagstaff and Wheatcroft (2014) published their work on content analysis of variables in television appeals. While valuable, the authors did not consider analysis of the relationship between the variables they found; in other words, how these cues relate to each other. There have been advocates to using multiple cues rather than focusing on individual variables, and researchers who have acknowledged the difficulty in typifying and computing deceptive behaviours (Leal, Vrij, Mann & Fisher, 2010; Porter and ten Brinke, 2010; Vrij & Mann, 2004).

Duran, Dale, Kello, Street and Richardson (2013) and Van der Zee, Poppe, Taylor and Anderson (2015) studied variables in relation to each other. This presents a method of analysing cues of deception that moves away from frequency counts. An analysis of the relationship between content analysis

variables of lies and truths gives a means of visualising the level of similarity of individual cases of a dataset, as well as similarities between behavioural profiles of truth-tellers in comparison to liars. Compared to linguistic tools, which ultimately rely on statistical differences between variables and can be skewed by factors common in high-stakes lie detection data such as number of cases and low variable occurrence rates, information visualisation based on algorithms within a set of verbal variables in lie detection allows an examination into between-variable distances. This aspect will be further expounded in Chapter 3.



**Figure 1**

*A framework identifying influences in veracity judgments*

## CHAPTER 2

---

### INFLUENCES IN VERACITY JUDGMENTS

## 2.1 Introduction

As briefly discussed in the opening chapter, the general path people seem to use to make veracity decisions is through an inspection of the source (sender or statement). Somewhere between this inspection, the detectability of this source and the various characteristics of it, veracity decisions are formed. The literature review in the current chapter is to establish where errors can be made and understand the pathway in which veracity decisions are formed.

## 2.2 Source (A)

Characteristics of the source itself, how the source is presented and source content can elicit what cues an observer deems important to them, which can in turn cause biases to surface. The following paragraphs methodically discuss several examples that have been pre-empted in literature under this category.

### *2.2.1 Source characteristics*

An aspect that can bias veracity judgments accuracy is characteristics of the source, for instance, the attractiveness and appearance of the source (sender). Stokemer and Praisio (2015) recently found evidence that their participants, who were asked to vote for a political candidate with limited information about the individuals, were inclined to cast their ballots for the better-looking candidate. In contrast, participants who were allotted a detailed account of the contenders' political competence were inclined to choose the more experienced politician. They concluded that participants who were less informed politically used the attractiveness of the candidates as a cue to help them make their decisions.

In the same way, in making innocent-guilt decisions, participants may not at times be privy to information about the sender at their disposal. With limited information and an uncertain outcome, observers may well rely upon the appearance and attractiveness of the source in making their veracity assessments. There are various aspects to how an individual bases their assessment of how attractive someone is, and usually, these assessments are made upon physical appearance and sound (Baron, Branscombe & Byrne, 2008). These interpersonal

attractions are made by the evaluation of a person's attractiveness, to form positive or negative impressions (Hinton, 1993).

It is well documented that the attractiveness of defendants has an effect in courtroom decisions and jury decision-making. Efran (1974) reported that defendants who were more attractive received judgments with less certainty of guilt than their unattractive counterparts. The advantage of possessing good looks seems to continue even after the conviction process into sentencing. Leventhal and Krate (1977) discovered that defendants who were recognised as attractive received shorter sentences as compared to their counterparts, who were given longer terms. Patry (2008) stated that participants in a mock jury study were more disposed to assign a guilty judgment to unattractive defendants compared to attractive ones.

While a wealth of research points to the many advantages of attractive, attractive-sounding, well-appearing individuals, there are a set of opposing findings. These findings assert that being attractive is linked to *less* favourable judgments of credibility (Johnson, Podratz, Dipboye & Gibbons, 2010). In any case, the general implication from the majority of research in social psychology and in criminology shows that one's attractiveness and appearance can impact veracity judgments by means of observers expending this cue in making their judgments. Consequently, it is important to study the role of perceived attractiveness in the present thesis.

### *2.2.2 Source presentation*

How a Source is presented is another aspect that is hypothesised to bias a judge's veracity judgments accuracy. Presenting the source in pairs is an example. While existing research focuses on how groups lie more than individuals for monetary and honesty gains (Cohen, Gunia, Young Kim-Jun & Murnighan, 2009; Sutter, 2009), or cues used by pairs of liars and truth-tellers (Vrij, Mann, Leal & Granhag, 2010), no research has empirically tested whether paired or solo deceptions are more successful at misleading observers. However, Cohen et al.'s (2009) research hinted that groups can appear to be models of trustworthiness (Sutter, 2009). Bornstein, Kocher, Kugler and Sutter (2002) found that groups are

just as trustworthy as individuals in a game involving monetary incentive, thus on par in their ability to deceive observers. Zhou and Zhang's (2008) study on computer mediated communication propose that group size, whether in pairs or threes, mediates deceivers' behaviours.

While most research, such as that of Granhag, Strömwall and Jonsson (2003) found differences in the consistency and preparation of pairs of liars and truth-tellers, subjective perception of these pairs and biases that may occur when judging more than one innocent or guilty person is not yet available in literature.

Rozin and Royzman (2001) posited the 'negativity bias', which they theorised is an intrinsic tendency. Of great interest is the authors' term of negativity dominance, where the combinations of negative and positive entities generate evaluations that are more inclined towards the negative rather than the simple arithmetical addition of both idiosyncratic values. According to the authors this bias infers the ability and potency for a negative or positive entity to contaminate and combine albeit both in dissimilar amounts.

Taken together it is evident here that there is a considerable possibility in advancing research in this area in the context of determining innocence or guilt in high-stakes situations. For example, when a guilty individual who presumably exudes negativity and a sense of 'danger' to beware of, is paired with an innocent individual who presumably emanates positivity, negativity may dominate when both entities combine, whereas an innocent person may be perceived as less truthful via observers using a contagion heuristic and as a result pair bias may arise. In the present thesis, a contagion heuristic is defined as a mental shortcut whereby people identify someone as being of a certain trait or impression simply by close relationship with another person who exudes a similar trait or impression, and/or when both individuals are perceived in combination with each other. In other words, the assessment of Individual A influences how Individual B is perceived merely by close relationship or proximation (i.e., a husband and wife) or when both are perceived in concert with each other. A pair bias may arise as a result of this heuristic (Ng & Youngs, 2016). A pair bias is when an assessment of a person is erroneous due to the employment of a contagion heuristic. As the

example given above, a lying husband may be mistakenly seen as more innocent when paired with his wife who is telling the truth.

In line with this, the presentation format of the source itself is an aspect that may affect veracity judgments. Frank et al. (2004) suggested that a reason as to why lie detection experiments yield poor results is possibly due to the mode of presentation or communication medium. Lie detection studies assessing the accuracy of audio/video versus audio judges indicate contrasting findings. Zuckerman, Amidon, Bishop and Pomerantz (1982) discovered in their research that an individual's voice tone is less well-regulated and therefore easier to observe leakages, in comparison to an individual's face. Bond and Rao (2004) found that individuals in the audio-visual group performed better in terms of accurately judging lies and truths compared to those in the audio-only group, when the speaker was from another culture. Ekman (1989) posited that audio-only cues remain more indicative of the actual veracity of a speaker as observers are not distracted from perceiving these visual unreliable and erroneous elements of deception. Davis, Markus and Walters (2006) established in their studies that in scenarios when observers are presented with an audio-only stimulus for false statements, they were more accurate in their judgments and reported lesser unbiased responses. Ekman (1981) noted that the audio format is a particularly salient mode of communication because we constantly listen to words and counter them as a reply; thus, observers may more habitually and efficiently perceive this channel as compared to non-audio modes. DePaulo, Stone and Lassiter (1985) noted that prediction of honesty is subjective to the medium the participants were shown, whether the target is presented in transcript form, audio-only format where observers merely listen to the stimulus without a visual element, visual-only format where observers only watch the stimulus without any audio recording, and audio-visual format. They found that under high motivation conditions, lies were detected better in channels that included non-verbal cues, namely the audio, visual and audio-visual channels compared to the verbal channel.

On the other hand, Sweeney and Ceci (2014) used an unrehearsed pro-social lying paradigm, where university students and older adults lied

‘realistically’ to defend a research assistant. Again, this scenario was videotaped and shown to another group of participants. The authors found no significant difference between the audio-visual group who achieved an accuracy rate of 68% and the audio-only group who achieved an accuracy rate of 65%.

From most of the studies mentioned in this section, it can be construed that the type of medium presented to a judge would influence their truth-lie accuracy rates (Bond & DePaulo, 2006). Bond and DePaulo’s (2006) meta-analyses revealed that truth-lie judgment accuracy declines when judging visible compared to audible lies. The authors suggested that one of the reasons for this finding is because the typical typecast of someone who is lying is most easily brought out in a visual presentation mode. Additionally, when someone must judge the veracity of a sender with limited information or evidence other than a given video, they do not have many other alternatives but to resort to their pre-existing understanding and stereotypes (Bond & DePaulo, 2006), which may potentially lead to biased judgments. Currently, it is still largely unknown if comparable results will also be revealed for innocence-guilt type judgments.

### *2.2.3 Source content*

The content of a Source is another aspect that can lead to biases in a judge’s veracity judgments accuracy. An example of content of source is how much sense a story makes, which interplays with a judge’s perception of whether it does or not and to what level it does or does not.

In Canter, Grieve, Nicol and Benneworth, (2003) work on narrative plausibility, they examined what they termed ‘narrative plausibility’ by altering the sequence of a story structure in one condition, and keeping the story structure following Stein and Glenn’s (1979) model which follows the ‘Abstract, Setting, Initiating Event, Attempt, Consequence and Reaction’ structure. In the second setting, they reversed the order of the first, taking care that the narrative plausibility of the story in terms of syntax and grammar remained intact. The researchers found that to judge the plausibility of narratives, there are internal and external factors or attributions. The narrative structure is an internal element; it is essentially how the story develops and how it is organised. The

external factors are ‘anchors’, rubrics commonly agreed to be true such as ‘drug addicts are always crooks’. They found that external factors can radically or even subtly change a receiver’s judgment of a speaker’s credibility even though his or her narratives appear ‘internally consistent’.

Following the results from this study, what happens when statements lack internal consistency in the current context? An international study directed to uncover stereotypes of liars conducted by the Global Deception Research Team found that among the top five invalid cues that people report using to judge perceived deception is speaker incoherence (2006, Study 1). A high number of participants from nearly 60 countries incorrectly hold the belief that if a statement is incoherent and does not make sense, it is more likely to be false. There is a lack of research in this area in innocence-guilt detection; evidently there exists a possibility in examining this issue further in a high-stakes context.

### **2.3 Judges (B)**

In examining the possible errors that may occur while judging a source stemming from the judge itself, Tversky and Kahneman (1974) is of interest in initiating this understanding. The authors proposed that decisions will be especially prone to mental shortcuts, known as heuristics, when attempting to cope with probabilistic judgments of an event with uncertain outcomes such as the guilt of a defendant.

Heuristics are used as a resort to reduce probability tasks with high complexities and in predicting these values into simpler judgment processes. While most heuristics serve to save cognitive effort and are generally useful, they can at times lead to inaccuracies in judgment. Tversky and Kahneman (1974) claimed that making a veracity judgment on something as uncertain in its outcome as in determining someone’s innocence or guilt, receivers may make judgments and decisions based on heuristic rules. Mental shortcuts can be viewed as rational and limit the demand on cognitive effort with the advantage that they can be made without full knowledge or information on a case; however, they can also cause deviation and biases. Biases are not the result of cognitive laziness, but rather it ensues from a distorted memory recollection or from external bases of information

that then guides and impacts judgments of the task ahead (Tversky & Kahneman, 1974).

Evans (2007) stated that information can be processed by people either quickly where it does not demand very much of our cognitive reserve (Chaiken, Liberman & Eagly, 1989), or systematically where it demands much more effort from this reserve. Chaiken et al. (1989) proposed that processing information heuristically has a higher tendency to lead to biases, as opposed to systematic processing. Tamborini et al.'s (2007) study was introduced in Chapter 1. The authors applied one of the many heuristic-analytic models in their study, and in this study participants were asked to judge the innocence or guilt of the sender rather than an evaluation of their truth-telling or lying. They used a model outlined in Eagly and Chaiken (1993), Chaiken and Trope (1999) and Chaiken et al. (1989). Hence, it could be said that the same theoretical frameworks that underlie lie detection are pertinent and applicable to innocence-guilt assessments.

The current thesis stresses that to understand how people come to make their judgments, researchers must first consider the context of the decision, as well as the social and cultural norms of the situation to assess why certain behaviours may be judged less favourably. For example, the influence of emotional display illustrated in the section above can be explained by the expectancy violation theory (EVT), which suggests that people hold expectations regarding which behaviours are considered normal in a given social scenario based on personal experiences, cultural and societal norms. When these expectations are violated, individuals who engage in that behaviour are judged less favourably (EVT; Dahl, Enemo, Drevland, Wessel, Eilertsen, Magnussen, 2007; Olsen, Røese & Zanna, 1996).

### *2.3.1 Truth and lie bias error*

Two widely remarked biases in lie detection literature are the truth and lie biases (Levine, Park & McCornack, 1999; Ekman, O'Sullivan & Frank, 1999). These shows at least two types of biases that observers can be subject to in the course of the decision-making process. While the truth bias is defined as the propensity to 'actively believe or passively presume that another person's

communication is honest independent of actual honesty', the truth *default* is a cognitive default in passively drawing assumptions about a statement or sender. Levine (2014, p. 380) defined as having a "passive presumption of honesty due either to a failure to actively consider the possibility of deceit at all or as a fall back cognitive state after a failure to obtain sufficient affirmative evidence for deception". The truth bias is an observable measure (that may or may not be initiated by an unconscious truth default) where the possibility that a statement or sender might be untruthful comes into consciousness when actively prompted (Levine, 2014).

Put differently, the truth default occurs without conscious reflection and can be empirically measured via the observation of a truth bias. Essentially, the truth *bias* can be made with or without conscious reflection (Levine, 2014). Truth bias needs not be a cognitive default, so a Judge can arrive at a truth bias without requiring a truth default; however, a truth default helps explain why a truth bias transpires. If a judge has a truth default, then they may be more inclined towards a truth bias in their veracity judgments. The truth default theory (Levine, 2014) is in direct contrast with the ALIED perspective as the latter claims that there is no such cognitive default, rather, both truth and lie biases arise from the same underlying processes (Street, 2015). Burgoon, Blair and Strom (2008) offered a justification for this bias. The authors stated that individuals usually face a higher number of truthful messages than deceptive ones in everyday life. This results in a pre-set anticipation of honesty and a cognitive bias towards trusting messages and statements.

Levine et al. (1999) repeatedly found in a series of studies that when sender veracity is truthful, in other words when the person is telling the truth, the receiver more frequently tended to correctly judge the veracity because receivers have a bias towards truth (McCornack & Levine, 1990; Levine & McCornack, 2001). In contrast, Ekman et al.'s (1999) findings suggested that police officers tended to demonstrate a lie-bias, instead of a truth bias. They claimed that it is perhaps instinctual that law enforcement officers are in general more distrustful than their lay counterparts due to the nature of their occupation. However, Masip, Garrido and Herrero (2009) contended that although participants indicated a

strong truth bias during briefer conversations due to heuristic judgments, there was a decrease in the truth bias after longer communications were introduced. They suggested that the evidence of truth bias in previous studies of detecting deception could stem from the usage of stimulus material that are too short and those that are drawn from unhelpful behavioural samples.

Vrij and Mann (2001a), Mann and Vrij (2006), Mann, Vrij and Bull (2004) and Vrij, Mann, Robbins and Robinson (2006) did not find evidence of the truth bias in their studies. Street and Masip (2015) recently found that individuals making several within-statement judgments were less likely to be truth-biased. The authors argued that this is not due to a time effect, rather it is the amount of evaluations being made step by step that provides the best explanation for this change.

### *2.3.2 Innocence default and bias*

If, then, someone is asked to make an innocence-guilt judgment, an innocence bias is a possible outcome. As introduced in Chapter 1, while Kassin et al. (2003) found prejudgments or expectations of guilt and innocence it is largely unknown whether there would be an intrinsic tendency for people to presume innocence or guilt. An inherent predisposition to a presumption of innocence would be akin to a truth default, with the latter outlined by Levine (2014). Tamborini et al. (2007) also discussed a term they coined the 'presumed innocence', which is defined as a mental shortcut in believing that a defendant is innocent until proven guilty. Their study seemed to substantiate the existence of a presumed-innocent mental shortcut and that it was expected to impact the way one perceives the judicial system and its processes. An example of this is the belief concerning the credibility of witnesses. An innocence default is defined in the present thesis as a mental default that passively presumes the innocence of a sender or a statement. An innocence default may be possible (with an *innocence bias* being a measurable outcome of this default, that also may or may not have stemmed from this default), which still needs distilling and examination in a high-stakes context. Similarly, a guilt bias in this thesis is defined as a disposition or passive presumption of guilt towards a statement or sender.

Related concepts to an ‘innocence bias’ will be discussed here. One such concept is the ‘halo effect’. The halo effect is a mental bias where a judge’s overall perception of someone influences their feelings and thoughts about that person or product’s character (Thorndike, 1920). The halo effect has been studied in its association with concepts of source characteristics as well as its implications for and in the judiciary system. One such concept is attractiveness (Wade & DiMaria, 2003). Wade and DiMaria’s (2003) study would reveal that how attractive someone is can influence people’s perception of this person’s level of accomplishment and of their character. Likewise, Dion, Berscheid and Walster (1972) found a ‘what is beautiful is judged to be good’ phenomena, where someone who are good-looking are judged to have more favourable and good characteristics. Another similar concept was introduced by Paunonen (2008) who also explored a model of honesty that interrelates with other judgments. In this study, the author found that a sender who is truthful was judged to be more attractive and more likeable.

These studies would suggest an interrelation of other judgments of the person (i.e., attractiveness) together with the ultimate verdict of them, whether this verdict be of character (Wade & DiMaria, 2003; Dion et al., 1972) or political knowledge and sophistication (Palmer & Peterson, 2012) for example. In this way, the concepts of ‘halo effect’ and ‘what is beautiful is good’ are comparable to that of an ‘innocence bias’ in the sense that all are robust concepts relating to person perception. They concern how we think and process information about a person, and interplays with feelings and thoughts are potentially present and influences the ultimate verdict of that person (Aronson, Wilson, Akert, 2010). This interrelation seems to also transpire vice versa. It could also be that people who are judged to be more honest, are then seen as more attractive, for example (Paunonen, 2008). These studies would suggest an interaction between a source (i.e., source characteristics), judges and their veracity judgments.

The story model by Pennington and Hastie (1986, 1988) was put forward to explain how people come to make innocent-guilt judgments. It describes how people use source content (i.e., evidence), for example, as a basis to decide upon a guilty or non-guilty verdict. This model puts forward that people make decisions utilising three steps (Pennington & Hastie, 1993). The first involves forming the

story, the second involves recognising verdict options and the last consists of making a verdict. In the story formation phase, a juror attempts to organise evidence presented to them into an order that makes sense in terms of its content and chronology. They structure the story causally, using what they personally know of similar events such as personal experience, their expectations of what makes a complete story and the evidence presented during trial.

Pennington and Hastie (1988) posited that jurors make decisions by reasoning the evidence presented and constructing an interpretation of this evidence, which then becomes the basis of an ultimate verdict.

Pennington and Hastie (1988) postulated that ‘coverage’, ‘coherence’, ‘uniqueness’ and ‘goodness-of-it’ contribute to the decision-making process in deciding whether a story is satisfactory or otherwise. Coverage denotes how much evidence is justified and explained by a specific story. The more evidence that can be accounted for, the higher the likelihood that the story will be accepted and the more confidence a juror is likely to have in that story. The story model proposed that consistency, plausibility and completeness all determine a story’s level of coherence. If a story has no internal and external contradictions, is parallel with the juror’s pre-held knowledge and contains all parts of what a juror believes makes up a story, the higher the possibility that the juror will accept the story. Jurors then map this causal model of an accepted story onto the verdict options available. The better the story fits onto the verdict options, the higher the juror’s confidence in the verdict-story matched-ness and it exceeding the required level of proof, and the more likely they will choose the verdict. There seems to be no mention of whether people have a predisposition towards awarding an innocence or guilt verdict, however, and this has not been previously studied if there are indeed such biases in Pennington and Hastie’s (1993) story model.

The judicial system in the United States entails a presumption of innocence until proven guilty, ensuring the legal right of a fair and unbiased trial to protect suspects (Helgeson & Shaver, 1990). Helgeson and Shaver (1990) found in a mock trial experiment that their participants did not assume the sender’s innocence. The results from this study would suggest that irrespective of

what 'legal instructions' they were given, participants were still influenced by experimentally-induced biases. An assumption of innocence is where a suspect is to be considered innocent until proven guilty (Mueller & Laird, 2009).

Incidentally, it is important to examine if people have an innocence bias or not, and whether thinking is different implicitly if judges are put under different experimental conditions and how these relates to their explicit judgments of the sender's innocence or guilt.

In the lie detection literature, the truth bias posits that this accuracy increases above chance level only with truthful messages, and hovers around 50% or lower for deceitful ones (Levine, 2014; McCornack & Levine, 1990; Levine & McCornack, 2001). To recapitulate, it also postulates that truths are most often correctly identified as honest, but errors predominate when lies are judged. This thesis uses the same cut-off point in determining the occurrence of an 'innocence bias' as in lie detection literature (Levine, 2014; McCornack & Levine, 1990; Levine & McCornack, 2001), with a rate of above 50% indicating an innocence bias and a rate of below 50% indicating a guilt bias. The absence or presence of an innocence bias (as well as a guilt bias) will carry implications for the psycho-legal arena if can be further developed, supported or challenged.

### *2.3.3 Meta-emotion bias*

The present thesis proposes that this is another possible error that conceivably contributes to lower veracity detection accuracy. Meta-emotion is the degree of sympathetic response *towards* an emotional sender (Granhag, 2006; Ng & Youngs, 2016). This is a person's assessment of their felt emotions, which can bias their cognitive mechanisms while making veracity judgments.

Schwarz (2012) and Srull (1984) postulated that individuals may use their feelings as a source of information particularly when under the circumstances where they consider their feelings as a highly significant source of information and also when more information about the source is not available. In highly emotional scenarios such as press conferences for missing or murdered relatives, observers may falsely deduce that their own felt emotions are a highly relevant component in making veracity assessments, especially when only fractional information is

known about the appellant. In broader terms, an observer may rely on their mood state as a heuristic cue, such as 'If I feel very sad watching the appeal, hence they must be innocent'. As found in criminology literature, jury members may be more likely to seek out cues that back up the emotion they are feeling and misinterpret this unreliable or invalid cue to be consistent with this emotion, subsequently affecting their veracity judgment of the speaker (Salerno & Bottoms, 2010).

## **2.4 Judgments (C)**

### *2.4.1 Implicit veracity assessments*

As illustrated in Figure 1, judges can make a veracity decision implicitly and explicitly. In this thesis, explicit veracity assessments are defined as dichotomous 'He is guilty' or 'He is innocent' choices, which participants are allowed to make after viewing a video clip of a sender (Granhag, 2006). Implicit assessments, in contrast, are where observers are asked to evaluate a sender's statement, appearance, and/or their own emotional reactions towards watching the video clip of the sender for example. In other words, any other questions asked apart from the statement or sender's ground truth may be considered to be implicit judgments (Granhag, 2006).

It should be kept in mind that the idea that there is a considerable implicit factor in what is known as 'implicit lie detection' today is debatable (Granhag, 2006). The author further argued that the available research (while few and far between) does not concern the fact that individuals are better at lie detection by primarily exercising their implicit knowledge, rather, this measure is essentially an approach to facilitate the understanding of individuals' responses to any other question except for 'whether the sender lying or not'.

Some studies show advantages of measuring implicit, rather than explicit veracity judgments. Findings by Vrij et al. (2001), who ran a between-subjects study using police officers as their participants, showed that participants in the explicit group rated liars as somewhat more likely to be lying, on the other hand participants in the implicit group rated whether liars had to think harder than truth tellers and subsequently were more accurate in their implicit judgments. Here it can be seen that more information of cue perception is generated from

implicit veracity assessments rather than a dichotomous rating. Similarly, Landström, Granhag and Hartwig (2005) found in their study that when their participants were asked to make explicit veracity judgments they were not able to discriminate liars from truth tellers with an average accuracy score above chance level (50%) but performed better in implicit ratings, where they rated liars as thinking harder than their truth-telling counterparts and this difference, although not statistically significant, was close to. They also rated honest statements as having richer details, more plausible and convincing compared to deceptive ones. In summary, these studies claim that observers perform better when asked to make implicit assessments of veracity rather than explicit ones as they are able to distinguish certain aspects such as the sender's statement or appearance for example. In contrast, Bond, Levine and Hartwig (2014) found no evidence that indirect lie detection provides better lie or truth discrimination compared to explicit decisions.

The advantages proposed in asking judges to make implicit evaluations are that 1) this is an indirect method of lie detection that is less 'artificial' than asking judges to make an instant explicit judgment (Bond & DePaulo, 2006), 2) regardless of whether it increases accuracy of judges' veracity assessments, it will enhance our understanding of perceptions and consequent biases of observers in lie detection research. DePaulo and Morris (2004) recommended that implicit assessments be favoured instead of a sole measure of explicit judgments. The argument behind examining implicit assessments is to understand the underlying cognitive mechanisms that can help shed light onto low detection accuracy rates.

Both Street and Vadillo (2016) and Granhag (2006) argued that substantiation for *only* an indirect method of lie detection (purely implicit) is limited. However, current lie detection research mostly applies explicit lie detection methods. The biases and heuristics proposed earlier in this chapter cannot be explored under the sole reliance of explicit veracity assessments. The inclusion of both types of assessments will allow heuristics and cognitive operations to be studied. Granhag (2006) asserted that neither type of assessment is better than the other; but that it is important to recognise that judgment is, at times, ultimately the artefact of interaction between both types of assessment.

Granhag (2006) proposed that there are many types of implicit assessments, such as verbal i.e. “truth tellers tell stories that make more sense”, nonverbal i.e. “liars are more tense”, mental process of participant leading to their judgments, outcome of an observer’s perception of sender which is ‘object-level’ i.e. sender is ‘thinking too hard’, result of an observer’s reflections on their own cognition (meta-cognition) or emotions (meta-emotion) i.e. assessing how sympathetic the sender is, or how comfortable the observer was in reaction to the statement, as introduced in Section 2.3.3. Support for an implicit measurement of veracity via meta-emotions is the hot hypothesis approach. Ask and Landström (2010) proposed that the above studies are missing the ‘hot’ component and therefore considered ‘cold’ hypotheses. Thagard (2006) explained the term ‘hot thought’, defined as emotional cognition and held the assumption that a viewer integrates the emotional aspects from a source into their judgment processes. In their study, they investigated not only statement believability, but also participants’ sympathetic response towards rape victim statements and how this mediates their credibility judgments of the victim, the sympathetic response being the added ‘hot’ component. They found that affective display mediated participants’ sympathetic responses and affected their perception of the victim’s veracity.

This sympathetic response is synonymous with the concept of ‘absorption’. Kujipers et al. (2014) compiled previously-researched, similar definitions into four major facets, and comprised them into the Narrative Absorption Scale which will be further expounded on in Section 6.2.5.

#### *2.4.2 Conscious vs. unconscious information processing*

Related to the section above is the concept of conscious and unconscious information processing where judges can make a veracity decision via systematic processing, and/or through a more autonomic route. As introduced earlier, the truth default theory is a cognitive default hypothesised to occur without conscious reflection and can be empirically measured via the observation of a truth bias (Levine, 2014). This theory claims that there is a strong possibility of a pre-selected response bias. If this theory can be proven accurate, then the truth bias stems from a cognitive default that occurs unconsciously. In contrast, the ALIED

perspective claims that there is no such pre-selection or predisposition (Street, 2015). Street (2015) went a step further to claim that so far available empirical evidence for unconscious lie detection is lacking. Street and Vadillo (2016) maintained that where an advantage can be dependably observed in terms of indirect lie detection accuracy, this outcome can be inferred to the conscious and deliberate part (in the absence of an explanation pertaining to the unconscious one).

## CHAPTER 3

---

### RELATIONSHIP BETWEEN CUES OF LIES AND TRUTHS

As presented in Chapter 1, another aspect apart from context-specific heuristics and biases that has not been studied using high-stakes stimulus material is the relationship between cues of lies and truths in regard to their message content analysis. Other than the efforts of Whelan et al. (2014), research on content analysis in lie detection show much potential in the area of lie detection, but lacks an examination of its verbal components and still heavily relies on credibility assessment tools. The following paragraphs focus on what the main credibility assessment tools used are, and their associated problems. It is essential to consider these tools and the principles behind their usage, as much of the rationale behind this is helpful in understanding deception detection.

### **3.1 Current content analysis methods**

Statement Validity Analysis (SVA) is a validity assessment tool which focuses on variables of an account that are correlated with truthfulness (Richard, 2008). The purpose of the SVA is to provide an assessment of the validity of the recorded statement, not of the general credibility of the witness. The hypothesis of SVA is that accounts based on real events, often subject to unintentional errors, would consist of different features than the ones based on a deliberate lie (Vrij & Mann, 2004). Both true and false accounts are affected by stable personal factors as well as situational factors albeit different in substance and source. The content of an account based on a real event may be driven by stress, fear, forgetting, verbal and sensoric deficits as well as inappropriate interview techniques and the one based on a deliberate lie may be driven by motivation, general credibility of the person, personality disorders and plausibility of the lie, or the credibility of the statement as well as the person.

SVA was built on detailed consideration of statement content using criteria-based content analysis (CBCA). While the criteria has been shown to be beneficial in helping to discriminate truthful and false statements (Zaparniuk, Yuille & Taylor, 1995; Akehurst, Köhnken & Hofer, 2004), it stands that this validity checklist can be improperly used because of the difficulties in applying it (Gumpert & Lindblad, 1999). Gumpert and Lindblad (1999) comprehensively studied a Swedish application of SVA, with results revealing individual differences

among ten experts using the statement criteria in SVA and each emphasised different aspects of cases in order to reach their final judgment. In fact, most experts in this Swedish sample used an alternative hypothesis to reach their conclusions. While drawing from a small sample the results from this study draws a red flag towards the issues in using SVA.

The main problems with tools such as SVA are that they lack standardisation and they are sensitive to contextual biases (Bogaard, Meijer, Vrij, Broers & Merckelbach, 2014; Risinger, Saks, Thompson & Rosenthal, 2002) where a set of occurrences shifts an individual's decision threshold by function of expectancies they implicitly generate based on information derived from a context. An example of contextual bias is the confirmation bias, wherein individuals are susceptible to search for evidence that reinforces a pre-existing conviction while disregarding suggestions that contradicts it (Findley & Scott, 2006). The same problems can be said of Scientific Content Analysis (SCAN; Sapir, 2005).

SCAN (Sapir, 2005) maintains that liars and truth tellers will use different language and developed a criterion list to assist in differentiating their statements. Bogaard et al. (2014) argued that there is no concrete evidence for SCAN's capacity to differentiate these supposed differences. The authors also asserted that a further problem is that individuals will differ in the way they use SCAN, which is not standardised. Again, the same problem of individual differences in users can be said for CBCA and Reality Monitoring (RM; Sporer, 1997). Sporer (1997) introduced RM and set eight criteria encompassing aspects such as realism, space and time details, sensory information and vividness. Moreover, RM is not used in criminal investigations unlike CBCA and SVA (Granhag & Vrij, 2005).

Overall, the usefulness of content analysis in lie detection can be found in Vrij et al.'s (2010) work, where the authors maintained that truth tellers use a different verbal strategy as compared with liars. Those telling the truth were more likely to be fixated upon what had transpired and provided more detailed responses, whereas liars would prepare their answers for the questions to come if they could and would choose to deliver replies that were less comprehensive.

In the recent years computer analysis programs have been developed to aid content analysis, showing much potential in lie detection research. Pennebaker, Francis and Booth (2001) developed the Linguistic Inquiry and Word Count (LIWC), which is a computerized method to analyse an individual's use of functional and emotional words. It is essentially a word counting program that computes the ratio of words within several dozen classifications used in a transcript. This program searches for predefined clusters of words that correspond with the different classifications.

Onyebadi and Park (2012) identified common characteristics of 419 scam emails. Using a computer program called Diction 5.0, the authors noted that these messages frequently contain various grammatical errors. Diction 5.0 operates similarly to LIWC, in that it is a computerised word count software with in-built content dictionaries. It analyses five core variables – certainty, optimism, activity, realism and commonality. Both Pennebaker et al. (2001) and Onyebadi and Park (2012) found that liars used different word patterns than truth tellers.

### **3.2 Conclusion**

These abovementioned tools and linguistic software lack a method in studying the relationship between all cues that differentiate behaviours of liars and truth tellers; a new method may offer a more rounded approach in studying deceptive and truthful content. The general findings and summaries from deception detection studies posit that behavioural cues, alone at least, are not useful. Over the past few decades, meta-analytic research in lie detection has shown that behavioural cues to deception tend to be unreliable and weak (DePaulo et al., 2003). The authors examined 158 behavioural cues and found there were no consistent cues that diagnostically distinguished liars from truth tellers, and no cues that reliably associate with lying with very small effect sizes and small numbers of estimates. Most differences were very small, difficult to discriminate in real life situations. Therefore the strength in using a method that provides a visual account of all speech variables is perhaps the much needed advancement from a reliance on individual cues alone.

## CHAPTER 4

---

### IMPLICATIONS FOR THE PRESENT THESIS AND OUTLINE OF CHAPTERS

Following the discussion of literature established in the previous chapters, the current chapter first summarises the key implications and below sets out the aims of the thesis.

#### **4.1 Innocence-guilt judgments using high-stakes scenarios**

Hartwig and Bond's (2014) meta-analyses would seem to challenge longstanding belief in the field of lie detection (Cohen, Beattie & Shovelton, 2010; Porter et al., 2008; Vrij, Akehurst, Soukara & Bull, 2004; Vrij, Edward, Roberts & Bull, 2000; Ekman, O'Sullivan & Frank, 1999; Miller & Stiff, 1993) that low detection rates are partially due to the usage of low stakes scenarios lacking in ecological validity. Nevertheless, little is known of the how people make innocence-guilt verdicts, how accurate they are and if any biases would emerge using televised press conferences of relatives appealing for the return of a loved one who is either missing or murdered.

#### **4.2 Considering norms**

The use of television appeals also underscores an example of a context-specific study, and is discussed further in the next chapter. Within this specific context, considering the norms of the particular context of television appeals is crucial. This gives a unique meaning and direction for the material to be studied that can also offer generalised implications.

Sporer and Schwandt (2007) asserted that before being able to be applied practically, beliefs about observed behaviours need to first be studied and recognised. The notion in the present thesis is that before consideration of what subjective perceptions and expectations held by observers is possible, there is a pressing need to first explore the norm behaviour of a specific situation, in order to understand and compare the two. Television appeals feature individuals who are presumably distraught and grieving over the loss of their loved ones, either temporarily in a missing person's case, or indefinitely in a murder case. Due to the specific nature of using television press conferences in the present thesis and in order to consider the norms of a family member appealing for a missing or murdered case, theories models of grief and their relationship with subjective

perceptions of a griever must be visited. As perceivers make veracity decisions, subjective perceptions that they deem as important (i.e. stereotypes and/or prior expectations) may guide these decisions, especially if not much other information is given to them (Bond & DePaulo, 2006). Fiske and Taylor (1984) proposed that these stereotypes result from dependence on pre-formed beliefs stored in memory. Individuals employing stereotypes remain close to their pre-established information structures when considering incoming data.

### **4.3 Addressing heuristics and biases**

Following expectations, observers may form or fall into particular heuristics other than stereotypes while mentally encoding information when watching a television appeal. At the most basic level of their theory, Tversky and Kahneman (1974) postulated that decisions of guilt or innocence will be particularly susceptible to using mental heuristics because the judgment has an uncertain outcome. Individuals tend to rely on 'rules of thumb' in the face of uncertainty in order to make sense of the given situation (Gigerenzer, Todd & The ABC Research Team, 1999). While Granhag and Strömwall (2004) were keen to study underlying cognitive processes by requesting observers to verify their veracity assessments, all the processes and heuristics identified in the literature have never been properly pooled together and studied using high stakes stimulus material. Little is known of the perceptions and biases of television appealers.

As addressed in Chapter 1, a variety of biases stemming from the employment of mental shortcuts is possible and the general literature presents these biases scattered under a variety of frameworks. Yet none of these have been collectively investigated in the context of television appeals. These biases include (but are not limited to) appearance biases, emotional display biases and the possibility of innocence/guilt biases (Peace & Sinclair, 2012; Paunonen, 2006). Additionally, the innocence bias is also as a measurable outcome of a possible innocence default that stands to be tested in this context.

So far, it has been hypothesised that to distinguish an appealers who is innocent from one who is guilty, one must first study the norms of bereavement,

loss and distress. Observers' perceptions and subsequent biases of grieving television appealers then give way to implicit and/or explicit veracity decisions.

#### **4.4 Explicit and implicit decision measurements**

To observe whether people do indeed think differently when given stimulus material with controlled (different) induced biases, implicit questions can be asked and studied. As explained in Chapter 2, there exists a continuous reliance on only explicit judgments in most lie detection studies, with only a handful of studies investigating implicit judgments in lie detection research. Implicit judgments have been shown to facilitate the understanding of how perceivers think, and offer more insight than pure explicit judgments (Granhag, 2006).

The present thesis recommends that innocence-guilt research should move towards including an implicit evaluation process, not because it is better or worse than its explicit counterpart but because the two processes may work in concert. While this topic is not exactly new, existing research studies tend to isolate explicit and implicit components of lie detection and examine them separately such as in Vrij et al. (2001). This study separated explicit and implicit lie detection where one group of police officers were asked to assess video clips explicitly (whether the suspect is guilty or not guilty) and the other were asked to do so implicitly (whether the suspect seemed to 'think hard'). Other studies that isolated explicit from implicit components in lie detection research include research by Landström et al. (2005), Hurd and Noller (1988) and Anderson et al. (1999).

As an alternative, measuring *both* decisions by the same group of judges can help investigate which implicit cues are prognostic of correct dichotomous explicit judgments. To measure implicit decisions in the current thesis, a questionnaire will be developed and this will be further expanded in Chapter 6.

##### *4.4.1 A measure of meta-emotion*

The measure of meta-emotion is an example of an implicit assessment of veracity. The study of felt emotions in information processing is belatedly recognised as important in studies of veracity judgments, and further substantiated as so in dual process theories (Slovic, Finucane, Peters &

MacGregor, 2007). Vrij et al. (1996) posited that experienced emotions also influence accuracy in veracity assessment. An observer will be stimulated to experience certain emotions when the possibility is raised that deception is at hand. This would be even more so when emotional stimulus material such as television appeals are shown. When the observer experiences these emotions, it is possible that they will closely search for behavioural cues that would reinforce their belief of this 'deception', consequently influencing the accuracy of their veracity judgment. Whether this measure is applicable and extendable to innocence/guilt-type veracity judgments is of interest in the current thesis.

#### **4.5 Relationship between variables**

In missing or murdered relatives' cases, the stakes are extremely high for appellants. By them choosing to deliver a public press conference, this provides a distinctive context to study. This context is often accompanied by unique emotions hypothesised to be attributable to grief and/or anxiety. It provides an avenue for advancing knowledge in the field as little is yet known about the language of liars and truth tellers in high stakes condition.

In conjunction with this, a final proposal in the present thesis is in regard to variability in occurrence of certain cues in particular contexts, specifically cues that may occur but with relatively low frequency. This sometimes results in a failure for such cues to be recorded in published studies at all. A method to study the relationship between all cues that differentiate behaviours of liars and truth-tellers in spite of their frequencies has yet to be employed in high stakes detection deception literature. This way, cues with very low occurrence may also be studied in relation to cues that are higher in occurrence to form a more robust representation of all cues that occur within a particular context. Specifically, multidimensional scaling procedure allows the exploration of relationships between ordered variables, revealing any underlying structure that may surface. It has been widely used for examining structures in ordinal level variables derived from qualitative material. This procedure is further explained and employed in the second part of the present thesis in Chapter 12.

#### 4.6 THESIS AIMS

Considering the issues and implications argued thus far and as acknowledged above, the present thesis will explore the following key issues.

1. As its first aim, this thesis seeks to examine the impact in using ecologically valid and high-stakes stimulus material on the accuracy of judges in detecting innocence and guilt (in experimentally manipulated conditions). The link between explicit innocence-guilt verdict with other implicit judgments is of interest. Key issues and objectives within this main aim are:
  - Will judges perform above chance level if the Source's stakes are high and certain cues are made more salient?
  - In examination of the meta-emotion bias, by experimentally manipulating the level of emotional display at Source level, will judges' implicit and explicit judgments be influenced by this and be less accurate in detecting innocence-guilt in a specific high stakes context?
  - In examining the appearance bias (Source characteristics), will judges be influenced in both their explicit and implicit judgments (Account) by whether they are able to visually see the appellant or not (controlling for Source presentation)?
  - Will judges be influenced by a story sense bias and will this impact their explicit and implicit judgments accordingly?
  - What happens to implicit and explicit decisions when judges are exposed to more than one liar? What happens to these decisions when they are exposed to a liar and a truth-teller at the same time?
  - Will there be innocence bias when judges are exposed to more than one liar?
  - Will judges reveal different underlying cognitive processing across all studies due to saliency of different cues in different conditions? In other words, do different conditions evoke judges to detect different implicit cues?
2. A second aim of this thesis is to investigate implicit veracity judgments in the high stakes scenario of television appeals and its relationship between with explicit judgments. The predictive value of individual implicit judgments towards explicit decisions is also of interest. Key issues and objectives within this main aim are:

- What implicit judgments help judges make accurate explicit judgments? (i.e., will judges show a high degree of meta-emotion towards the appellant, and will this be reflected in their explicit decisions?)
  - What implicit judgments help predict explicit judgments when these explicit decisions are accurate? (i.e., will the judges feel emotions towards the appeal, and will this help them make an accurate explicit veracity decision?)
3. The final aim of this thesis is move away from the subjective perceptions of television appeals and in its place, to examine the actual cues that this high-stakes scenario may expose. Instead of focusing on the invalid cues observers may rely on, the aim is to uncover verbal cues television appellants actually express. There are key issues to explore and objectives within this main aim. This includes utilising a methodological approach that will allow not only using frequency counts in investigating the cues to deception but also in examining cues in relation with each other. The question to pursue an answer for within this aim is ‘Will using forms of multidimensional scaling reveal an underlying structure using the sample of television appellants, with the ability to distinguish genuine from false appeals with a high degree of accuracy?’

#### **4.7 Organisation of studies**

In light of the key issues acknowledged above, the following studies are conducted in two parts. All studies utilise a questionnaire further developed in Chapter 6 that contains both explicit and implicit veracity assessments.

Each study explores a different bias (all elaborated in Chapter 2) by experimentally manipulating the stimulus material. These experimentally manipulated materials exaggerate the differences between each study conditions to their extremities with the aim of revealing judges’ possible reliance on different implicit judgments (and heuristics that leads to response biases). In other words, by carefully manipulating the Source scenarios, these studies aim to observe which biases are more likely to emerge and if any differences in cognitive processes between experimental conditions do in fact exist.

Part 2 will explore subjective judgments and/or biases of judges of the truthfulness via implicit veracity assessments of these appeals, based on theories specifically expounded in each study. This part will investigate the stereotypes and heuristics that pertain to making veracity decisions. Television press conferences are typically very brief, generally approximately 30 seconds to a minute in length. The next chapter introduces the stimulus material in full. Because video clips of press conferences are generally very short, moreover manipulated variables are present in each study, it is expected that judges will resort to heuristics (due to limited information about the appellants) such as meta-emotion and attractiveness especially when the wrong cues are accented. However, when these wrong cues are removed (i.e. removing visual cues in audio-only group or a pair bias in Solo condition) accuracy is hypothesised to be higher.

Not only is the current thesis interested in subjective participant judgments but to also investigate diagnostic cues of liars and truth tellers, and how these cues form in relation with each other. Whether these differences are in fact distinguishable as well as statistically detectable is also of interest. In Part 2, the norms of grieving and theories of grief in general will also be explored. Investigating from the angle of grief theories underlying the current context has not yet been studied in lie detection research. Part 3 departs from perceived judgments, stereotypes and invalid cues judges more than often use, examining what verbal cues these liars and truth tellers reveal using high stakes scenarios.

## **PART 2: SUBJECTIVE CUE PERCEPTION**

### **4.8 Outline of Studies 1 to 4**

#### *4.8.1 Study 1: The perception of emotions in appeals*

While people often use emotions as a cue to decide of whether someone is telling the truth or not, this is not a diagnostic cue (Vrij, 2000). Schwarz (2012) proposed the feelings-as-information theory, which maintains that we pay attention to how we feel and use it to inform ourselves of the circumstances at hand and as a source to make judgments.

Some studies have examined the role of emotion as a heuristic. Ask and Landström (2010) found that an emotional victim can affect credibility judgments as a heuristic with a group of police participants, and the magnitude of this effect was heightened with induced cognitive load. Landström, Ask, Sommar and Willén (2013) found that emotionality of a child shown when reading an account boosts participants' judgments of credibility of that child, finding support for what the authors call the 'emotional victim effect'. Reinhard and Schwarz (2012) found interplay between affective states on the process of detecting deception. The emotion participants feel at the time of the experiment influenced how they judge the truthfulness of incoming source content.

The process behind this bias, where the emotional content is used as a cue to judge the veracity of the story, is relatively absent in innocence-guilt detection literature. By selecting highly emotional appeals and comparing them to appeals with low emotional content, it may be possible to highlight a meta-emotion bias in people having to make an innocence-guilt judgment.

#### *4.8.2 Study 2: The perception of Source presentation in appeals*

While the appearance of television appealers cannot be manipulated, the manner of source presentation can be studied and judges' responses pertaining to appearances can be recorded. Study 2 explores the possible bias in judgments due the availability of a visual element or the lack of it in judging the innocence or guilt of an appealers. It may be possible to highlight a visual bias in this study, if judges display different levels of accuracy in their veracity assessments with and without the aid of visual stimuli. It has been acknowledged that visual cues are particularly unreliable in lie detection (DePaulo et al., 2003). It has also been found that lie detection accuracy is higher in audio-only mode compared to ones with a visual element (Bond & DePaulo, 2006). It is of interest if innocence-guilt detection accuracy will decline, remain at chance level, or increase in an audio-only condition.

#### *4.8.3 Study 3: The perception of story sense in appeals*

Very few lie detection studies have employed manipulation of verbal content in a statement. Reinhard and Sporer (2010), one of the few known studies to have exercised manipulation of verbal content, manipulated the consistency and plausibility of the transcripts in their study. While *story sense* is frequently used by observers as a cue to aid them in deciding a speaker's veracity, there is a lack of research in this area in innocence-guilt detection; evidently there exists a possibility in examining this issue further in a high stakes context. The implicit judgment of whether a story makes sense versus when lacking sense, also lacks research in this area. This main aim of Study 3 is to investigate if story sequencing (thus altering level of story sense) has an impact on observer judgments of appellant veracity in the context of television appeals. It is conceivable that when unable to integrate or comprehend an appeal effortlessly, judges may turn to different implicit veracity judgments and potentially utilise mental short cuts. By controlling for the variable of story sense level and by making the availability of this cue higher in one condition than the other, it is expected that implicit and explicit judgments differ for both groups. It may also be likely that biases that form as a result of mental shortcuts be evident.

#### *4.8.4 Study 4: The perception of paired versus solo appellants*

Vrij et al. (2010) proposed that it is crucial to study liars and truth tellers in pairs as it reflects reality in the criminal world, where criminals usually work in groups or networks and not by themselves. As of yet, innocence-guilt judgments of pairs versus solo liars is still unknown. There are no studies to date which investigate whether two collaborating false television appellants are more successful at deception, as compared to when they appear solo even. Accuracy of veracity judgments has not been explored in these manipulated settings, and neither have implicit judgments been studied in this context. The main aim of Study 4 will be to discover how accurate judges are in their innocence-guilt judgments of appellants who are pleading for the return of their relatives, when these appellants are presented separately and as a pair; essentially whether paired liars are more successful in deception than when appearing solo.

## **PART 3: OBJECTIVE VERBAL CUE ANALYSIS**

### **4.9 Study 5: Objective Verbal Indicators of Truthful and Deceptive Television Appeals**

As expounded in Chapter 3, verbal content analysis shows much potential in discerning lies from truthful statements, whereas behavioural methods may offer less 'tell-tale' signs (Vrij et al., 2006). In the current context, truthful television appeals are made by individuals who are in distress and grieving either the momentary loss of a missing relative or the permanent passing of one murdered. This accordingly provides a framework for studying prospective indicators of a deceptive appeal. While of course grief is displayed through physiology and behavioural actions, verbal aspects were chosen to be focused on in the current thesis as speech content has been validated to indicate the foremost importance and mental states of the appealer; such as finding the victim urgently or fervently attempting to cover up a crime (Pennebaker, Mehl & Niederhoffer, 2003).

Grief produces psychological, emotional and social change within an individual (Vance, Najman, Thearle, Embelton, Foster & Boyle, 1994) which manifests in their language (Pennebaker et al., 2003). These changes include altered relationships with society and changes in roles within the family of the victim. Study 5 tests the hypothesis that verbal indicators differentiate high-stakes true from false statements. Verbal markers of emotion and grief will be drawn from the literature as the basis for a content analysis of the television statements, generating eight aspects of variables. In addition, this study employs a methodological approach that allows for the examination of verbal cues in relation with each other, so that even cues with very low frequency count included in analyses and assume a role in perspective with other cues that emerged.

### **4.10 Ethical approval**

All studies presented in the current thesis received approval from the University of Huddersfield Ethics Board Panel. All participants who were recruited to take part in the experimental studies were over 18 years of age and gave informed consent to take part. Participants in the experimental studies

viewed television appeal and interview videos showing individuals pleading for help in the case of a missing or murdered relative aired on television news programmes. These appeal videos were acquired and purchased from news and media sources via the internet. An information sheet was given to participants before they took part in the studies, which clarified the objectives of the experiment, the nature of videos participants will be shown, that they can withdraw at any time should they feel the need to, that their participation is entirely voluntary, that their data will be treated with confidentiality and anonymity, meaning they cannot be identified (see Appendix B). Once their participation in the experimental studies were completed, then were handed out a debrief (see Appendix C).

## CHAPTER 5

---

INTRODUCING THE STIMULUS  
MATERIAL:  
TELEVISED PRESS CONFERENCES

As presented in the opening chapters, several researchers have put forward that one of the concerns in lie detection research is the lack of a highly ecologically valid and naturally occurring data source (Vrij, 2004; Porter & ten Brinke, 2008; DePaulo & Morris, 2004). It has been alleged that basing research data with low ecological validity runs the risk of not being able to detect the more idiosyncratic deception cues. The issues that were believed to arise with using laboratory-based studies in lie detection research include generalisability and relevance of results from the demographic typically used in such studies (i.e. university college students) to a wider population. The view was that by using highly ecologically valid and high-stakes stimulus material, results presented will yield a higher practical application to real-world investigations. Because high stakes deception has been theorised to lead to better detection rates due to a higher level of cue saliency (Miller & Stiff, 1993), it is possible that important cues remain undetectable in low ecologically valid datasets. As such, the cues exhibited by deceptive individuals were expected to be more palpable in a high-stake situation than in a low-stake one. Accordingly, it was regarded that participants' accuracy rates may suffer when being exposed to laboratory-based stimulus as they would be unable to detect certain cues that are less salient. Conversely, Hartwig and Bond's (2014) meta-analyses would seem to challenge this argument (Cohen et al., 2010; Porter et al., 2008; Vrij, Akehurst, Soukara & Bull, 2004; Vrij, Edward, Roberts & Bull, 2000; Ekman, O'Sullivan & Frank, 1999; Miller & Stiff, 1993) that low detection rates are partially due to the usage of low-stakes scenarios lacking in ecological validity. Nevertheless, how people perform and how they make innocence-guilt judgments still warrant an examination in a high-stakes context. The present thesis uses real televised press conferences of relatives appealing for the return of a loved one who is either missing or murdered.

DePaulo et al. (2003) stated that certain cues may only be relevant, useful and/or occur in a specific context which may not be generalisable to other situations, which is an important consideration in the present thesis. Whelan et al. (2014) highlighted the context-specific disposition of certain deceptive behaviour, stressing the importance of analysing ecologically valid lies and truths in the particular framework in which they occur. The authors maintained that

seemingly inconsistent discoveries in high stakes lie detection literature may be partly due to the nonexistence of a context-based effort. Other researchers propagated that the saliency of a cue and the production of it depend on the factors that come with a given situation (Porter & ten Brinke, 2010), and that therefore cues that are salient in one situation may not be so in another. Not having a contextually-focused high stakes research scenario may also pose the risk of excluding certain cues from appearing. In some contexts, not only the presence but the direction of a certain cue will may also differ from another, thus it is important to take context into consideration.

### **5.1 Why are television appeals chosen over other stimulus material?**

One such example of high-stakes stimulus material are televised press conferences. Television appeals ‘naturally occur’, where people are under pressure to give an apparently honest account of the situation as they understand it. The essence of television appeals is that the person making the appeal has a close relative who has either disappeared or been found murdered. In a press conference the appellant asks for help in finding the missing person or that victim’s killer. Throughout the English-speaking world law enforcement agencies use these appeals to encourage members of the public to come forward with relevant information and in the hope that the perpetrator will give themselves up. In several cases, the individual appealing is later found to be the culprit. Recordings of these appeals therefore provide a unique possibility for a naturally occurring experiment in which genuine and false appeals are available in the same format. Televised appeals are considered high-stakes because deceptive appellants have a very strong motivation to successfully dupe viewers and investigators of the case seeing as the consequence of not being able to do so may result in long-term imprisonment or death. Having a strong motivation to succeed in lying is theorised to generate a higher salience in deception cues (DePaulo et al., 2003).

Televised appeals are chosen over other forms of high stakes data in the current thesis for several reasons. In the experiments that follow in the present thesis, lay participants (non-law enforcement officers) will be asked to assess these appellants after observing these televised clips, having had no interaction with the

appealer. This is similar to how judgments are made in the actual investigation of the missing or murdered relatives' cases, where both law enforcement authorities and television viewers will be establishing their explicit and implicit judgments while witnessing these appealers. Researchers Vrij and Mann (2001) used police participants in their study to observe these televised appeals, allowing the examination of explicit accuracy in law enforcement agents' detection of deception in a high stakes scenario. Instead of being constrained to focusing only on police performance, using these appeals will also permit expanding the investigation of how non-law enforcement agents perform in detecting deception in this context.

In addition, these videos are easier to obtain without the need for police clearance and obtaining approval for other forms of restrictions, in comparison to material in the forensic context such as, for example, police interviews and emergency calls (Harpster, Adams, & Jarvis, 2009; Mann & Vrij, 2006; Mann et al., 2004, 2006; Vrij & Mann, 2001a; Vrij et al., 2006). Due to the public nature of televised appeals, where the appeal is broadcast on national television to gain the attention of the local community for help with the case, these appeal videos are in the public record and freely obtainable. Ten Brinke and Porter (2012), Whelan et al. (2014) and ten Brinke, Porter and Baker (2012) have all utilised television appeals as their target material.

Furthermore, to date, there has been no formal recommendations as to how these appeals are conducted, how they should be televised, what questions should be asked, and how ought these veracity judgments to be formulated. Compared to police interviewing which receives considerable attention (i.e. in outlining interview style suggestions in maximising the prospect of detecting deception in that context) (Baldwin, 1993; Dando & Bull, 2011), television appeals have yet to receive a proposal in policy or script to guide police officers involved in these missing or murdered relatives' cases. The significance of having such a benchmark is reinforced by the fact that these appeals are conducted and broadcasted nationwide for public view. Even 911 calls and television shows such as COPS and Jeremy Kyle (Reynolds & Rendler, 2010), all of which were used in other high stakes lie detection research, follow a reasonably scripted guideline. If conducted right, the very way the appeals are managed can aid in the investigation of these

cases. Investigating this unique set of high stakes material is taking a step forward in gaining understanding in how these appeals transpire. Formal recommendations can subsequently be put forward.

Television appeals are a distinctive category of high stakes scenario, given that the individuals themselves typically contact the police to go forward with a public broadcast (Canter & Youngs, 2009). To reiterate, this lends itself a unique context and background to analyse. While there are cues which distinguish true from false accounts that appear more consistently across a high number of diverse deception situations and while it remains vital to determine these (Harpster et al., 2009; ten Brinke & Porter, 2012, Reynolds & Rendler-Short, 2010; Toma & Hancock, 2010; Van Swol & Braun, 2014; Pennebaker et al., 2003), a number of researchers stressed the importance of taking account of the situation in which the lies and truths are told in order to identify indicators of deception (Whelan et al., 2014; Sporer & Schwandt, 2006; Porter & ten Brinke, 2010; DePaulo & Morris, 2004). The significance of taking context into account is due to the anticipation that a variety of dynamics specific to this environment may affect what cues become prominent, what cues are generated, and in which direction they are expressed, which may be different from other high stakes contexts. Seemingly inconsistent discoveries in high stakes lie detection literature have also been attributed to the nonexistence of a context-based effort (Mann et al., 2002). Other researchers have also suggested that how salient a cue is, and the production of it, depends on the factors that come with a given situation (Porter & ten Brinke, 2010), therefore cues that are salient in one situation may not be so in another. Not having a contextually focused high stakes research scenario may also pose the risk of excluding certain cues from appearing.

Evidently, there may be cues that may appear in the context of this real-life situation, exclusive only to this genre. Although these cues may only appear in this one context, the value of detecting such cues are advantageous given the severity of the case where a missing or murdered relative is involved and the possibility that either an innocent appellant may face a wrongful conviction or a guilty appellant walks free. Particularly, the cue of emotions has the potential to be remarkably prominent in this genre in view of the situation at hand, where a

loved one may be kidnapped or possibly dead. As such, the present stimulus material offers an interesting context to study the perceptions of norm in grieving and the cognitive biases that may transpire, as well as the prospect of studying diagnostic affective cues that distinguish truthful from false appeals.

## **5.2 Existing research using television appeals**

Vrij and Mann (2001) were first to show their participants these television appeals for missing or murdered relatives. This exploratory research remains an invaluable step in uncovering observer judgments in high stakes conditions. These television appeals take place in a natural setting and are not performed in a mock setting, breaking certain limitations associated with lie detection accuracy and ability. Researchers who have followed suit in examining this particular real world data include ten Brinke and Porter (2012), Whelan et al. (2014) and ten Brinke et al. (2012) who all utilised television appeals as their analysis material. While there are palpable strengths in the Vrij and Mann (2001) study, there are limitations and critiques to be acknowledged in the authors' methodology. Firstly, the researchers' participants were police officers from Netherlands as well as from the United Kingdom. Participants were consigned to and tested in groups of up to five members at a time. Group dynamics and deliberation in the study generates concerns with extraneous variables which may have affected the study results as well as the group's accuracy rates (Patry, 2008; Zajonc, 1965), which the authors acknowledged as well. Secondly, a sample pool of only policemen and women in Vrij and Mann's (2001) study poses as a limit, making the study un-generalisable to the lay population. Television press conferences are revealed to the public sector, and in some cases even rely on the public's tips and aid. Thus, it is also imperative to begin grasping a layperson's judgments and their ability to make them in a high-stake lie detection context. Lastly, the appeals used in this study were all made in Britain, but the participants' first language was Dutch. An issue that may arise here is born from research on language, culture and their relation to emotion. It is possible that the participants in the study may not have fully comprehended the emotions of the appellant thus hindering their full ability in

accurate judgments, despite arguments that emotions are basic (Ekman, 1977; Mesquita and Walker, 2002; Sarter, 2012).

An extensive study published on television press conferences was conducted by Ten Brinke and Porter (2011). False appeals are public, often televised appeals by individuals posing as victims, pleading for the safe return of the real victim who may be missing or murdered, when they are in fact the offender (Canter & Youngs, 2009). Ten Brinke and Porter (2011) examined three modes of communication, namely speech, body language and emotional facial expressions or affect. Using 78 false appeal videos gathered from all over the world of individuals pleading to the public for the safe return of a family member, they found that liars used more tentative words than truth tellers, and also spoke less. The use of tentative words seem to be the acquiescence that their family member will not be found alive, circumventing commitment to the lie, and to decrease psychological discord they would suffer from the inconsistencies between their what they know and what they choose to display. The researchers suggested cognitive loading and psychological distancing to be the likely cause but this hypothesis was not supported. Approximately half of the pleaders in the false appeal videos were false appealers, the other half being truthful and not culpable of the any crime in relation to the disappearance or murder of their relative. They found no evidence for body language cues that is different for genuine and false pleaders, possibly due to successful camouflaging. Genuine pleaders are more hopeful and committed to the safe return of their family member in the words that they use.

Ten Brinke and Porter (2011) also found that false appealers tended to display feelings of disgust rather than sadness during their pleas. During the course of the plea, genuine appeals expressed sincere, full-face sadness and distress. The authors proposed that this reveals their candid emotions as well as to garner the empathy and the support required to lead their loved ones back home. In analysing the videos, the authors found that false pleaders were more likely to display a raised upper lip to express the emotion of disgust. It has been theorised that disgust is an innate, deep-seated reaction to the murderous act false pleaders have committed, even guilt and shame over their actions, or antipathy towards the victim (Chapman, Kim, Susskind & Anderson, 2009). Ten

Brinke and Porter (2011) also found that false pleaders blinked approximately twice as quickly more than genuine pleaders, which supports their previous findings in ten Brinke and Porter (2008) that arousal that stems from the finding that camouflaging emotions was correlated with higher blink rates. This challenged the more widely accepted idea that cognitive load is the only source of changes in blink rate during a lie (Leal & Vrij, 2010). Whelan et al. (2014) also explored television press conferences in their study and found behaviours that discriminated between truthful and false appeals. They found that false appealers used more equivocal and ambiguous language, tended to avert their gaze more, and engaged in higher incidences of head shaking as well as speech errors. They also identified truthful appealers made more emotion-related references, expressed more hope in finding the missing loved one alive, showed higher levels of positive emotions towards the victim and vocalised a higher amount of concern or pain. They also avoided brutal language in their appeals.

While undoubtedly invaluable, these research efforts had neither considered subjective perception of observers in depth viewing these clips while investigating possible implicit biases that may occur during judgment and decision making processes, nor had they considered a multivariate approach in analysing diagnostic verbal cues – two key aims of the present thesis expounded in Chapter 4.

### **5.3 Ground truth**

Rather than assuming ground truth, the present thesis applies a clear delineation of cases that qualify this definition to control for internal validity. Only cases substantiated with certainty beyond reasonable doubt will be used so as to also set a base line, thus subsequent results can be applied with a higher degree of confidence in cases where it is not then known in future. Appeal cases with sufficient ground truth has been verified with complete certainty corroborated by evidence beyond reasonable doubt to disconfirm the authenticity of their statement and to ascertain that the individual was implicated in the murder or disappearance of the victim. Usually in these cases a trial has been carried out and the verdict has also been assigned to the individual in question,

either whether they were not culpable for the death or disappearance of their relative or that they were in fact deceptive – unequivocal evidence both discredited their veracity of their appeal and ascertained their culpability for the murder or disappearance of the relative they were appealing for which subsequently led to their conviction and imprisonment. The initial criteria were set in Vrij and Mann's (2001) study, which was described in the section above, and were later modified by Porter and ten Brinke (2012). Ground truth was already established by ten Brinke and Porter (2011) as they utilised the video clips also used in the present thesis. An additional step in examining ground truth was also carried out in the present thesis for these clips via a search on the Internet from news sites and webpages pertaining to each missing or murdered relatives' cases.

The types of evidence include discovery of DNA such as body fluids, possession of weapon such as a gun or other items from the crime scene, CCTV/video/speed/security recording, un-recanted confessions, discovery and/or knowledge victim's body/body parts' location, search history on computer in regards to the location of crime or body, post mortem evidence, unsatisfactory or inconsistent alibis, public tips, phone records, forensic evidence such as soil, expert evidence in court, fibre, blood spatter pattern and arson tracks. For instance in the case of one appellant pleading for the safe return of his wife and daughter, evidence of forged emails and him purchasing tarpaulins and an electric chainsaw, including an un-recanted confession and the discovery of the victims' bodies instigated his arrest and conviction for both their murders. A more comprehensive table outlining evidence criteria met for all appeals can be found in Appendix H.

The present thesis will utilise video recordings of individuals appealing during television press conferences for the return of their relative, to search for what happened, and to find the killer or abductor of their relative. All clips will be obtained from a list of television appeals provided by ten Brinke (personal communication, 15th January 2014) to the current researcher who acquired permission to use some of these videos by independently purchasing these from news sites. The clips will consist of both female and male pleaders. They will be drawn from the United Kingdom, United States, Canada, New Zealand and

Australia. Appeals that do not contain adequate information, possesses questionable evidence that do not meet the criteria outlined in Appendix H or ones with no verdict yet from trial were not included in the sample.

#### **5.4 Length of stimulus material**

The length of video clips that are used in the present thesis is to be considered as well. Ambady and Weisbuch (2010) coined the term ‘thin-slice vision’, which proffers that people often only need not more than ten seconds of visual exposure to make non-random judgments about a person’s character, internal states, sexual orientation, popularity, vulnerability and other traits. This brief exposure and interaction is enough for a complete stranger to paint inferences from someone’s nonverbal behaviour, but these researchers suggest that it comes with its own limits to what one can infer from thin-slice visions, but in reality, individuals make split judgments. Slepian, Bogart and Ambady (2014) maintained that thin-slicing actually centres judges’ focus to a specific group of cues, namely non-verbal ones. This focus will either increase veracity decisions’ accuracy or decrease it, providing helpful non-verbal cues exists within the stimulus material. Otherwise, accuracy in veracity judgments is likely to decrease (Street & Masip, 2015). The authors further suggested that thin-slicing holds the possibility of aiding or deterring the detection of deception, not due to the fact that they tap into the automatic and intuitive cognitive processes instead of the systematic and deliberate one, but merely because they deflect focus from either diagnostic or invalid cues.

# CHAPTER 6

---

## CONSTRUCTION OF QUESTIONNAIRE ITEMS

To reiterate, not only is the present thesis interested in measuring explicit veracity assessments of ‘is the appellant guilty or innocent’, it is also interested in measuring implicit judgments in effort to gather more information about what cues they may be using in making their veracity decisions. The questionnaire aims to incorporate items that capture what valid and invalid responses judges may utilise. In the current chapter, implicit judgments that guide the derivation of the questionnaire used in the four experimental studies will be discussed. These range from implicit judgments of perceived emotions and meta-emotions, verbal elements i.e. ‘liars’ stories do not make much sense’, and judgments of the sender’s appearance.

The aim is to study which questions contribute to judges’ explicit judgments as well as their accurate explicit innocence-guilt judgments. Guided by literature reviewed in the opening chapters, the questionnaire ultimately endeavours to uncover cognitive aspects of how judges may come to perceive both the appeal and the appellant.

### **6.1 Defining the ‘implicit’ nature of question items**

To avoid any confusion of the term ‘implicit’ used in the current questionnaire and throughout the present thesis, a definition of what this term means is specified here. First, a definition of intuition is discussed. While intuitive judgments are usually defined as instantaneous and requiring minimal mental effort (Hogarth, 2001; Myers, 2002), implicit judgments in the current questionnaire may involve either automatic processing or systematic ones that are more cognitively deliberate.

Granhag (2006) further theorised that measures in assessing whether a sender appears sympathetic involves more ‘intuition’, for example. Intuition may certainly contribute in the process of making implicit judgments for some question items more than others; however, Granhag (2006) proposed that there may not be a strong relationship between the two concepts. Furthermore, the available empirical evidence for unconscious lie detection is lacking thus far. Street and Richardson (2015) maintained that where an advantage can be dependably observed in terms of lie detection accuracy, this outcome can be inferred to the

conscious and deliberate part (in the absence of an explanation pertaining to the unconscious one). Besides, by asking judges a response to a question that may have initially been embedded in their ‘unconscious’, this may well surface it to the conscious level in any case. The present thesis adopts the definition of implicit cues of being responses by judges to “every other question but the statement’s ground truth” (Granhag, 2006, p. 188).

## **6.2 Questionnaire measures and items**

The questionnaire will comprise of 16 complementary questions divided into different components: meta-emotion, appearance, emotion and cognition of source content. The questions are all rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Bearing in mind that regression analyses will be conducted for implicit judgments, the number of questions included in the questionnaire had to be concise while still retaining the foremost implicit cues of importance. Participants will be given a questionnaire pack, which contains a briefing, consent form, the main questionnaire, a demographic questionnaire, answer sheet and debrief. These can be found in Appendix B to G accordingly. For the Audio-only condition, a specific question sheet can be found in Appendix E. To examine guilt and innocence biases, an additional dichotomous ‘Do you think the appellant is guilty or innocent (of being culpable in the murder of the victim)?’ question choice is included to assess binary veracity decisions.

These cues are chosen because literature has depicted that they have been used by participants in deciding a speaker’s veracity, however these questions are not overtly asking participants whether they think the appellant is guilty or not except perhaps for the item pertaining to appellant credibility (which is closer to an explicit question than an implicit one). The concepts of cognition of source content, perceived appearance of source, emotional perception of source and meta-emotion for and/or with source will be operationalised and be studied in Studies 1 to 4 (Chapters 7 to 10). The following paragraphs discuss the literature pertaining to each question item that will be included.

### *6.2.1 Credibility*

Yuille (1989) wrote that credibility assessment refers to any attempt to ascertain truthfulness. When one makes a credibility judgment, one aims to determine the truth of a statement or story. In the present case credibility judgments can also be made in a police investigation, for example, of a television appealer pleading for the return of his missing wife. The judgment would consist of a judgment of whether he or she is credible in general, of not being culpable of the death or disappearance of his wife, and that he or she is not held responsible of the crime. While the story an appealer gives may make sense and may be plausible, they may still be judged as lacking credibility. Schwarz (2015) stated that in judging whether an account is truthful or otherwise, people typically pay attention to certain decisive factors. One of these factors include how credible the sender or source is. Other variables in the questionnaire are further segmented and expounded below in their appropriate sections.

## **6.2.2 Cognition of Source content**

### *6.2.2.1 Story sense*

Granhag (2006) proposed that certain measures of implicit veracity assessment pertain to the sender's verbal behaviour. An example of a comparatively 'verbal' implicit assessment is that the account of truth-tellers will tend to make more sense. To measure the cognition of Source content, a series of questions were designed. Validation for including this component comes from epistemological research suggesting three vital clauses of truth: consensus or consistency, correspondence and coherence (Baudet, Jhean-Larose & Legros, 1994). For speech to be deemed truthful, according to these researchers there should be no contradiction within the story content. The subject matter within this content needs to also be clearly interconnected. McAdams (2006) wrote that stories that illustrate characters without a clear motive, vague plot, no clear cause or consequence, or closure can be judged as incoherent as well. Not only does coherence pertain to the structure and organisation of a story, it also refers to the content of the story. Furthermore, the author proposed that an account that opposes or challenges an observer's expectations and experiences can appear incoherent to them as much as an account that does not follow 'internal' structural

norms. Lam (2001) described that in an attempt for someone to assign a probability of guilt or innocence towards another, people first try to rule out stories or events that are incoherent and do not make sense. Hence, the question item ‘Does the appellant’s story make sense?’ is included in the current questionnaire.

#### *6.2.2.2 Story plausibility*

Lam (2001) wrote that post establishing the stories that make sense or are coherent, jurors then would assign a degree of plausibility to the stories. This very much follows a Cartesian argument (Street & Kingstone, 2016) where understanding comes prior to believing or disbelieving, whereas the Spinozan rationale (Gilbert, Krull & Malone, 1990) would argue that individuals initially assume truthfulness and only later revise it necessary. In relating the variable of ‘story plausibility’ with ‘story sense’, Schwarz (2015) posited that one of the factors that leads people tend believe a story more is when it is internally organised and lacking in inconsistencies.

Statement Validity Analysis postulates that an assessment of the validity of a statement is not an assessment of the general credibility of a speaker (Vrij & Mann, 2004). Credibility of the speaker and validity of the statement are therefore two distinct criteria of truthfulness. While an observer can perceive that the statement is valid, he or she may still doubt how truthful the speaker is. Nahari et al. (2010) defined plausibility as sufficient but not necessary for credibility judgments, thereby firmly extricating both terms and designating them to be regarded separately. Something may be evaluated to be false even though it is plausible in theory. Chung (2010) described plausibility as the likelihood that the events could have transpired in the way in which the appellant presents it. Connell and Keane (2006) defined a scenario that is highly plausible as one that matches existing understanding well, and has diverse resources to corroborate.

In any case, plausibility remains to be a good linguistic cue for veracity and considerably affects credibility judgments. Thus, the question ‘Is the appellant’s story plausible?’ will be included in the current questionnaire.

#### *6.2.2.3 Well-thought-out*

The Oxford Dictionary defines ‘well-thought-out’ as ‘carefully considered and planned’ (Well thought out, 2017). Vrij et al. (2010) proposed that liars tend to prepare a story when they can, if they are expecting to be interviewed. Well-prepared liars are harder to detect than those who are not (Bond & DePaulo, 2008). This variable expects to measure how carefully people think the appellant has considered and planned their speech before giving the appeal on television with the question item ‘The appellant appears to have thought out well his or her appeal before appearing on television’.

#### *6.2.2.4 Organised*

The item ‘The appellant gives an organised speech’ intends to measure whether judges think the internal order of the appeals are arranged in a systematic manner. While this variable is closely related to the variables of appeal sense and planning, it pertains more specifically to the structure and order of the actual appeal itself. According to the Oxford dictionary, ‘organised’ is defined as ‘arranged or structured in a systematic way’ (Organised, 2017). Structure of stories, which is comprised of episodes, affects a jury member’s understanding and how they make innocence-guilt decisions. The structure of these episodes corresponds to the structure of people’s existing knowledge about how the structure should be and of similar events related to the story (Pennington & Hastie, 1992). The temporal and causal order of a story can affect a judge’s perception of the story (Lam, 2001). Pennington and Hastie (1986, 1988) found in one of their experiments that the ease at which stories could be constructed in terms of its order influenced a juror’s decision-making and what verdict outcome they make.

#### *6.2.2.5 Wording*

In relating ‘story organisation’ and ‘story sense’ with the variable ‘wording’, a story that is internally consistent will be processed with much more ease (Johnson-Laird, 2012). The variable ‘The appellant words his or her sentences in a manner that is pleasant to hear’ will also be included in the questionnaire to measure if judges find the appeal’s sentence structure pleasing

to listen to, and if perhaps this could contribute to the rating of appellant's innocence. To also note, Fernald (1993) found that infants as young as five-months old were able to show negative affect to a language they were less familiar with compared to one language vocalised in English, to which they showed higher positive affect. While of course the purpose of this variable to be included is not to measure languages but rather in wording agreeableness, the same principle of 'favouring the familiar' applies here. The hypothesis, similar to that for 'story sense' and 'coherence', is that the more pleasant-sounding a sentence is, the higher the chance of a judge appointing innocence towards a speaker.

#### *6.2.2.6 Practice*

While the content of an appeal could make sense, and its structure be organised and carefully planned, the perceived aptitude or proficiency in one giving the actual appeal is a related variable to consider. The Oxford Dictionary defines 'practice' as the 'repeated exercise in or performance of an activity or skill so as to acquire or maintain proficiency in it' (Practice, 2017). In terms of how well individuals judge liars and truth-tellers, Bond and DePaulo (2008) stated that practised liars are harder to detect than those who have not had much practice. The item 'The appellant seems to have had practice in making television press conferences before' intends to measure how much time or effort judges think appellants spent working on mastering and refining their appeal or appeal-presenting skills before making appearing on television. Inclusively, all five variables expect to evaluate implicit components of how judges cognitively process and view the story content.

#### **6.2.3 Emotional perception of Source**

As extensively reviewed in Chapter 2, a number of research from multiple areas have shown that not only perception of emotional expression is associated with decision making processes, emotional content can bias these processes (Forgas, 1995; Doss, 2002; Slovic et al., 2007; Shafir et al., 1993). Schwarz (2015) proposed that besides source credibility (as introduced in Section 6.2.1), another

decisive factor in assessing the truthfulness of an account is the congruence of the story with the judge's pre-existing beliefs and knowledge. The two variables that will be included in this questionnaire for this component is expected to reveal judges' beliefs about Sources, balanced with the believability of speech while corresponding this with the expressivity of emotions ('The appellant's emotions do not seem to match his or her story' and 'The appellant displays less emotions than seems appropriate for their situation').

#### **6.2.4 Attractiveness/appearance perception of Source**

The interrelation between appearance and other factors, concerning how we think and process information about a person and how these may interplay with feelings and thoughts are discussed at length in Section 2.3.2. Two questions pertaining to attractiveness and appearance will be included in the questionnaire. They are 'The appellant has a pleasant voice' and 'The appellant has an attractive face'. Literature has shown that a judge is instinctively capable (Fink & Penton-Voak, 2002) of using various aspects of a Source's appearance, including their attractiveness and the sound of their voice (Baron et al., 2008; Paunonen, 2006) to make veracity assessments of that person.

#### **6.2.5 Meta-emotion**

As introduced in Section 2.3, the present thesis draws attention to considering the context in making veracity judgments. Researchers must first consider this, as well as the social and cultural norms of the situation to assess why certain behaviours may be judged less favourably. This highlights the importance of how well an incoming stimulus integrates with pre-established norms and beliefs (Schwarz, 2015). As put forward in Section 2.3, for instance, the influence of emotional display can be explained by the expectancy violation theory (EVT), which suggests that people hold expectations regarding which behaviours are considered normal in a given social scenario based on personal experiences, cultural and societal norms (EVT; Dahl et al., 2007; Olsen et al., 1996).

The more sympathetic personal reactions of the observers to the honest appellants (such as, 'feel the appellant's pain' and 'feel sorry for the appellant'), and

their corresponding lack of sympathy for the deceptive appealers, are also of particular interest in relation to recent developments in neuroscience. Studies on mirror neural mechanisms suggests that having this facilitates an experiential system that helps one comprehend what is observed in others, and this allows one to grasp someone else's emotions (Rizzolatti, Fogassi & Gallese, 2006). Therefore, it is arguable that this kind of emotional synchronisation is more likely to occur with genuine emotions than faked emotions; hence one's own emotional reactions to people can be used as 'relevant data' when making credibility judgements about them.

A scale put forward by Kujipers, Hakemulder and Doicaru (2004) is a usable measure of this meta-emotion concept. Kujipers et al.'s (2004) scale focused on how absorbed and emotionally engaged one is into stories and films. Their scale includes all previously explored absorption-like concepts including previously investigated flow, immersion and narrative engagement were classified and summarised in their absorption questionnaire to form the absorption experience. The meta-emotion scale that will be used in the present study is adapted from Kujipers' (2014) original questionnaire to contain five questions, relating to the level of emotional identification participants experience when watching each appeal video. The author tested their own model using Confirmatory Factor Analysis (CFA) and reported the overall reliability of the scale to be  $\alpha = .954$ , which is highly satisfactory.

Adaptation of Kujipers' (2014) questionnaire is necessary for several reasons. Firstly, to include the full version of this questionnaire runs the risk of incurring maturation and practice effects, keeping in mind that judges would need to answer the same question items for each appeal video watched. Subsequently, maturation and practice effects are likely to threaten the internal validity of the current questionnaire. Secondly, both instruments are not appropriately worded for the unique scenario in the current thesis. Kujipers' (2014) instrument was developed for text-based stimulus material (a short story) and not video or film. Thus, some words in the question items has to be changed. For example, the item 'I felt sympathy for the main character' will be modified to 'I felt sympathy for the

appealer' to reflect the specific situation that the current thesis is interested in. The substance of the item will not be altered, however.

Kujipers (2014) posited that emotional engagement is an important aspect of this scale that pertains to feelings receivers have for and/or with characters (in the current case, for and/or with appealers). These feelings include empathy, sympathy and character identification. One of the objectives of a television appeal is to gain cooperation from the public in solving a missing or murdered relatives' case. As aforementioned, television appeals are typically emotional broadcasts that potentially instigate sympathy and worry in viewers. The hypothesis is that the higher this emotional identification and level of meta-emotion felt for the appealer, the more likely a viewer will judge them to be innocent. To measure emotional engagement with the Source (appealer), three items are selected: "I felt sympathy for the appealer"; "I was worried about what was going to happen to the appealer and the missing person"; and "The appealer makes an appeal that affected me emotionally". These specific items are selected to reflect the specific nature of television press conferences.

Attention is another aspect in their scale that is defined not as a character trait but as an effortless deep concentration of receivers feel while engaging in a sender (Kujipers, 2014). As a measure of attention, the items "I felt absorbed into the story the appealer was telling" and "When I was listening to the appeal I was focused on what happened during the appeal" are selected from Kujipers' (2004) questionnaire as well. In Kujipers' (2014) original scale, the Cronbach's alpha for Emotional Engagement was .914 and the Cronbach's alpha for Attention was .905.

### **6.2.6 Supplementary questions**

Other questions will also be included in the current questionnaire to ensure participants are not familiar with the outcome of the appeals or with the appealers themselves so as to eliminate the possibility of biased responses. These questions are 'Do you personally know anyone who featured in the appeal you've just watched?' (Question 18), 'Do you know of the appeal you have just watched i.e. its outcome; the case?' (Question 19), 'If you ticked yes to knowing the appeal, what did you think the outcome was?' (Question 20) and 'If you have any other

comments or notes for this particular appeal, state below' (Question 21). The advantages of using the present questionnaire are discussed further in Chapter 11 in Section 11.7.4.

### **6.3 General methodology across Studies 1 to 4**

A total of 29 MSc students in Investigative Psychology from the University of Huddersfield were involved in the data collection process, as part of a module they completed for the Masters course entitled 'Empirical Research and Research Writing'. They were assigned the task to acquire a minimum of ten participants and up to as many as they wanted to. All MSc students involved in the data collection were given a questionnaire pack prior to collecting data. This pack contained a briefing, consent form and the main questionnaire. These can be found in Appendix B to G and Appendix E. They were given the appeal videos pertaining to the experimental group they were randomly assigned to, to be shown to participants. These students were also given a standardised SPSS file to enter the raw data they collected, and this marks the extent of their involvement in the present thesis. These individual SPSS files were then transferred to the current researcher. Data analysis was conducted independent of the MSc students' involvement.

A breakdown of the number of students who collected data for each condition is reported here. Four students collected data for the Audio-only condition in Study 2, and another four collected data for the Audio-visual condition in the same study. Three students collected data for Study 1 (which was a within-subject design). Six students collected data for the High Sense condition in Study 3, and five students collected data for the Low Sense condition in the same study. Three students collected data for Paired condition in Study 4, and four students collected data for the Solo condition in the same study. All data across all four studies were collected during the same period of time, from September 2014 until January 2015. The total numbers of participants for each condition in each study are individually reported and can be found in the Methods section in their respective study. Descriptive information of participants is also reported in each individual study in the same section.

The present methodological design employed a double-blind technique. Neither the Masters students running the experiments nor the participants in the present thesis knew what any of the conditions represented. Participants were also randomly allocated a condition, removing order effects. This decreased any threats to demand characteristics, which refers to participants attempting to behave in a certain way because they believe that this is expected of them (Orne, 1962). This technique also lowered the chances of any experimenter effects from occurring as it was not possible for the Masters students to bias participants in any conditions. Furthermore, the employment of various experimenters in and of itself avoids experimenter bias (Orne, 1962). All conditions in all studies consisted of four video clips of individuals appealing for a missing or murdered relatives' case (except for Study 1 which employed a within-subjects design; there were only two video clips for the condition of High Emotionality and a further two for Low Emotionality). The questionnaire given to participants to complete was exactly the same for each video clip.

The same instructions as to how to conduct the experiment were given to all MSc students and were as follows: Before beginning to watch the video clips, participants were given definitions of 'story plausibility' and 'credibility'. 'Story plausibility' here was defined as "the possibility that the appellant's story could be true or could have taken place" and 'credibility' was defined as "your perception of how truthful the appellant is regardless of the story plausibility or even after it is taken into account". All participants were asked to first watch or listen to a clip, and afterwards fill in the matching questionnaire before moving on to the next clip, and so on. This was carried out sequentially to not encourage simultaneous comparison of the appeals. Participants could watch the videos as many times as required and to change their decisions on the answer sheet at any point during the process. They were instructed beforehand that some of these clips would be genuine, while other appellants would be responsible for the death or disappearance of their relative. Participants received no training in lie detection. The experiment will be carried out on individual participants in a quiet room, with a laptop and headphones from which the clips could be played

via a PowerPoint presentation. This presentation will be carried out in a quiet room, for a clear audio experience.

A concern to be acknowledged is the small number of trials carried out in all experimental studies in the present thesis. Unquestionably, a higher number of trials tends to lead to fewer unexplained variations in the sample and averages out some (or lowers) random error (Winokur, 2005). This ensures the best out of 'signal-to-noise' ratio which will enable interpreting outcomes with higher accuracy. Here, a signal depicts that a variable really influences the dependent variables, and noise depicts any random errors that may surface (Slutz & Hess, 2017). In other words, it is difficult to draw accurate conclusions from a small number of trials. While a higher number of trials unable to be conducted in the present thesis due to time constraints, this could be improved in future studies. However, the following studies presented in Chapters 7 to 10 remain important and valid as a first step in uncovering the way people think when making guilt-innocence judgments using television appeals as stimulus materials.

## PART 2

---

### EXPERIMENTAL STUDIES OF SUBJECTIVE CUE PERCEPTIONS

# CHAPTER 7: STUDY 1

---

## PERCEPTION OF EMOTIONS IN TELEVISION APPEALS

Figure 1 in Chapter 1 depicts how a Source is presented and what it constitutes can affect outcomes in veracity judgments. A lateral point to be made is that observers may base their judgments on their normative expectancies of a stimulus based on their culture and previous experiences (Bond et al., 1992).

While emotional display can affect veracity judgments, so can sympathetic emotional responses *towards* this display. According to Vrij (2008), there are several factors that can influence cues to deception, one of them being emotional reactions. Observers may then use their own emotional responses as an effortless but potentially faulty indicator in regulating their judgments in favour or to disapprove of the appellant in question in a difficult scenario where the outcome is ambiguous (Epstein, 1994). Also in Figure 1, this reliance on meta-emotions, or felt emotions, can then ultimately contribute to veracity judgements. Slovic et al. (2007) describes this reliance on such feelings and consigning a great significance of it in consequential decision-making the 'affect heuristic', which is synonymous with the concept of meta-emotion.

In the present study, since television appeals are typically emotional, one of the most salient and available cues available for observers to pick up from a television appellant are the emotions displayed. When watching something as emotionally salient as a television appeal observers may, as one might expect, develop a sympathetic emotional response towards the appellant. Resulting from this display of emotions (or the lack of) observers may, people rapidly form emotional imprints from their experiential thinking mode rather than a comprehensive cognitive-based assessment (Slovic et al., 2007) and evaluate their impression of the appellant – either favouring or disfavouring them (Shafir et al., 1993). In the current thesis, expectancies of a grieving appellant are considered together with its relationship to level of emotional displays and veracity judgments. To understand and distinguish the emotions a true appellant shows and the ones a false appellant shows, the literature on what emotions a truly grieving person goes through is considered.

## 7.1 Perception of emotion in grieving individuals

If observers may form impressions and judgments of a television appeal based on their normative expectancies of a stimulus based on their previous experiences, consideration first needs to be given to what these normative expectancies might be, considering the specific context in the current scenario. In the present case, the normative expectancies cover the grounds of a grieving individual who has just temporarily or indefinitely lost a close family member. This provides a theoretical basis from a grief perspective in understanding judgment and decision-making processes of an observer in a high stakes deception detection context. To date, grief theories have not been applicably explored in this area of research. As of yet, the role of grief perception in television appeals, along with public expectations of how an appeal is 'supposed' to appear is currently unknown. In this section, consideration is given to theories and public expectations from available literature.

Early theories of grief include Freud's (1917) and Lindemann's (1944) psychoanalytic model, and Bowlby's (1961) attachment model where conditions surrounding the death of a loved one affected the features, severity and period of the bereavement course. The second generation of grief theories include psychosocial transitions by Parkes (1972), Kubler-Ross's (2005) five stage model, and Worden's (1983) tasks of mourning where the tasks were the work that the mourner must do in order to advance through the course of grief. The major issue with these models was that they insufficiently explained variations of grievers who experienced similar encounters of bereavement, appealing a new zeitgeist of theories.

In later models of grief, individual differences in grieving were recognised. Penman, Breen, Hewitt and Prigerson (2013) suggested that grievers may experience variations in emotional expressions such as sorrow varied with anger, behavioural changes and in thinking, sometimes experiencing disbelief of their situation or perplexity. Relating the principles of these theories to public expectations, social psychology research on grief has shown that participants assess emotional expression of grief to be significant (Costa & Stewart, 2007). Participants were able to identify diversity in grief responses, and considered

avoidance behaviours to be problematical. Boelen, van den Bout and van den Hout's (2003) findings also concurred in that their participants found mourners who engaged in avoidant coping styles formed negative impressions and interpretations of the griever.

Theories aside, the medical terminology of the grief experience addresses grief in the Diagnostic Statistical Manual-V (American Psychiatric Association, 2013). The recent changes of bereavement diagnosis outlines, with the bereavement exclusion removed, that in grief painful emotions are experienced in waves, are usually amalgamated with positive reminiscences of the departed.

Kubitz, Thornton and Robertson (1989) found that the type of death (in their case, sudden or anticipated) experienced by griever can impact participants' expectations and evaluations of grief. In television appeals, most cases are sudden rather than anticipatory, and even in deceptive appeals the appellants who have had culpability in the murder of their relative try to portray the type of death as sudden. Kubitz et al. (1989) also studied interpersonal attractiveness of the griever as a moderating effect, factorising gender of the griever, sudden or anticipatory type of death, and the intensity of the symptoms as exhibited by emotional expression. In their study, participants read scripts describing an individual grieving over the death of a friend and then rated the bereaving individual on a supplied questionnaire. Participants perceived males who exhibited less intense emotions for a sudden death as functioning better. They perceived females who exhibited less intense emotions for an anticipated death as more attractive, and lastly perceived an exhibition of higher emotional expression as appropriate in a sudden death vignette. In contrast, Penman et al. (2013) investigated public expectations of grief following bereavement after DSM-V changes and found no evidence that type of death influenced participants' expectations of grief.

Studies have also suggested that participants assume an unfounded traditional stage model for the grief process (Breen & O'Connor, 2007; Costa & Stewart, 2007). Public expectations of grief also follow an unsupported time period, where grief is expected to lessen over time (Lensing, 2001; Costa, Hall, & Stewart, 2007). The general public have clear expectations concerning the

appropriate duration of grief, and in Costa et al. 's (2007) findings provided evidence that participants viewed time as an important factor in their judgments of an 'appropriate' grieving model.

In the context of television appeals, while no research has explicitly been done on the perception of emotional display, it is known that Susan Smith murdered her two young sons and proceeded to falsely appeal on television for a period of nine days for their return. She appeared to be earnestly crying in most of the appeal videos whilst blaming an African American man for the crime (Doss, 2002). Here, Doss (2002) claimed viewers of the appeal appeared to believe her to be credible. This was ostensibly due to her believable level of emotional display, which was seemingly full of despair. Similarly, Anderson et al. 's (1999) study showed that participants reported emotional involvement while telling the story (i.e. "...he looked really sad when telling the story..."; "She did not look genuinely happy...") was used as a cue to determine the veracity of speakers.

In consideration of the literature on grief theories and public expectation of grievers, it is posited in the current study that television appealers who are visibly distressed and expressed high levels of emotion will more likely be thought to be more credible. Judges may assess the probability of whether a sender is guilty or innocent based on how much they believe he or she represents someone who is grieving, and this is subsequently based on stereotypes of how a griever should behave. Because a high level of emotion is a particularly salient cue and readily picked up by observers, they may use this specific cue in regulating their veracity judgments. This may especially be the case when the outcome of a situation is ambiguous and little information is given about the appealers to judges.

## **7.2 Meta-emotion bias**

The meta-emotion bias is not a new discovery, with previous research efforts reporting that emotional presentations can influence perceived veracity and character believability. This bias is where people try to use emotions as a cue in making veracity judgments (Peace & Sinclair, 2012). Baldry, Winkel and Enthoven (1997) found that rape victims who exhibited higher levels of emotion such as being in hysterics, crying, shaking and exhibiting clear despair were

seen as more truthful and were believed more than those who were neutral during their statements, an unfounded bias. These results were replicated by Bollingmo, Wessel, Eilertsen and Magnussen (2008). Wessel, Magnussen and Melinder (2013) again found similar results with those working in the child protective service area. In this study, participants rated eight mock videos of children acting and reading out accounts of being abused with differing levels of emotional expression for each account. Participants were strongly swayed by the level of emotional display and this consequently affected their credibility judgments of the child. In a study by Bollingmo, Wessel, Sandvold, Eilertsen and Magnussen (2009) which also used a mock scenario, the authors found that giving participants the added instruction of not using displayed emotion as an indicator of truthfulness causes participants to adapt their decision-making and reduce the effect of this bias on subsequent credibility judgments they make. In Wessel, Bollingmo, Sønsteby, Nielsen, Eilertsen and Magnussen's (2012) study, the actor was a young male accused of rape in a mock trial scenario. Mock juror participants judged the actor's accounts. Findings from this study were that credibility judgments were highly affected by emotional display instead of story content.

In light of these studies, this chapter aims to assess the possibility of a meta-emotion bias through the use of implicit assessments, to examine whether this bias occurs in the context of television appeals.

### **7.3 Attractiveness/appearance bias**

As reviewed in Chapter 2, the attractiveness and appearance of a sender can be used as a mental shortcut when having to make decision tasks about the sender. While it is extensively known that attractiveness can influence judgments, this has yet to be explored in the current context.

### **7.4 Innocence bias**

As expounded in Chapter 6, explicit ratings will be measured as well as the more 'implicit' version of credibility.

## **7.5 Cognition of Source bias**

As introduced in the opening chapters, perceived level of a sender's preparation and rehearsal for a story may bias observers' veracity judgments (Akehurst et al., 1996; Strömwall & Granhag, 2003). Vrij et al. (2010) proposed that liars tend to pre-formulate a story if an interview is anticipated since they must practice delivering the lie consistently. Deceitful individuals tend to rehearse stories so that it follows a logical consistency in hopes of being believed. In the present study, perceived preparation is measured implicitly via several questions pertaining to perceived practice, how well thought out the appeal seems to be, how organised it is and whether the wording of sentences is pleasing to hear – all outlined in Chapter 6. Whether the preparation bias occurs when level of emotional display is manipulated has never been investigated in the current context.

## **7.6 Key issues to explore**

The key aim of this study is to uncover how both emotionally charged and emotionally deprived television appeals of missing or murdered relatives' cases impact upon judgments of appellant veracity, and whether differences in underlying cognitive processes exist between conditions. In consideration of all factors mentioned above, an observer watching television appeals may anticipate and form expectations of certain emotions or states of mind from someone who has just had their relative kidnapped or murdered, and potentially be absorbed into and identify with the appellant, and these expectations and identification in turn may affect their judgments of how honest or deceptive the appellants are. This expectation may regulate the implicit judgments of perceived emotional display and meta-emotion.

As abovementioned, social psychology research on grief has shown that judges assess emotional expression of grief to be significant (Costa & Stewart, 2007). This implies that television appellants who are visibly distressed and expressing prominent level of emotion will more likely be thought to be more credible. Because a prominent level of emotion is particularly salient and available to observers to pick up on, especially when information on the appellant is scarce,

the hypothesis is that they will judge appeal videos showing high emotional display to be more innocent. Further, when the highly emotional appeals contain verbalised emotions such as 'love' and 'miss' towards the victim, an additional verbal cue is present for observers to pick up on. A secondary objective here is to explore whether other known biases, such as the attractiveness and appearance bias, preparation bias and innocence bias, will be significantly different between both conditions.

## **7.7 Methods**

### *7.7.1 Participants*

This sample consisted of 51 individuals. All participants answered questionnaires pertaining to four appeal cases each, leading to a sum of 204 cases in total. None of the participants had prior knowledge of the appeal or appellant. Participants were recruited by snowball sampling. They were from varying occupations ranging from dentist, prison officer, company director to farmer, psychologist, radio DJ and more; as well as nationalities/ethnicities (British, Asian, African, Middle Eastern, European and American). Participants who answered 'not sure' to other implicit variables were taken note of and results are presented in Appendix I.

### *7.7.2 Material and Procedure*

As briefly introduced in Chapter 6, Bond and DePaulo (2008) highlighted that detectability and credibility are sometimes mutually exclusive, in that certain individuals' veracity is obvious and some others indecipherable. Certain individuals just appear sincerer than others, and some others just appear more 'guilty' regardless of their credibility. Even honest grievors can be mistakenly judged to be malingering. Kassin (2012) reported the case of Amanda Knox (now exonerated) where she was mistakenly judged by police and the court to be guilty of the death of her roommate because she did not exhibit any emotions as a reaction to the death. The issue of culture ties in with sender detectability as well, where Vrij and Fischer (1995) argued in their study of simulated rape cases that

not every individual or culture exhibit clear symptoms of distress when retelling an upsetting event they have experienced.

To curtail this issue of sender detectability, throughout Part 2 in the present thesis, where appropriate all judges in different conditions were shown video clips of the same appealers. In other words, while the appeal content differed in some way or form for judges in different conditions through experimental manipulation, the speaker in both conditions remained the same where possible. The only study where this was not feasible was in the current study, as it was not viable to use the same appealers for the conditions of High and Low Emotionality (i.e. if an appeal is highly emotional in an appeal video, it was impossible to manipulate the video-clip of the same appeal to be less emotional).

Hendriks and Vingerhoets (2006) found that observers in their study felt more sadness and emotional support as a response to being shown faces that were tearful compared to faces that were neutral. In the current study, the facial expression of crying was chosen to denote a 'high emotionality' and a neutral facial expression was chosen to denote a 'low emotionality'. Highly Emotional appeals were selected based on the presence of crying (i.e., one appeal used tissues to wipe her tears, blowing persistent watery mucus discharge from the nose) and the presence of a cracking voice due to distress while appealing. In contrast, Low Emotional appeals were selected based on neutral expressions and an absence of visible distress in facial expressions and no crying or blowing/wiping water mucus discharge present or cracking voice. While noting that in some cases people feel sad and yet do not express this (Vrij & Fisher, 1995), these facial expressions were chosen not only because they are qualitatively different expressions, but they have been shown to evoke different feelings and thoughts in people who view them (Hendriks & Vingerhoets, 2006). Two highly emotional appeals were selected and two appeals with low emotional qualities were selected. In addition, the two highly emotional appeals contained spoken emotions such as 'miss' and 'love' whereas the low emotional appeals did not. All judges experienced the same four videos in a non-randomised order.

A difficulty in choosing highly ecologically valid data is that emotional display cannot be manipulated. In other words, the same appeal cannot be

manipulated to show high levels of emotional display in one condition and in another condition showing less emotion while giving the same appeal. This means that the current study cannot control for appeal content, which will be different for all appeals chosen. To counterbalance this limitation, the study design is maintained as within-group. Nevertheless, it cannot be firmly concluded because appeal content was not completely controlled for. Additionally, it was problematic to find guilty appellers displaying the emotional levels that corresponded with the selected definitions of what constitutes a 'Highly Emotional' and 'Low Emotional' appeal for this particular study. To recapitulate, Highly Emotional appeals were selected based on the presence of crying (one appellant used tissues to wipe her tears, blowing persistent watery mucus discharge from the nose) and the presence of a cracking voice due to distress while appealing. In contrast, Low Emotional appeals were selected based on neutral expressions and an absence of visible distress in facial expressions and no crying or blowing/wiping water mucus discharge present or cracking voice. Thus, all appeals selected in the present study were innocent.

One male and one female were chosen for each level of emotionality so as to balance for any gender effects. Appeal video 1 was 37.46 seconds in length, Appeal video 2 was 25.00 seconds in length, Appeal Video 3 was 26.20 seconds in length and Appeal video 4 was 44.97 seconds in length. Care was taken to select appeal with lengths as similar as possible, with an average of 33.41 seconds ( $SD = 9.53$ ). The general methodology of how the experiment was conducted is outlined under Section 6.3 in Chapter 6.

Because individual differences were anticipated in appeal emotional level appraisal, only judges who agreed with the author's categorisation of the emotional level were retained for further analyses. This step was taken so that we can be confident that judges viewing the videos assigned as 'highly emotional' videos also found them to be highly emotional, and the videos assigned as low in emotional content were also perceived by judges to be not very emotive in nature. This is to ensure that the independent variable sets out to measure what it was intended to measure.

Analyses were first conducted for explicit decision to determine accuracy rates and if an innocence-guilt bias occurred. As for implicit responses, a Kolmogorov-Smirnov test for normality (Massey, 1951; Smirnov, 1948) was conducted, and because the K-S test showed non-normal data (as it is used to assess two samples with a reference probability distribution) a Wilcoxon's Rank Test for within-subjects cases was then conducted. To test if there was a statistically significant relationship between the accuracy of judges' explicit decisions and level of appeal emotionality shown to them, a Chi-square was computed.

Logistic regressions for within-group analyses were also carried out to determine which implicit responses judges used to predict their explicit decisions. This analysis was used to explore the relationship between explicit and implicit decisions, regardless of whether judges may have attended to implicit cues consciously or unconsciously and meaningfully or otherwise. Regression was chosen as an analysis as it allows an understanding of which among the independent variables (implicit cues) are related to the dependent variable (explicit decision), and to explore the forms of these relationships. Variance inflation factors (VIF) are also analysed. VIF values that exceeds 10 and Tolerance values being less than 0.10 are often regarded as indicating multicollinearity (Dart, 2017).

In the current study, it was acknowledged that the number of independent variables observed was high relative to sample size. Overfitting can result in meaningless variables entering the model merely by chance, even if the model seems to have a good fit with the data used. Babyak (2004) stated that the issue with overfitted models is that results may not be replicated with a different sample, and thus findings with the current sample should be interpreted with caution. Nevertheless, all predictor variables included in regression analyses were each derived (and justified inclusion) from an extensive review of the literature expounded in the opening chapters of the present thesis. While the possibility exists that too many variables in a model leads to the potential issue of overfitting, Zhang (2014) stated that including a high number of variables as possible to

examine the association between a predictor and outcome variable can be beneficial and controls for confounding variables.

The bootstrapping process is also applied as it is very useful for small sample sizes (Adèr, Mellenbergh & Hand, 2008) and works by making extrapolations of a population from a sample data. This process samples this data and then makes extrapolations about a sample from re-sampled data (Efron & Tibshirani, 1993). In the present study, this process is repeated 1000 times. Both bootstrapped models for high and low emotionality will contain the same twelve independent variables (plausibility, story sense, credibility, practice, well-thought-out, wording, less-emotions, emotions-matched, organised-speech, attractiveness, voice attractiveness and meta-emotion level).

### *7.7.3 Ground truth*

Appeal 1 shows an innocent female appealing directly to the killer who murdered her daughter. Appeal 2 shows an innocent female appealing directly to the abductor of her missing daughter. Appeal 3 shows an innocent male appealing for the return of his granddaughter. Appeal 4 shows an innocent male appealing for the return of his missing nephew. All appellants made a direct appeal to the killer or captor, the missing person to get in touch, or to the public for information or assistance to search for the missing or murdered relative. In all four cases, substantial evidence was found to reinforce the appellants' claims concerning the fate of their relatives. For Appellant 1, another man was convicted of killing her daughter, with evidence of an un-recanted confession, eyewitness accounts, surveillance video footage, public tip, and recovered bone fragments. For Appellant 2, another man was charged with coercing, enticing and transporting a minor to engage in sexual activity with the appellant's daughter, with evidence of confession and eyewitness accounts. For Appellant 3, his granddaughter's body was found and post mortem examination suggests suicide with no evidence of foul play involved. Finally, for Appellant 4, his missing nephew returned home with no evidence of foul play involved.

## **7.8 Results**

### 7.8.1 Participants

After removing cases where participants had knowledge of the appeals, there were 47 participants but only a total of 108 cases remaining for further analysis. The final sample consisted of 28 participants (45 cases) for the High Emotional condition and 19 participants (63 cases) for the Low Emotional condition. As some participants knew of more than one out of the four appeal cases, the total number of cases were lower than the overall number of participants. As individual differences were anticipated in appeal emotional level appraisal, the final sample only included participants who agreed with the question ‘The appellant displays less emotion than seems appropriate for their situation’ for low emotional appeals and disagreed for highly emotional appeals. Also, as individual differences were anticipated in appeal emotional level appraisal, the final sample only included participants who agreed with the question ‘The appellant displays less emotion than seems appropriate for their situation’ for Low emotional appeals and disagreed for High Emotional appeals. The final pool of participants ranged from 20 to 61 years in age ( $M = 29.61$ ,  $SD = 10.53$ ).

### 7.8.2 Explicit veracity judgments

As shown in Table 1, Highly Emotional appeals reported a veracity rating of 56.00%, whereas Low Emotional appeals reported a rating of 40.60%. An innocence bias did not appear to be present. Implicit assessments of veracity are examined in the next section.

**Table 1**

*Judgment rates for appeals for High and Low Emotionality conditions*

	High Emotionality	Low Emotionality
Appeals	56.00%	40.60%

### 7.8.3 Relationship between level of appeal emotionality and explicit accuracy rates

Chi square results were  $\chi^2 (1) = 31.410, p < .05$ . This revealed that there was a statistically significant association between accuracy of judges' explicit decisions and level of appeal emotionality shown to them. The two variables (accuracy of explicit ratings and High versus Low Emotionality) are dependent and/or related. Phi values indicate the strength of this relationship and are chosen over Cramer's V as both variables used in the sample are dichotomous (Jones, 2009). The strength of association between the variables was high ( $\phi = .54$ ), demonstrating the presence of a substantial relationship (Davis, 1971).

#### *7.8.4 Implicit veracity judgments*

Results for the K-S test for normality indicated that the score distributions deviated significantly from a normal distribution ( $D = .123, p < .005$ ). Due to the within-subjects design of the study and the assumption of normality being violated, a Wilcoxon Signed Rank test was conducted (Wilcoxon, 1945). Descriptive statistics for participant truth scores between High and Low Emotionality, results from the Wilcoxon test, and effect sizes can all be found in Table 2. Cronbach's alpha, the estimated internal consistency, for each scale items can be found in Table 3. Implications are reviewed in depth in Chapter 11 under Sections 11.6 and 11.7. The Wilcoxon test would suggest differences in perceptual differences between both conditions. Story plausibility was judged significantly higher when judges were watching appealers who were visibly distressed, with a mean of 2.73 and a standard deviation of 1.116 in comparison with appealers who lacked signs of visual distress at a mean of 1.71 ( $SD = 1.300$ ). Similarly, the appealers' story was perceived to make more sense significantly more in highly emotional appeals with a mean of 2.69 ( $SD = 1.104$ ) compared to low emotional ones which recorded a mean of 2.06 ( $SD = 1.189$ ). The same is said for perceived appealers' credibility. Judges recorded a significantly lower mean for emotion matching in the low emotionality condition at 1.53 ( $SD = 0.815$ ). The two conditions recorded significantly different meta-emotion levels, with appealers who were visibly distressed receiving a higher mean of sympathetic reaction and character identification at 11.93 ( $SD = 4.539$ ) and those who lacked visible distress at a mean of 8.18 ( $SD = 2.949$ ).

**Table 3***Cronbach's alpha coefficients for Study 1*

<b>Meta-emotion</b>	<b>Cognition</b>	<b>Appearance</b>	<b>Emotions</b>
0.74	0.47	0.30	0.54

*7.8.5 Logistic regressions analyses*

Throughout Studies 1 to 4, Guilt was coded as 0 and Innocence was coded as 1. To assess what implicit cues judges may have used to predict their explicit veracity judgments within both conditions, binary logistic regressions were performed for both. Logistic regressions were carried out to investigate these perceived indicators of veracity as well as participant accuracy as dependent measures for both conditions. In this study, tolerance was greater than .10 for all items, and the variance inflation factor was also less than 10 for all items. This suggested that multicollinearity was not an issue in this study (meta-emotion, Tolerance = 0.64, VIF = 1.55; credibility, Tolerance = 0.67, VIF = 1.49; plausibility, Tolerance = 0.66, VIF = 1.52; practice, Tolerance = 0.88, VIF = 1.14; voice attractiveness, Tolerance = 0.80; VIF = 1.09; organised speech, Tolerance = 0.69; VIF = 1.44; face attractiveness, Tolerance = 0.91; VIF = 1.10; less emotions, Tolerance = 0.60; VIF = 1.66; wording, Tolerance = 0.84; VIF = 1.19; well-thought-out, Tolerance = 0.74, VIF = 1.34; emotions match, Tolerance = 0.62, VIF = 1.62; practice, Tolerance = 0.88, VIF = 1.14).

First consideration was given to only Highly Emotional appeals. This helps predict which implicit veracity judgments judges could have used in parallel with their explicit decisions. The full regression model indicates that it was able to distinguish between respondents who explicitly reported a Guilty judgment from the ones who reported an Innocent judgment, with results reported as  $\chi^2(12, N = 108) = 332.77, p < .001$ . The model explained between 32% (Cox and Snell  $R^2$ ) and 42% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 79% of cases.

Consideration was then given only Low Emotionality appeals and accordingly which implicit veracity judgments did participants utilised supported their explicit one. The full model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent with results reported as  $\chi^2(8, N = 108) = 32.34, p < .001$ . The model explained between 27% (Cox and Snell  $R^2$ ) and 37% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 81% of cases.

Although VIF values suggest that multicollinearity was not an issue in this study, precautions are also taken to reduce false positive findings. False positive findings are also synonymous with the fishing and error rate problem, or alpha inflation (Parker & Szymanski, 1992). This threat occurs when a high number of multiple statistical comparisons are being conducted and the possibility of making a Type I error increases with each comparison in attempt to test hypotheses and finding a significant effect (Parker & Szymanski, 1992). To lower this risk of false positives and to increase confidence in results, a more stringent cut-off point will be employed for the regression analyses ( $p < 0.01$ ). As shown in Table 4, none of the independent variable made a unique statistically significant contribution to the model using this cut-off level for both condition.

#### *7.8.6 Not-sure answers frequencies*

As described in the Methods section, not-sure answers for all implicit decisions were recorded for the remaining cases retained for analyses. Judges who reported 'not sure' for the question item pertaining to emotion level were already removed to ensure only those who were sure were included in this study. For the remaining implicit cues, only percentage differences between both conditions that are higher than 10% are discussed. A table can be found in Appendix I showing all percentages of not-sure answers for both conditions. From this table, it can be seen that when low emotional appeals were shown, judges found it harder to rate the appealers' story plausibility and level of practice.

**Table 2***Wilcoxon's Rank implicit assessments between High and Low Emotionality conditions*

	High Emotionality means	Low Emotionality means	<i>z</i>	<i>r</i>
<i>Truth score</i>				
<b>Credibility</b>	<b>2.222 (1.396)</b>	<b>1.403 (1.137)</b>	<b>-2.482*</b>	<b>-0.24</b>
<i>Cognition of Source Content scores</i>				
<b>Plausibility</b>	<b>2.733 (1.116)</b>	<b>1.714 (1.300)</b>	<b>-3.526*</b>	<b>-0.34</b>
<b>Story Sense</b>	<b>2.689 (1.104)</b>	<b>2.064 (1.189)</b>	<b>-2.482*</b>	<b>-0.20</b>
Practice	1.511 (1.160)	2.063 (1.256)	-1.758	NA
Well-thought-out	2.222 (1.259)	2.095 (1.411)	-0.418	NA
Organised	2.578 (1.138)	2.476 (1.189)	-0.132	NA
Wording	1.933 (1.194)	1.667 (1.191)	-1.535	NA
<i>Emotion scores</i>				
<b>Less Emotions</b>	<b>1.644 (0.484)</b>	<b>3.286 (0.633)</b>	<b>-5.640*</b>	<b>-0.54</b>
<b>Emotion-match</b>	<b>2.933 (1.095)</b>	<b>1.533 (0.815)</b>	<b>-4.580*</b>	<b>-0.44</b>
<i>Appearance scores</i>				
Voice attractiveness	1.600 (1.355)	1.571 (1.214)	-0.263	NA
Face attractiveness	1.181 (1.062)	0.935 (0.939)	-0.640	NA
<b>Meta-emotion score</b>	<b>11.933 (4.539)</b>	<b>8.175 (2.949)</b>	<b>-4.313*</b>	<b>-0.42</b>

Significant *p* values are in bold ( $p < 0.01$ ). Standard deviations appear in parentheses.

**Table 4**

*Bootstrapped binary logistic regression models for explicit judgments between High and Low Emotionality conditions*

Variables	Model 1: High Emotionality				Model 2: Low Emotionality			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	.157 (.0240)	1.170	.730, 1.873	0.900	.494 (0.218)	1.640	1.069, 2.515	0.024
Plausibility	.487 (.0264)	1.627	.969, 2.732	0.066	.017 (0.230)	1.017	.648, 1.595	0.942
Story Sense	.114 (5.981)	1.132	.680, 1.884	0.691	.359 (0.228)	1.433	.916, 2.240	0.115
Practice	-.442 (0.235)	0.643	.405, 1.020	0.060	-.147 (0.209)	0.864	.574, 1.300	0.483
Wording	.163 (0.258)	1.177	.710, 1.951	0.529	.277 (0.234)	1.319	.834, 2.084	0.237
Well-thought-out	-.038 (0.222)	0.963	.623, 1.487	0.864	-.390 (0.193)	0.677	.464, .989	0.043
Organised	.138 (0.240)	1.147	.717, 1.838	0.567	-.032 (0.242)	0.969	.603, 1.557	0.896
Less Emotions	-.491 (0.230)	0.612	.390, .960	0.033	-.359 (0.210)	0.698	.463, 1.054	0.087
Emotions Match	.029 (0.233)	1.030	.652, 1.627	0.900	.009 (0.255)	1.009	.612, 1.662	0.973
Face	-.090 (0.281)	0.914	.527, 1.587	0.750	-.135 (0.249)	0.874	.536, 1.422	0.587
Voice	-.128 (0.223)	0.880	.633, 1.071	0.633	-.055 (0.227)	0.946	.606, 1.478	0.809
Meta-emotion	.136 (0.084)	1.146	.971, 1.352	0.107	.034 (0.075)	1.035	.894, 1.198	0.649
<b>Model description</b>	Model is statistically significant: $\chi^2 = 31.87$ ; df = 12; Nagelkerke's $R^2 = .423$ ; Hosmer-Lemeshow test: $\chi^2 = 7.66$ ; df = 8; p = .467				Model is statistically significant: $\chi^2 = 32.94$ ; df = 12; Nagelkerke's $R^2 = .38$ ; Hosmer-Lemeshow test: $\chi^2 = 13.93$ ; df = 8; p = .084			

Note: Significant  $p$  values are in bold;  $p < 0.01$ . Standard errors appear in parentheses.

## 7.9 Discussion

The present chapter examined implicit judgments in relation to the variable of emotional display in the context of high stakes television appeals. Findings would tend to suggest support for the main hypothesis of this study; if an appeal had low emotional content (visually and verbally) judges were more likely to believe the appellant to be guilty.

With explicit accuracy rates, it was found that judges assessed appellants to be more innocent when they were crying or sobbing with the presence of a cracking voice, while verbally declaring that they 'love' and 'miss' the victim. In contrast, appellants who lacked visible distress in their facial expressions with no crying or cracking voice present and no verbalised declaration of affection towards the victim were more likely considered to be lying. Judges could have based their explicit judgments on how similar to a stereotype of a grieving individual they were familiar with. This stereotype included salient physical and emotional symptoms of acute grief that includes crying, sighing and shortness of breath for example, as exemplified in earlier theories of grief. In the review of literature individuals tended to favour a particular 'stage' of grief, more often an acute form. Maciejewski, Zhang, Block and Prigerson (2007) stated how this stage theory of grief has become well-accepted and putative, widespread to generalise a wide variety of losses, and this unsupported stage process is relied upon even in medical education. Here in the current chapter, judges seemed to have used their personal expectations of how someone who has just lost a relative should behave. When the appellants' emotional behaviour matched their expectations, they tended to favour these appellants by awarding them a higher innocence rating.

In the within-group analyses of judges' implicit veracity judgments, several implicit judgments emerged as significantly different between both groups with varying effect sizes. Story plausibility, story sense and overall appellant credibility were judged to be higher in the Highly Emotionality appeals when compared with the Low Emotionality ones, indicating that perceived level of appeal emotionality and verbalised affection towards the victim influences even how plausible a story sounds. A caveat mentioned in the Methods section was that the appeal content differed for all appeals, which could have affected judges' veracity judgments. In

other words, these differences could be due to a clearer and convincing appeal rather than a higher level of emotional display that led judges to believe an appellant is more credible. For example, appeals that were low in emotional display level could have simply made less sense and were less plausible than those that were high in emotional display. Alternatively, it could also be that the presence of verbalised emotions portrayed the highly emotional appellants as more honest in their affections and longing for the victim, thus making their cases more compelling.

The results in the current study would seem to suggest that there are differences in explicit ratings, and there are differences in thinking when people watch a highly emotional video versus one that is not as can be surmised from Table 2. Theories and studies discussed in the literature review in the present chapter would seem to suggest that people tend to favour highly salient emotional displays and judge them to be more innocent (Bollingmo et al., 2008; Baldry et al., 1997; Breen & O'Connor, 2007; Costa & Stewart, 2007; Peace & Sinclair, 2012). Judges' meta-emotion ratings were significantly different for both conditions, suggesting that Highly Emotional appeals increased judges' own identification with the appellant in comparison to the appeals that were low in emotional display. However, sympathetic reactions and character identification did not predict their explicit decisions of whether the appellant was innocent or guilty, in both conditions. Therefore, while a visibly distressed appellant elevated sympathetic reactions of judges this did not regulate their cognitive processes, inferring that this did not contribute to explicit decisions (whether knowingly or otherwise).

There was also no appearance and attractiveness bias present at least in the current sample, probably in part due to the lack of manipulation of this variable in this study. No conscious effort was made to find a set of appellants who were attractive and two who were not, partially because it is difficult with a limited sample of high stakes data and partly because it was not a main aim of the present study. Rather, this implicit judgment was included to examine the likelihood it was used in regulating veracity decisions via superficial processing.

Tentative inferences can be drawn from the present study, but warrants further studies with a higher number of videos and trials. This conjecture would

benefit from further research as to whether level of emotional display really causes a change in the way information is processed by people and whether this causality subsequently alters and/or affects guilt-innocence decisions.

An issue with using only innocent appealers was that it cannot be determined from the regression analyses as to whether results are predicting accurate explicit decisions, or simply what implicit judgments participants were using for the appeal videos. This issue will be addressed in the next chapter. The next chapter will examine the role of presentation format to remove or at least reduce both the meta-emotion and appearance biases. Though this chapter presented limitations in working with high stakes data, the importance of using ecologically valid stimulus material cannot be overvalued.

Another limitation of the present methodology considers the base rate argument and chance veracity assessments. For example, should the base rate of the stimulus to be assessed be 90% truthful and judges guessed truth all the time, judges would obtain 90% 'accuracy' simply by guesswork. This accuracy is simply an artefact (Burgoon, 2015a), rather than judges achieving a 90% *accuracy*. Thus, using only truthful appeals makes it problematic to study an innocence detection accuracy (true negative rate), or specificity. The remainder of studies in the present thesis will employ a 50/50 ratio of truthful and deceitful appeals to control for innocence and guilt biases. Using this methodology, an examination of not only guilt accuracy (sensitivity) also known as true positive rate, can be reported as well as innocence accuracy (specificity) or true negative rate (Burgoon, 2015a).

Another apparent limitation was that content could not be controlled for in the appeals used in this study. Content of highly emotional appeals differed from those with low levels of emotion (it was simply unrealistic to locate a highly emotional appeal and a low emotional appeal with the exact same content), and this may affect veracity judgments separately to the emotional component. While chi square results would suggest a relationship between how accurate judges were in their explicit decisions and the level of emotionality, these results are of course still subject to further replication and investigation. Content differences still remain an issue, and in order to address this, the subsequent study will follow a different methodological approach, employing a between-subjects methodology.

Lastly, by applying the phase of only including Judges who agreed with the author's categorisation of the emotional level in the chosen videos, this raises the possibility of rendering certain cues as correlational. For example, a Judge may think that an appealer is telling the truth and subsequently that they look emotional, or that the appealer is lying and subsequently that they do not look emotional either, or even that they look emotional and therefore they must be telling the truth. Currently there is no way to tell whether cue A affected cue B, or the other way around is true (see Section 9.9 for a brief discussion of this). This inter-relational process of the association between cues still requires further distilling in future studies.

## CHAPTER 8: STUDY 2

---

### PERCEPTION OF SOURCE PRESENTATION IN TELEVISION APPEALS

From the results presented in Study 1 (Chapter 7) it was recognised that perceived emotional display has, to a certain extent, an impact on eliciting mental shortcuts. For example, judges felt significantly higher levels of meta-emotion with highly emotional content compared to ones lacking them. At times, some of these biases then did not contribute to explicit veracity decisions (i.e. one may have felt significantly more sympathetic towards a highly emotional applier but it may have no bearings towards how trustworthy they think they are, as shown in the previous chapter). Results from the previous chapter may have been convoluted by the differences in the actual content of the appeals. The current chapter refines its experimental methods and proposes that presentation format may prompt certain heuristics to be utilised and others to be suppressed.

The rationale in studying presentation mode stems from several research studies which allege visual bias in regulating judgments and decision making, as will be expounded below in the literature review section of this chapter. Ekman (1981) stated that non-verbal modes of communication are particularly useful in detecting deception, but only so with the presence of emotions. Moreover, in situations that present relatively low motivation to deceive, the possibility is lower that emotions should emerge. Conversely, television appeals present high motivation to deceive if guilty and to solve the investigation if innocent, with appellars relying on a visual channel to plead their case.

In order to understand how visual cues may affect innocence-guilt judgments, existing literature on presentation format are discussed, together with their impact on cognitive processes, sentencing and veracity judgments.

## **8.1 Presentation format**

Chapter 2 showed how Source content can alter and influence its believability. The current chapter examines manipulation of Source characteristics – its presentation mode.

As discussed in Chapter 2, Bond and DePaulo's (2006) meta-analyses revealed that truth-lie judgment accuracy declines when judging visible compared to audible lies. The authors suggested that one of the reasons for this finding is because the typical typecast of someone who is lying is most easily brought out

in a visual presentation mode. When someone has to judge the veracity of a sender with limited information or evidence other than a given video, they do not have many other alternatives but to resort to their understanding and stereotypes (Bond & DePaulo, 2006), which may potentially lead to biased judgments. Currently, it is still largely unknown if comparable results will be revealed for innocence-guilt judgments using television appeals.

In addition, support for reviewing presentation format comes from other studies in the lie detection literature such as one by Anderson et al. (1999). Using pairs of friends as participants, results indicated that some the cues generated from their open-ended responses were from the appearance of their friend such as overall demeanour (i.e. they were less comfortable when telling the story), salient emotions (they were crying) and visual cues (i.e. they looked at me in the eyes). Their study showed that participants tended to mention more verbal cues, such as the content of the story or plausibility when judging an honest account, and in comparison, more likely to mention cues pertaining to appearance and demeanour such as nervousness and eye contact level when judging a deceptive account.

Anderson et al. (1999) offered two explanations for why this difference in tendencies to mention different cues for truths and lies exist. The first explanation they posited was that perhaps their participants were observing reliable cues; however, the participants were not using said cues in their assessments. For example, the authors contended that if their friend's story content was truthful, they perhaps picked up on aspects of the story content that were honest. When the story content is not so convincing and lacks believability, they may have been doubtful and sought after other cues to deception such as visual cues. Anderson et al. (1999) contended that it is possible observers first note verbal cues, and only when this cue is unsuccessful in persuading them, they move to a different cue. The authors also speculated from these results that their participants might have not fully understood the meaning of their own assessments. For example, they failed to recognise that the movement from pursuing story plausibility and consistency to visual cues could possibly be a hint that their friend may be lying. The authors also presented the possibility that another reason why different cues were mentioned for truthful versus deceptive stories were not due to differences

in attention, but in the memory retrieval processes and/or in the cue reporting stage. They suggested that participants who heard honest accounts perhaps encoded and/or retrieved memories relating to story content comparatively at a higher level than visual cues.

Anderson et al.'s (1999) study indicated that a difference in implicit information processing occurs for truthful and deceptive accounts, and that the sources endorsing the implicit judgments were either visual or verbal. In relation to the present thesis and in consideration of the previous chapter's limitations, the question is then raised as to whether implicit judgments will be affected when the visual cue of television appeals is removed and whether these differences further extends to appeals made by innocent and guilty senders. In addition, removing the visual cue potentially eliminates the probability of certain biases and heuristics occurring such as the appearance/attractiveness bias.

## **8.2 Cognitive loading**

Additional support for increased accuracy when presented with only one channel comes from cognitive loading theory. Besides being more prone to incorrect reliance on visual cues, observers may also undergo an increase in cognitive load when perceiving an added visual element in the audio-visual format, let alone multiple appealers at times. An increase in cognitive load implies either an increase in the amount of information elements or sources to be processed or that a higher demand for processing is consigned for the same amount of information coming from the same source (Lavie, 2005; Sweller, 2010). The cognitive load theory (CLT) posits that at any one time, a working memory can process only a limited amount of information pieces (Miller, 1956). Chandler and Sweller (1996) asserted that CLT does not always recommend the integration of information from multiple sources. It is only required when the split information sources are incomprehensible when not integrated and do not supply all necessary information.

Research in several fields has shown that imposing cognitive load decreases performance, accuracy in performing tasks, motoric functions or memory (Vredeveldt, Hitch & Baddeley, 2011; Logie, Gilhooly & Wynn, 1994; Yoge-

Seligmann, Rotem-Galili, Dickstein, Giladi & Hausdorff, 2011; Longstaffe, Hood & Gilchrist, 2014; Mann & Vrij, 2006; Weaver & Stewart, 2012). Otten, Alain and Picton (2000) found that visual loading impacts auditory processing, enhancing task performance. Hoffmann, von Helversen and Rieskamp (2013) recently discovered that cognitive load can, on the contrary, be advantageous rather than detrimental. In their study, they found that cognitive load encouraged a change to a similarity-based judgment strategy, which subsequently led to more accurate judgments. The authors argued that the outcome of cognitive loading boils down to the specific circumstance and/or type of load.

Based on the number of studies mentioned above, research suggests that cognitive load impairs performance and judgment on critical tasks. Removing the visual cue in television appeals in effect imposes less cognitive load onto observers. By separating the visual dimension from the audio in the current study, it can be observed whether observers can perceive veracity and deceit through one channel of source information. Having examined the literature, including those on cognitive load theory, it is hypothesised that participants exposed to the Audio-only condition will achieve higher explicit rating compared to those in the Audio-visual condition.

### **8.3 Attractiveness and appearance bias**

As aforementioned, by presenting an audio-only condition to observers this, to all intents and purposes, sees to the removal of a visual cue bias. Encompassed within this visual cue bias is the attractiveness and appearance bias. To reiterate, it is well documented that the attractiveness of a speaker has an effect in veracity decision making as exemplified in Chapters 2 and 7. The general implication from much of research in social psychology and in criminology shows that one's attractiveness and appearance impacts judgments of those judging their veracity.

In the last chapter, it was found that some judges did not agree with the operational definition chosen for what constitutes a highly emotional appeal and what constitutes a low emotional appeal (see Section 7.9). This step was taken to ensure that the intended independent variable is really measuring what it is supposed to. However, in the current chapter (while still aspiring to assess the

visual influence on decision making), a way to bypass this issue is to manipulate presentation mode instead. In doing so, the question as to whether implicit judgments will be influenced by visual cues can be evaluated.

In the current chapter, the implicit variable of voice attractiveness is a point of interest. While observers may not be able to visually see how attractive an appealers is, a person's voice can also contribute to how attractive a person is perceived to be (Ekman, Friesen & Scherer, 1976). Sound or vocal attractiveness can be referred to as how pleasant or unpleasant someone sounds to listeners (Bruckert, Bestelmeyer, Latinus, Rouger, Charest, Rousselet, Kawahara & Belin, 2010). Hughes, Dispenza and Gallup Jr. (2004) found that how participants rated the attractiveness of a voice correlated and predicted sexual behaviour and body configuration. Another study with regards to the effects of sound and how it affects judgement was conducted by Surawski and Ossoff (2006), where participants rated the physical and vocal attractiveness of politicians. The authors found that if a politician had an unattractive voice, the ratings for this politician were low even if the politician was physically attractive. Therefore, even with the removal of a visual appearance and attractiveness bias, the perceived attractiveness of an appealers voice may lead to superficial information processing. Moreover, this potential bias has never been tested in an audio-only format in comparison with an audio-visual one.

#### **8.4 Innocence bias**

In lie detection literature, Buller, Strzyzewski and Hunsaker (1991) observed that face-to-face communication was more easily affected by truth bias whereas participants viewing through other media were not as susceptible. It is not known whether an innocence bias will carry over when judges are asked to make innocence-guilt assessments.

#### **8.5 Meta-emotion bias**

Similarly, a meta-emotion bias may be affected due to the elimination of visual cues, which can prompt differences in implicit judgments according to how television appeals are presented format-wise. Hussain, Calvo and Chen (2013)

proposed that viewers experience cognitive loading when presented a task which requires an emotional response. The emotional response impedes and impacts upon their accuracy in performing the task as well as their behavioural response towards it forming a circular argument. When watching an emotionally laden television appeal, for example, observers may perceive and experience certain emotions when they believe that the applier is lying, which possibly prompts then to search for visual cues that may point to their deception. The removal of a visual element may reduce perceived and experienced emotions to some level. In turn this can result in fewer searches for invalid indicators of deception to support the experienced emotions.

### **8.6 Cognition of Source content bias**

Likewise, these implicit variables introduced in previous chapters may also be influenced depending on the presentation format.

### **8.7 Key issues to explore**

The general premise of the current study is that by presenting a television appeal in an Audio-only format, the observers will be removed from unnecessary and even misleading visual variables (that may or may not be related to deception). By removing the visual element, not only is cognitive load hoped to be reduced, the issue of using a visual-based heuristic is made unavailable. The main hypothesis is that there will be a difference in explicit accuracy rates for both mediums, with the accuracy rates improved when the appealers are presented audio-visually compared to an audio-only channel. An additional hypothesis is to uncover predictors of veracity judgments, with the hypothesis being predictors will vary for both channels, due to expected underlying cognitive mechanisms in both channels. Observers may adapt and use different implicit cues to aid themselves in making veracity decisions – in a condition where both visual and verbal cues are available and salient to them and differently in a condition where the visual cue is not available.

### **8.8 Methods**

### *8.8.1 Participants*

This sample of participants totalled to 180 participants for both groups, with 720 cases as each participant judged four cases of these appeals. None of the participants personally knew any of the featured appellants in this study but some reported knowledge of certain appeals from watching or hearing it on television. They were recruited using a mixture of snowball and opportunity sampling from the UK and Europe. They were from varying occupations ranging from financial advisor to engineer to chemist to carpenter to cleaner to gardener to insurance broker and more; as well as varying nationalities/ethnicities (British, Asian, African, Middle Eastern and European).

### *8.8.2 Material and procedure*

A methodological consideration in the current study is the usage of a between-subjects design instead of a within-subjects one as in the previous chapter. The advantage of using a between-subjects design is the ability to determine not only what implicit decisions were used by judges, but also what generated and predicted accurate explicit decisions. In addition, this design controls for appeal content, in that both conditions will be exposed to the same appellants with the same appeal context, thus there is no issue of sender detectability here. This was impossible for Study 1 as one cannot manipulate or change an appellant's emotional display in real life high stakes data. The appeal content will remain the same for this study. By using two different presentation modes with the same appeal content, it can be said with a higher level of certainty that implicit judgments are directly related with the experimental conditions. Furthermore, the use a between-subjects design decreases demand characteristics compared to a within-subjects design (Rubin & Badaea, 2010; Orne, 1962).

Another refinement in the present study is the use of both true and false appeals as an independent variable. Evidence from previous research suggests that differences exist in processing and veracity judgments, with regards to the actual veracity of the source (Just & Carpenter, 1976; Johnson-Laird & Savary, 1999). Not sure answers for all implicit decisions will be recorded for the remaining cases retained for analyses.

As a between-subjects experimental design, two sets of press conferences will be selected for the purpose of this study. In the first set, judges will be exposed to press conferences in an Audio-visual format. In the second, the same press conferences will be viewed by a different group of judges without the visual element. Thus, all judges will experience the exact same Audio content of the appeals in a non-randomised order; the only difference being that one group will watch the appeal with a visual element of the appellants pleading and the other group did not. The videos will be put into a PowerPoint presentation, and range from 34.21 seconds to 160.09 seconds in length with an average of 69.74 seconds each. The general methodology of how the experiment will be conducted is outlined under Section 6.3 in Chapter 6.

Regarding the content of the appeals, some of the appeals are interviews and some others are direct appeals; this is due to the limited choice of appeals present for the current study. It is recognised that presentation of videos may confound with veracity. There is a likelihood that interviews could be more persuasive (or otherwise) than direct appeals. That said, since the current study executes a between-subjects experimental design, all judges in both conditions will experience the same set of appeals be it interviews or direct appeals in terms of appeal content. Thus, the fact that not all press conferences selected are direct appeals will not affect between-subjects responses. Judges will follow the same instructions and procedure as Study 1 in the last chapter. The same method of analysing results will be conducted in this study. To test if there was a statistically significant relationship between the accuracy of judges' explicit decisions and presentation mode, a Chi-square will be computed.

As the previous chapter, logistic regressions for within-group analyses will also be carried out to determine which implicit responses predict their explicit decisions. This analysis will be used to explore the relationship between explicit and implicit decisions, regardless of whether judges may have attended to implicit cues consciously or unconsciously and meaningfully or otherwise. Regressions will also be conducted to determine which implicit responses predict accurate explicit decisions. Variance inflation factors (VIF) will also be analysed to check for multicollinearity (Dart, 2017). Two additional Mann Whitney U tests will be

conducted for honest and false appeals between both conditions as the present study utilises two true and two false appeals and judges may utilise different implicit cues for truthful and false appeals.

Unlike the previous chapter, the present study will only contain ten independent variables in all bootstrapped models (plausibility, wording, organised-speech, story sense, credibility, practice, well-thought-out, voice attractiveness and meta-emotion level). The variables of emotions will be removed as judges in the Audio-only condition will not be able to visually appreciate these. The bootstrapping process was applied here (Efron, 1979). This process operates by generating a new sample from a given sample size. It is very useful for small sample sizes (Adèr, Mellenbergh & Hand, 2008) and works by making extrapolations of a population from a sample data. This process samples this data and then makes extrapolations about a sample from re-sampled data (Efron & Tibshirani, 1993). In the present study, this process is repeated 1000 times.

### *8.8.3 Ground truth*

Appeal 1 shows an innocent female giving an account of her murdered daughter and directly appeals to the killer who murdered her daughter. Interview 2 shows a guilty female giving an account of what happened the day her boyfriend went missing. Interview 3 shows an innocent male giving an account of what happened the day his wife went missing and directly appeals to the public for information. Appeal 4 shows a guilty male giving an account of what happened the day his granddaughter went missing. In all four cases, substantial evidence was found to either discredit false appellant's claims or reinforce true appellants' claims concerning the fate of their relatives. For Appellant 1, another man was convicted of the crime of killing her daughter; for Interviewee 2, various pieces of evidence were presented in court including a gun, rented car, phone, friends, camera, palm print DNA; for Interviewee 3, his alibi stands and another killer was convicted of raping and murdering his wife; and for Appellant 4, a tip from a member of the public and the found body were the main evidence.

## 8.9 Results

### 8.9.1 Participants

After removing cases where participants had knowledge of the appeals, 616 cases remained for analysis. The number of participants was  $N = 180$ , with 94 participants (317 cases) for the Audio-only condition and 86 participants (299 cases) for the Audio-visual condition. The final pool of participants ranged from 17 to 75 years in age ( $M = 29.96$ ,  $SD = 12.62$ ).

### 8.9.2 Explicit veracity judgments

As presented in Table 6, judges in the Audio-only condition recorded an overall rating of nearly 70%. This rating was higher than what was reported by those in the Audio-visual condition which, at 53.45%, was very near chance level.

### 8.9.3 Relationship between source presentation and explicit judgment rates

Chi square results were  $\chi^2(1) = 13.761$ ,  $p < .05$ . This revealed that there was a statistically significant association between accuracy of judges' explicit decisions and presentation mode. The two variables (accuracy of explicit ratings and Audio-only versus Audio-visual) are dependent and/or related. The strength of association between the variables was .163, indicating a weak relationship (Davis, 1971). Phi values indicate the strength of this relationship and chosen over Cramer's V as both variables used in the sample are dichotomous (Jones, 2009). The strength of association between the variables was small (.163), demonstrating the presence of a weak relationship (Real Statistics Using Excel, 2017).

### 8.9.4 Innocence bias

Table 5 shows judgment rates for true and false appeals for both conditions. The results of the current study did not support the presence of an innocence bias. Furthermore, accuracy was lower for truthful appeals compared to false appeals in the Audio-visual condition. In the Audio-only condition however, this pattern was reversed.

## Table 5

*Judgment rates for True and False appeals for Audio-only and Audio-visual conditions*

<b>Appeals</b>	<b>Audio-only</b>	<b>Audio-visual</b>
<b>False</b>	<b>65.70%</b>	<b>54.90%</b>
<b>True</b>	<b>72.60%</b>	<b>52.00%</b>
<b>Total</b>	<b>69.20%</b>	<b>53.45%</b>

*8.9.5 Implicit veracity judgments*

Non-parametric tests were applied to questionnaire items analyse in greater detail the differences in implicit veracity judgments. Results for the Kolmogorov-Smirnov test for normality (Massey, 1951; Smirnov, 1948) indicated that the score distributions deviated significantly from a normal distribution ( $D = .102, p < .005$ ). Cronbach's alpha, the estimated internal consistency, for each scale items are presented in Table 6. Again, implications are reviewed in depth in Chapter 11 under Sections 11.6 and 11.7. Results from the Wilcoxon test, and effect sizes ( $r$ ), can be found in Table 7.

**Table 6**

*Cronbach's alpha coefficients for Study 2*

<b>Meta-emotion</b>	<b>Cognition</b>
0.56	0.35

Implicit decisions revealed several underlying perceptual differences of judges for the two conditions. Interestingly, the appellant's voice was perceived to be significantly more attractive in the Audio-visual condition with a mean of 2.30 ( $SD = 1.322$ ) compared to those in the Audio-only condition which recorded a mean of 1.99 ( $SD = 1.268$ ). The two conditions recorded significantly different meta-emotion levels, with judges who could only hear the appeals recording a lower mean of sympathetic reaction and character identification at 9.54 ( $SD = 3.568$ ) and those who were in the Audio-visual condition at a mean of 10.44 ( $SD = 3.123$ ).

Judges in the Audio-visual condition reported a higher mean for whether the appellant's wording was assembled in a manner that was pleasant to hear at a mean higher ( $M = 2.25$ ,  $SD = 1.05$ ) than those in the Audio-only condition ( $M = 1.87$ ,  $SD = 1.09$ ).

**Table 7**

*Mann Whitney U implicit assessment for Audio-only and Audio-visual conditions*

	<b>Audio-only mean</b>	<b>Audio- visual mean</b>	<b>Z</b>	<b>r</b>
<i>Innocence</i>				
<i>score</i>				
Credibility	2.046 (1.213)	1.777 (1.307)	-2.118	NA
<i>Cognition of Source</i>				
<i>content scores</i>				
Plausibility	2.025 (1.251)	2.127 (1.224)	-0.161	NA
Story	1.986 (1.319)	2.087 (0.899)	-0.516	NA
Sense	2.004 (1.125)	1.891 (1.389)	-0.57	NA
Practice	2.192 (1.188)	2.118 (1.253)	-0.167	NA
Well-thought-out	2.110 (1.100)	2.190 (0.880)	-1.118	NA
Organised	<b>1.870</b>	<b>2.250</b>	<b>-4.469</b>	<b>-0.18</b>
<b>Wording</b>	<b>(1.090)</b>	<b>(1.050)</b>		
<i>Appearance score</i>				

	<b>Voice</b>				
	<b>attractiveness</b>	<b>1.993</b>	<b>2.300</b>	<b>-4.381</b>	<b>-0.16</b>
		<b>(1.268)</b>	<b>(1.322)</b>		
<b>Meta-</b>					
<b>emotion</b>		<b>9.538</b>	<b>10.437</b>	<b>-3.949</b>	<b>-0.15</b>
<b>score</b>		<b>(3.568)</b>	<b>(3.123)</b>		

---

Note: Significant  $p$  values are in bold ( $p < 0.01$ ). Standard deviations appear in parentheses.

#### 8.9.5.1 Honest versus false appeals

As the present study utilised two true and two false appeals, judges may have detected different implicit cues for truthful and false appeals. In other words, judges could have picked up different cues for innocent and guilty appellants. To distinguish between judgments made for true appeals and comparing this with those made for false appeals and whether these differences are significant between both conditions, separate Mann-Whitney  $U$ s were conducted.

For true appeals, significant differences were found for perceived wording, with the Audio-only condition reporting a mean of 2.80 ( $SD = 1.21$ ) and the Audio-visual condition reporting a mean of 3.09 ( $SD = 1.06$ ). Significant differences were also found for perceived voice attractiveness, with the Audio-only condition reporting a mean of 2.15 ( $SD = 1.18$ ) and the Audio-visual condition reporting a mean of 1.90 ( $SD = 1.32$ ). There were no significant differences found for false appeals. Again, effect sizes were very small as indicated in Tables 8 and 9 in the  $r$  column.

**Table 8**

*Mann Whitney U item scores for false appeals between Audio-only and Audio-visual conditions*

	False appeals		<i>z</i>	<i>r</i>
	Audio-only means (SD)	Audio-visual means (SD)		
<i>Innocence score</i>				
Credibility	2.359 (1.257)	2.047 (1.338)	0.066	NA
<i>Cognition of Source content scores</i>				
Plausibility	2.295 (1.266)	2.496 (1.119)	-0.328	NA
Story Sense	2.089 (1.424)	2.322 (0.881)	-1.163	NA
Practice	1.925 (1.058)	1.906 (1.439)	0.214	NA
Organised	2.750 (1.258)	2.580 (1.130)	0.743	NA
Well-thought-out	2.034 (1.200)	2.189 (1.207)	0.078	NA
Wording	2.670 (1.025)	3.090 (1.116)	2.001	NA
<i>Appearance scores</i>				
Voice attractiveness	1.843 (1.332)	2.259 (1.304)	0.227	NA
<i>Meta-emotion score</i>	10.164 (3.046)	10.897 (2.925)	0.172	NA

Note: Significant *p* values are in bold; *p* < 0.01. Standard deviations appear in parentheses.

**Table 9**

*Mann Whitney U item scores for true appeals between Audio-only and Audio-visual condition*

	True appeals		<i>z</i>	<i>r</i>
	Audio-only means (SD)	Audio-visual means (SD)		
<i>Innocence score</i>				
Credibility	1.719 (1.077)	1.441 (1.191)	1.992	NA
<i>Cognition of Source content scores</i>				
Plausibility	1.741 (1.175)	1.667 (1.190)	-1.020	NA
Story Sense	1.879 (1.196)	1.794 (0.836)	0.299	NA
Practice	3.050 (1.102)	2.950 (1.230)	0.439	NA
Organised	2.930 (1.179)	2.879 (1.123)	-1.703	NA
Well-thought-out	3.260 (1.142)	3.190 (1.060)	-1.538	NA
<b>Wording</b>	<b>2.800 (1.212)</b>	<b>3.090 (1.060)</b>	<b>4.229</b>	<b>0.17</b>
<i>Appearance scores</i>				
<b>Voice attractiveness</b>	<b>2.150 (1.181)</b>	<b>1.902 (1.324)</b>	<b>-2.970</b>	<b>-0.11</b>
<i>Meta-emotion score</i>	8.885 (2.861)	7.886 (3.976)	2.230	NA

Note: Significant p values are in bold;  $p < 0.01$ . Standard deviations appear in parentheses.

### 8.9.6 Logistic regression analyses

Binary logistic regressions were conducted to assess which implicit cues judges may have used to predict their explicit veracity judgments within both conditions (rather than between conditions). The effect sizes between the differences for both conditions were very small in those tests but were useful in indicating where differences lie in judgments and whether this was for truthful or false appeals. Logistic regressions were carried out to investigate participant accuracy as dependent measures for both conditions. Again, predictor variables included in regression analyses were derived from an extensive review of the literature expounded in the opening chapters of the present thesis.

In this study, tolerance was greater than .10 for all items, and the variance inflation factor was also less than 10 for all items. This suggests that multicollinearity was not an issue in this study (meta-emotion, Tolerance = 0.81, VIF = 1.24; credibility, Tolerance = 0.78, VIF = 1.27; plausibility, Tolerance = 0.69, VIF = 1.46; practice, Tolerance = 0.85, VIF = 1.18; voice attractiveness, Tolerance = 0.91, VIF = 1.01; organised speech, Tolerance = 0.79, VIF = 1.42; less emotions, Tolerance = 0.75, VIF = 1.33; wording, Tolerance = 0.82, VIF = 1.22; well-thought-out, Tolerance = 0.91, VIF = 1.10; emotions match, Tolerance = 0.94, VIF = 1.10).

Although VIF values suggest that multicollinearity was not an issue in this study, precautions are also taken to reduce false positive findings. False positive findings are also synonymous with the fishing and error rate problem, or alpha inflation (Parker & Szymanski, 1992). This threat occurs when a high number of multiple statistical comparisons are being conducted and the possibility of making a Type I error increases with each comparison in attempt to test hypotheses and finding a significant effect (Parker & Szymanski, 1992). The result presents threats to statistical conclusion validity and drawing accurate conclusions about the relationship of the variables studied. To lower this risk of false positives and to increase confidence in results, a more stringent cut-off point will be employed for the regression analyses ( $p < 0.01$ ).

For prediction of explicit decisions in the Audio-only condition, the full model indicates that it was able to distinguish between respondents who explicitly

reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2$  (8,  $N = 108$ ) = 32.34,  $p < .001$ . The model as a whole explained between 18% (Cox and Snell  $R^2$ ) and 24% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 68% of cases. As shown in Table 10, only one independent variable made a unique statistically significant contribution to the model for this group. This predictor was how well-thought-out judges thought the appeal was, with the odds ratio ( $OR$ ) of 0.63.

Next, consideration as to which implicit veracity judgments measured in the present study contributed to the accuracy of their explicit judgment was also made. Another binary logistic regression with the actual accuracy of judges' explicit veracity judgments as the dependent measure in this condition was conducted, and the full model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2$  (8,  $N = 180$ ) = 16.49,  $p < .05$ . The model as a whole explained between 6% (Cox and Snell  $R^2$ ) and 8% (Nagelkerke  $R^2$ ) of the variance in accuracy, and correctly classified 69% of cases. As shown in Table 11, only one independent variable made a unique statistically significant contribution to the model for this condition. They were only accurate by using the implicit judgment of whether the overall appearance of the appellant was credible or not ( $OR = 0.68$ ).

For prediction of explicit judgements in the Audio-visual condition, the binary logistic regression model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2$  (8,  $N = 180$ ) = 54.14,  $p < .005$ . The model as a whole explained between 21% (Cox and Snell  $R^2$ ) and 28% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 73% of cases. As shown in Table 10, three independent variables made a unique statistically significant contribution to the model for this condition. These predictors were how well-thought-out they judged the appeal to be ( $OR = 1.76$ ), how attractive they judged the voice of the appellant was ( $OR = 1.75$ ), and whether they thought the appellant worded the appeal in a manner that is pleasant to hear ( $OR = 1.18$ ).

Consideration was then given to which implicit veracity judgments measured in the present study contributed to the accuracy of explicit judgment in

the Audio-visual condition. Another binary logistic regression with the actual 'accuracy' of judges' explicit veracity judgments as the dependent measure in this group was ran. The full model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2(8, N = 180) = 15.63, p < .05$ . The model as a whole explained between 7% (Cox and Snell  $R^2$ ) and 9% (Nagelkerke  $R^2$ ) of the variance in accuracy. There is no straightforward, commonly accepted measure for the fit of a regression model (Chatterjee & Hadi, 2015); the answer is governed by the aim of a particular study. If the main aim of the study was to determine which predictors are statistically significant and whether the changes in the predictors have a relationship with the change in the outcome variable, the  $R^2$  value is virtually completely immaterial (Frost, 2013). A low  $R^2$  value is only problematic when the study's aim is to produce precise predictions. Valuable inferences on the relationship between the changes in the predictor values being connected with the changes in response values can be made even with low  $R^2$  values, if statistically significant predictors exist (Frost, 2013). The rate of correct classification is an alternative method to assess the model's predictive accuracy. The current model correctly classified 63% of cases. A high proportion of correct classification indicates that the model works well (Chatterjee & Hadi, 2015). As shown in Table 11, only one independent variable made a unique statistically significant contribution to the model for this condition. This implicit variable was how organised the judges thought the appeal was ( $OR = .58$ ).

**Table 10**

*Bootstrapped binary logistics regression models for explicit veracity judgments between Audio-only and Audio-visual condition*

Variables	Model 1: Audio-only				Model 2: Audio-visual			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	.249 (.119)	1.282	1.015, 1.619	0.037	.243 (.131)	1.275	.986, 1.649	0.064
Plausibility	.250 (.113)	1.284	1.028, 1.560	0.027	-.131 (.138)	0.877	.670, 1.148	0.339
Story Sense	.083 (.110)	1.087	.876, 1.348	0.448	-.176 (.193)	0.838	.574, 1.224	0.361
Practice	-.132 (.127)	0.876	.682, 1.124	0.298	-.221 (.119)	0.802	.635, 1.012	0.063
Well-thought-out	<b>-.460 (.135)</b>	<b>0.631</b>	<b>.485, .822</b>	<b>0.001</b>	<b>.566 (.133)</b>	<b>1.761</b>	<b>1.356, 2.285</b>	<b>0.000</b>
Wording	.207 (.140)	1.229	.935, 1.617	0.139	<b>.167 (.170)</b>	<b>1.182</b>	<b>.847, 1.651</b>	<b>0.004</b>
Organised	-.096 (.140)	0.909	.690, 1.197	0.496	-.546 (.186)	0.580	.402, .835	0.965
Voice	-.091 (.118)	0.913	.725, 1.150	0.442	<b>.561 (.131)</b>	<b>1.752</b>	<b>1.357, 2.263</b>	<b>0.000</b>
Meta-emotion	.101 (.042)	1.106	1.019, 1.200	0.016	-.123 (.055)	0.884	.793, .986	0.012
<b>Model description</b>	Model is statistically significant: $\chi^2 = 55.29$ ; df = 9; Nagelkerke's $R^2 = .238$ ; Hosmer-Lemeshow test: $\chi^2 = 8.258$ ; df = 8; p = .409				Model is statistically significant: $\chi^2 = 54.143$ ; df = 8; Nagelkerke's $R^2 = .281$ ; Hosmer-Lemeshow test: $\chi^2 = 10.27$ ; df = 8; p = .247			

Note: Significant  $p$  values are in bold ( $p < 0.01$ ). Standard errors appear in parentheses.

**Table 11**

*Bootstrapped binary logistic regression models for accurate explicit veracity judgments between Audio-only and Audio-visual condition*

Variables	Model 1: Audio-only				Model 2: Audio-visual			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	<b>-0.383 (.121)</b>	<b>0.682</b>	<b>.538, 0.865</b>	<b>0.002</b>	-.103 (.049)	1.108	.769, 1.218	0.780
Plausibility	.146 (.118)	1.158	.919, 1.458	0.213	-.252 (.128)	0.780	.604, .999	0.049
Story Sense	.076 (.113)	1.079	.881, 1.370	0.504	.177 (.180)	1.197	.839, 1.698	0.326
Practice	-.068 (.128)	0.934	.864, 1.346	0.593	.000 (.104)	1.000	.815, 1.230	0.997
Well-thought-out	.128 (.129)	1.137	.883, 1.465	0.320	-.288 (.115)	0.750	.598, .940	0.013
Wording	.053 (.138)	1.055	.804, 1.384	0.699	.167 (.170)	1.188	.847, 1.651	0.326
Organised	.007 (.142)	1.007	.762, 1.330	0.963	<b>-.546 (.186)</b>	<b>0.580</b>	<b>.402, .835</b>	<b>0.003</b>
Voice	.108 (.121)	1.114	.879, 1.412	0.373	-.154 (.186)	0.863	.696, 1.060	0.148
Meta-emotion	-.033 (.042)	0.968	.891, 1.051	0.433	.103 (.049)	1.110	1.001, 1.220	0.036
<b>Model description</b>	Model is statistically significant: $\chi^2 = 16.49$ ; df = 8; Nagelkerke's $R^2 = .08$ ; Hosmer-Lemeshow test: $\chi^2 = 11.45$ ; df = 8; p = .178				Model is statistically significant: $\chi^2 = 15.63$ ; df = 8; Nagelkerke's $R^2 = .088$ ; Hosmer-Lemeshow test: $\chi^2 = 6.11$ ; df = 8; p = .635			

Note: Significant  $p$  values are in bold ( $p < 0.01$ ). Standard errors appear in parentheses.

### *8.9.7 Not-sure answers frequencies*

As described in the Methods section, not-sure answers for all implicit decisions were recorded for the remaining cases retained for analyses. Only percentages of 10% difference and higher between both groups. Only percentage differences between both conditions that are higher than 10% were reported and this able can be found in Appendix J. Audio-only condition found it easier to rate perceived overall appeal credibility compared to the Audio-visual condition.

## **8.10 Discussion**

In the present study, a distinction in explicit lie detection was found between both conditions: it seemed harder for participants to detect innocence and guilt when they could see the liar. The current study found better detection rates both explicitly and implicitly when the visual element of an appeal was removed. These results in this study would seem to replicate Bond and DePaulo's (2006) findings that truth-lie judgment accuracy declines when judging visible compared to audible lies, albeit in the current chapter this decline pertains to innocence-guilt judgments. Thinking patterns also appeared to be different for deceptive and truthful stimuli. Differences in processing veracity judgments seems to be influenced by the actual veracity of the source, showing the benefits of including a 50/50 ratio of innocent and guilty appealers.

Explicit veracity judgments were higher than chance level in the Audio-only condition when compared with the Audio-visual condition, suggesting that mode of presentation for these high-stake scenarios impacted decision processes. Looking at explicit innocence-guilt detection rates, judges who were in the Audio-visual condition reported a total rate of around 54.00%, whereas judges in the Audio-only condition reported a rating of nearly 70.00%.

There does not seem to be an innocence bias present for the Audio-only condition, with 72.60% rating for true appeals and 65.70% for false appeals, totalling to an overall of 69.20% rating. Overall, an innocence bias was not consistently found in the present study. However, the Audio-visual condition reported ratings that were similar to lie detection research, with ratings close to chance level. For false appeals the judges in this study achieved a rating of 54.90%

and for true appeals the percentage was 52.00% totalling to an overall rating of 53.30%. A breakdown of the overall results can be found in Table 5. In this condition, judges achieved higher percentages in detecting guilty appealers than innocent ones.

A difference was revealed in judges' explicit decision (with a rating of 72.60% in Audio-only compared to only 52.00% in Audio-visual). This either indicates that judges in both conditions were encoding some other implicit cues not captured in the questionnaire for only false appeals, or simply none of those cues were useful in detection deception for false appeals in this sample. Alternatively, this potentially highlights the distinction between explicit and implicit judgment processing. While this explanation requires further study, this draws attention to the occasion where judges cannot offer an implicit-type reasoning when offering an explicit decision to them (and vice versa). A case in point; Anderson et al. (1999) noted that participants in their study typically offered an implicit-type reasoning when they did not or could not explain with a significant explicit one.

In examining the regression results, judges in the Audio-visual condition detected cues such as voice attractiveness that judges in the Audio-only condition did not. The results in Table 10 would suggest that innocent appealers were judged to have higher voice attractiveness than guilty ones by those in the Audio-visual condition. They also seem to have used a higher number of cues compared to the Audio-only condition as shown in Table 10. Innocent appealers were judged to have presented a more well-thought-out and pleasant-sounding appeal by those in the Audio-visual condition. It could again imply that the questionnaire used in this study did not capture the actual implicit cues used by judges in the Audio-only condition. While still speculation at this stage, a possible explanation for the poorer performance in detecting innocence and guilt in the Audio-visual condition may perhaps be due to an increase in cognitive load when perceiving an added visual element. Judges in this condition could have conceivably experienced a wider expanse of information elements and/or sources to process, when the working memory can only process a limited amount of information pieces. The most common method of indirectly investigating cognitive load effects is analysing

performance outcome. In the current study, it can be said that the main performance outcome is judges' explicit veracity judgments.

One of the limitations in this study is that there was no direct measure of cognitive loading, so this hypothesis remains to be tested in future studies. A more direct method to objectively measure cognitive loading is the use of a dual-task analysis, which posits that when judges has to perform two separate tasks at the same time, and particularly if requires the same resources in verbal or visual working memory then the available resources would have to be distributed between the two tasks (Brünken, Plass & Leutner, 2003). The performance level in both tasks would be the outcome variable with the hypothesis being lower performance in the dual-task group compared to the single-task one. Yet another method is a self-reported stress level and/or self-reported difficulty of materials presented (Brünken et al., 2003). Another limitation of this study is that results could of course be idiosyncratic to these video clips, and the small number of videos used in the present study is noted. These results would be useful to be confirmed by further studies with a higher set of trials and different video clips to increase the study's generalisability.

In the between-group analyses of judges' implicit veracity judgments, their meta-emotion levels showed a significant difference between both conditions shown by Mann Whitney U tests, though with very small effect sizes. Judges in the Audio-only condition felt significantly less sympathetic towards the appellars in comparison to those in the Audio-visual condition. As can be seen in their explicit judgments, judges were much more accurate in judging an appellant's veracity both true and false (over chance level) in the Audio-only condition as compared with their counterparts who hovered around chance level. This conveys over to the next point of discussion which is how they felt about the appellant was effectively impacted by what they could see. In a sense, they were very much less persuaded by the implicit judgment of appearance when they were not able to see it. Judges adapted and relied on verbal cues, as suggested by Anderson et al. (1999) when visual cues were not available to them.

Regression results would suggest that presentation format influenced how judges thought, highlighting biases of emotional display when a visual element

was present. Chi square results would also infer this, although the strength of this relationship was weak. One of the reasons the strength of this relationship was weak is perhaps due to the length of the clips itself. The appeals presented to judges may have been too short; if judges were exposed to a longer stimulus material perhaps the strength of this relationship may have been larger and more prominent. Additionally, the length of appeal clips used in the present study varied considerably (from 34.21 seconds to 160.09 seconds). It is possible that this factor could have influenced making innocent-guilty judgments, with the content of longer appeals potentially having more influence over judgements and shorter ones having less impact (Masip et al., 2009). A fuller discussion as to the reasons of why this relationship was weak can be found in Sections 11.6 and 11.7 in Chapter 11. Although, it must be acknowledged that recent research by Street and Masip (2015) would argue that this is not explained by a time effect but rather due to the act of making within-statement comparisons which decreases truth bias.

In summary, differences in thinking were found for both experimental and the present study provided some insight into whether the innocence bias hypothesis stands in this type of high-stakes situations. Tentatively, results would tend to suggest that the absence of visual input aids the detection of innocence-guilt.

## CHAPTER 9: STUDY 3

---

### PERCEPTION OF STORY SENSE IN TELEVISION APPEALS

A worldwide study aiming to uncover stereotypes of liars conducted by the Global Deception Research Team found that among the top five invalid cues that individuals often used to judge deception was 'speaker incoherence' (2006, Study 1). Over 25% of judges from 58 countries around the world held the belief that 'deceptive statements are incoherent'. As can be seen from this percentage, a very salient and easily accessible cue available for judges to pick up from a television appeal is the verbal aspect of what the appellant says, and whether their story makes logical sense.

In the present study, attention is now paid to the Source content – verbal cues in particular – as Studies 1 and 2 addressed visual ones (appearance and emotional displays). Very few lie detection studies have employed manipulation of verbal content in a statement. Reinhard and Sporer (2010), one of the few known studies to have exercised manipulation of verbal content, manipulated the consistency and plausibility of the transcripts in their study. Implicit judgments of appeal believability may be influenced by the subjective impressions of the sequencing of a story (which alters the story sense). Canter et al.'s (2003) work on narrative plausibility operated on the assumption that a narrative schema was the baseline for how 'true' a story was and provided a foundation for comparison. Anything that diverged from this baseline was seen to have a lower plausibility. They maintained that for observers to make either implicit or explicit judgments there are internal and external factors or attributions that lent to these judgments. The internal factor in this study was the appeal structure, which was essentially how the appeal was organised and develops. The external factors included possible biases and subjective perceptions of these appeal structures and appeal sense.

Participants in Anderson et al.'s (1999) study, when asked to assess their friends' veracities, generated verbal cues such as story complexity (e.g. 'The story was so vague, I thought it was a lie') and story flow (e.g. 'They seemed unsure of their own story; it didn't flow') and a general category the authors labelled story content or structure of the story (e.g. "It was a good story; it all made sense"). Similarly, the current study enquires into definitions and pre-existing research on

perceptions of story sense and similar aspects such as story coherence, sense and story preparation.

### **9.1 Definitions of coherence**

Habermas and Bluck (2000) defined temporal coherence as activities that make orderly logic within an occurrence. They described ‘causal coherence’ as the ability to connect separate activities into causal sequences. ‘Thematic coherence’ was defined as the ability to extract an overarching theme or general message from a series of narrated events. Trabasso, Secco and van den Broek (1984) theorised that the coherence of a story is dependent on the causal cohesion, logically and causally, of the story’s individual events. The more cohesive the individual events are, the easier it will be to construct a coherent story. To understand a story is an effort to determine the causes that propelled the event, and the consequences of the event. The listener learns of the roots and outcomes of the event, directing them to an experience of the event in an orderly fashion as opposed to incoherent series of events. This association of sequences is where the authors claim a listener begins to deduce and generate responses.

Labov (1972) suggested that a narrative that is temporally ordered lends to easy comprehension and remembering. Socially speaking, the author contended that a story must contain sufficient coherence to ensure communication of the story message to the listener. Stories that do not have a clear temporal structure, cause and effect can cause a listener to judge it as incoherent. McAdams (2006) contended that an account that abruptly starts at the supposed ending, to the beginning, then juxtaposing again forward temporally, can be challenging for a listener to track. The anticipation for a complete story is for it to comprise of a clear start, central event and finish to the observer. The author wrote that stories that illustrate characters without a clear motive, vague plot, no clear cause or consequence, or closure can be judged as incoherent as well.

### **9.2 How does Coherence associate with Judgements of Plausibility and Story Sense**

In general, stories that are coherent are more likely to be received in a more positive light. Dictionaries offer definitions of ‘incoherency’ as a story that is incomprehensible and unintelligible, lacking logical and internal consistency (“Incoherent,” 2016). As expounded in Chapter 6, Lam (2001) described that in an attempt for an observer to conduct a probability assessment of whether someone is guilty or innocent, they first try to rule out stories or events that are incoherent and do not make sense. Having established the stories that make sense or are coherent, they then would assign a degree of plausibility to the stories. Inclusively all three variables evaluate implicit components of the general outcome of perceived innocence and guilt of the appellant.

In parallel with Lam (2001)’s findings, Pennington and Hastie (1988) found in a mock trial scenario that when evidence was presented in a story order making this story easier to construct, nearly four out of five participants returned a guilty verdict for the defendant in question. When this evidence was presented in witness order which makes it less easy for the story to be constructed, only one in three participants returned a judgment of guilt. This revealed that the structure and order of stories presented influences decision making. The story model is discussed at length in Section 2.3.2 of the current thesis.

In addition, plausibility closely relates to the distribution of pattern in how words and even sentences co-occur describing a story. Lapata, McDonald and Keller (1999) found that a certain association between words can make precise sentences seem more plausible due to the advantages of using those particular words. In other words, plausibility is associated with word and sentence coherence; altering the conceptual coherence of a story lowers its plausibility. However, while stories that make more sense are more likely to be judged as plausible, a speaker may still be judged as credible while their story be seen as implausible, make less sense and is incoherent.

Recently, Whelan et al. (2014) found that untruthful appellants more often told a story in an illogical or unclear sense than their counterparts. Similarly, DePaulo et al. (2003) discovered that the stories that liars tell make less sense and is less plausible. For example, liars tend to make inconsistent and ambiguous

statements. These literatures aside, the next section reviews the subjective perceptions of how much sense liars and truth tellers are *assumed* to portray.

### **9.3 Judges' Cognitive perceptions of Story Content**

In discussing how accurate subjective perceptions of liars and truth tellers are in terms of appraising story sense, results from existing studies on observer assessments and accuracies were mixed. It should be kept in mind that no studies have directly compared incoherent statements to coherent ones in assessing perceived veracity judgments, however there are studies comparing the level of story preparation. Vrij (2000) predicted that unprepared lies will be easier to detect compared to prepared ones, because an unprepared liar has not thought through the content and details of their story and may contradict themselves, making them easier to be identified. Strömwall, Granhag and Landstrom (2007) found that the participants in their study gained higher accuracy for unrehearsed stories (56.6%) compared to stories that were prepared (46.1%). Their senders were children aged eleven to thirteen and adult participant observers. Bond and DePaulo (2006) who examined a meta-analysis of the subjective impressions and evaluations of story preparation concluded that prepared stories were seen as more honest than those that were unprepared when the same person has given both types of statements. However, they also found inconsistent evidence when comparing between-subjects studies, where they found that prepared stories were less believable than unprepared ones. They further found that accuracy was higher for spontaneous stories compared to prepared ones, but only with small effect sizes. In other words, they found it was easier to tell the difference accurately whether someone is lying or not when the story is unprepared rather than prepared. This was attributed in part to the low stakes nature of most of the literature reviewed in their meta-analysis. Support for differences in perceived verbal cues can be found in Granhag and Strömwall's (2000) study where the authors found that disagreement within their sample as to whether certain consecutive statements were consistent or not.

An internally coherent story may not always be perceived as coherent by an observer and vice versa. Hasson, Simmons and Todorov (2005) proposed that

understanding story content does not require believing it. Overall believability can be influenced by other factors apart from internal story structure, for example, as well. McAdams (2006) wrote that to an atheist, the second coming of Jesus and the story of the resurrection is implausible and delusional, yet to many Christians this story is perfectly plausible. Subjective impression in appraising the plausibility and coherence of a story are subjected to norms and past experiences.

The concept coherence view posits that when an observer makes a plausibility assessment, they tend to relate the sender's story to their own past experience, make inferences from the story, and assess whether this is a good match to what they have experienced in the past directly or remotely (Lapata et al., 1999). Considering how an observer processes language is useful in understanding perceptions and the subsequent biases that may arise when encoding speech. Observers approach sentences in English with at least four basic assumptions (Bever, 1970) and subsequently devise strategies for dealing with what they hear. One of these basic assumptions is that people utter sentences that make sense. The listener would then use their knowledge of the world to pick the most likely interpretation of the sentence they hear (Aitchison, 2008). Burgoon, Jones and Stewart (1975) proposed the language expectancy theory, where the authors claimed that what a speaker chooses to say can significantly predict whether an observer is successfully convinced. This theory contends that people hold expectations as to what they consider normal language in a given situation, grounded in psychological and cultural norms of their immediate society, and bring these stereotypes into social situations. This means that a sender's words may present inconsistencies with the stereotypes observers carry. Due to this violation of expectancies the observer carries, a sender may be seen to be less liked or believed, as their language and/or behaviour is inherently less persuasive.

Similarly, the information processing theory (McGuire, 1968) contended that to be persuaded an observer must first attend to and understand a persuasive message. If an observer cannot attend to and understand the message, they cannot then apply their understanding and knowledge to correctly accept or reject the message. Interrupted structures in sentences or phrases offer another difficulty in comprehension.

The consensus in the literature suggests that a story that depicts events or happenings that defy an observer's understanding of how the world works and how human beings typically act, think, feel, and want, may seem to not make sense, in that it violates structural norms. The question which then arises in the context of television appeals is whether what grievors say can be excused, or accepted favourably, if it does not make sense. Or, alternatively, will observers be inclined to perceive grieving appealers as liars if their appeal is incoherent and interrupted? The language of grief is governed by emotional, psychological and social changes. As such, observers could expect, to a certain degree, what a griever should and would say, and little research exists on the subjective perceptions of language in a high stakes deception context.

### *9.3.1 Rehearsal and Structure*

The rehearsal of a story can lead to subjective impressions of how credible and believable a story is as well. The notion is that the more prepared a story is, the more coherent it may sound as it will contain more structure. Lie detection research has consistently found that liars always prepare their narratives ahead of time before an interview and studies report that a difference in accuracy is shown by observers for unprepared statements than prepared statements (Granhag, Andersson, Strömwall, & Hartwig, 2004; Vrij et al., 2009). When statements are unprepared they can appear incoherent, or at least less coherent. On the other hand, Vrij, Leal, Granhag, Mann, Fisher, Hillman and Sperry (2009) wrote that preparation can lead to responses that are excessively scripted.

One of the criteria in Content Based Criteria Analysis (as introduced in Chapter 3) for assessing statement credibility is unstructured reproduction (Vrij, 2008). Truthful accounts tend to have less of a prepared structure, contain less chronological content compared to untruthful stories, which are more likely to be overly scripted and chronological (e.g., "I did X, then Y happened and Z took place"). This is due to the liar memorising details of the account in chronological order to keep their story straight (Porter & ten Brinke, 2010). The more they fabricate, the more they must remember, and liars have difficulties recounting a story that is not in chronological order. Vrij (2000) offered that liars may find

difficulty in including certain criteria found in CBCA in their stories. For example, they may face difficulty in producing an unstructured account, which is typically a distinctive feature of a truthful account. Liars are also generally under the impression that if they admit a lack of memory or raise doubts about their own narratives, they will be seen as less truthful, affecting their credibility. Adams (2002) purported that when liars are experiencing stress, deviations from the practiced story structure can emerge.

The discrepancy can be seen here between how liars normally produce and relay a fabricated story, which are often chronologically and conceptually coherent due to their scripted nature, versus how an observer would encode and judge the coherency of a speech, where stories with a clear temporal structure are favourable and tend to be judged positively. However, a connection to the coherence of the story as well as observer judgments of this cue has yet to be explored in a high stakes deception context.

#### **9.4 Sender detectability**

Besides individual differences in appraising coherence, there are also individual differences in the ability in telling coherent stories. McAdams (2006) wrote that individuals vary in terms of their skill in narrating a story with thematic and causal coherence. The author further maintained that the coherence of a story depends upon the storyteller's ability to refabricate, imagine and integrate. The storyteller's memory of the event and the way they choose to tell of the event is notable, as stories that are coherent should have credibility to it too. To eliminate this issue, the same appealers will be shown in both high and low sense conditions in the present study.

#### **9.5 Other implicit cues**

Again, implicit cues such as perceived attractiveness and emotions felt by participants towards appealers expounded in previous chapters are studied here.

#### **9.6 Key issues to explore**

The key issue the present study sets to explore is the effect of reordering sentences and words of a coherent story, so that it quickly becomes nonsensical, thus potentially reducing its plausibility and intelligibility. A high-stakes sensible story is used but its conceptual coherence altered, with the hypotheses that Judges exposed to the High Sense condition will achieve higher explicit ratings, higher judgments of story plausibility, overall credibility, story sense and emotions felt than their counterparts.

## **9.7 Methods**

### *9.7.1 Participants*

The preliminary sample of participants totalled to 181 participants for both groups, with 724 cases, as each participant judged four cases of these appeals. None of the participants reported 'Yes' to question "Do you personally know anyone who featured in the appeal you've just heard?" and no one reported 'Yes' to the question "Do you know of the appeal you have just heard i.e. you've watched or heard it on TV?". Therefore, none of the participants had any prior knowledge of the appellants they listened to but some knew of the cases from watching or hearing it on television. Ten participants were removed from this initial sample because they failed to record explicit veracity judgments. Participants were recruited using a mixture of snowball and opportunity sampling from the UK and Europe. They were from varying occupations ranging from bus driver and dermatologist to advertising managers and civil servants; as well as nationalities/ethnicities (British, Asian, African, Middle Eastern and European).

### *9.7.2 Material and Procedure*

Granhag et al. (2003) wrote that liars' stories are usually highly consistent due to the contention that liars often prepare a story beforehand. If they do not prepare the contents of the fabricated story beforehand they may run the risk of being inconsistent especially in a police interview situation. In contrast, Gilovich, Savitsky and Medvec (1998) maintained that truth-tellers on the other hand prepare their stories less for they hold the view that the truth will be transparent without having to practice their story. Because television appeals are mostly

prepared beforehand, this element of rehearsal is probably similar across all appeals. In the current study, this internal element was edited and manipulated to observe what participants' subjective impressions of story 'rehearsal' were and how this impacts their veracity assessments.

Using a between-subjects experimental design, two sets of press conferences will be used in this study. In the first set, judges will be exposed to press conferences and interviews in their original format. In the second, the same press conferences and interviews will be viewed by a different group of judges, however in this condition the appeals were presented with reordered sentences. Sentences will not be truncated, but rather restructured to make less sense via Windows Movie Maker editing software. The effect of this video editing will be that the appellants appear to speak normally but their sentences will not be as coherent and make less sense in comparison to the unaltered version. In other words, all judges will experience the exact same audio content of the appeals; except one group will watch the appeals in a High Sense manner and the other will watch it in a Low Sense manner.

Appeals that are longer in content will be selected (included were two police interviews out of the four appeals) so as to ensure sufficient story content. An example of a High Sense order of interview that will be used in this study is 'There is a lot of evidence that places me at (the victim)'s house, and there is a reason for that, and the reason is that I did see (the victim) the day they passed away. A lot of things happened that day, I almost lost my life. There was an argument amongst some people, two individuals that one wanted to take my life. The guy with the gun was standing near me and I just remember holding my head, closing my eyes. He pulled the trigger, and nothing happened, just a click.' The same interview will be rearranged and as a result the Low Sense version of this same interview study will be 'That I did see (the victim) the day they passed away. There is a lot of evidence, and the guy with the gun was standing near me, and I just remember holding my head and closing my eyes, and there is a reason for that, I almost lost my life. There was an argument. A lot of things happened that day, amongst some people, two individuals that places me at the victim's house, that one wanted to take my life, and that reason is that he pulled the trigger and nothing happened,

just a click.’ Notice that while the appeal here has the same content as the unaltered version, it does not flow with ease, grammatical errors are present and generally the story makes less logical sense.

Two genuine appeals and two false appeals will be chosen (these will be the same appeals for each set). The first video depicts a female appealer who is guilty, the second shows a guilty male appealer, the third shows an innocent female appealer, and the fourth shows an innocent male appealer. The videos range from 34.13 seconds to 160.00 seconds in length, with an average of 95.12 seconds each. The general methodology of how the experiment will be conducted is outlined under Section 6.3 in Chapter 6.

To test if there will be a statistically significant relationship between the accuracy of judges’ explicit decisions and level of speech coherence they were exposed to, a Chi-square will be computed. Logistic regressions for within-group analyses will also be carried out to determine which implicit responses judges used could predict their accurate explicit decisions. This analysis will be used to explore the relationship between explicit and implicit decisions. As with the previous chapter, additional regressions were also conducted. Variance inflation factors (VIF) will also be analysed to check for multicollinearity (Dart, 2017). All bootstrapped models will contain the same twelve independent variables (story plausibility, story sense, credibility, practice, well-thought-out, wording, organised, lessemotions, emotionsmatched, face attractiveness, voice attractiveness and meta-emotion level). As in the previous chapters, the bootstrapping process is applied. It is very useful for small sample sizes (Adèr, Mellenbergh & Hand, 2008) and works by making extrapolations of a population from a sample data. This process samples this data and then makes extrapolations about a sample from re-sampled data (Efron & Tibshirani, 1993). In the present study, this process is repeated 1000 times (Efron, 1979; Miles et al., 1999).

### *9.7.3 Ground truth*

Appeal 1 shows a guilty female giving her accounts during an interview about her murdered boyfriend. Appeal 2 shows a guilty male giving his accounts during an interview about his missing granddaughter. Appeal 3 shows an innocent

female both directly appealing to the public in the case of their missing daughter. Appeal 4 shows an innocent male giving an account of what happened the day his wife went missing and directly appeals to the public for information. In all four cases, substantial evidence was found to either discredit false appellant's claims or reinforce true appellants' claims concerning the fate of their relatives. For False Appellant 1, various pieces of evidence were presented in court including a gun, rented car, phone, friends, camera and palm print DNA. For False Appellant 2, a tip from the public and the found body were the main evidence. For True Appellant 3, again various pieces of evidence were presented including a body that was found, DNA evidence (blood stains on clothing and vehicle) and forensic entomology reports. For True Appellant 4, his alibi stands and another killer was convicted of raping and murdering his wife.

## **9.8 Results**

### *9.8.1 Participants*

Because of individual differences in cue appraisal, the question item 'The appellant tells a story that makes sense' was analysed; only participants' answers that corresponded with the author's definition of sense level were selected for further analysis. This step was necessary to be taken to make sure participants' opinions match the author's definitions, and that the independent variables are truly measuring what they were intended to. This was to verify that judges perceived what the author construed as a highly sensible appeal to be actually so, and similarly, that an appeal that has low sense was also perceived as such. Otherwise, it will be difficult to offer meaningful interpretation of results. Furthermore, those who reported 'not-sure' to these questions were eliminated.

The final pool of participants ranged from 17 to 69 years in age ( $M = 32.11$ ,  $SD = 12.29$ ). 44.6% were male and 55.4% were female, with 219 cases for the High Sense condition and 89 cases for the Low Sense condition totalling to 308 cases.

### *9.8.2 Explicit veracity judgments*

As presented in Table 12, judges in the High Sense condition recorded an overall rating of almost 62.00% overall, higher than the Low Sense condition at 49.70%, which was around chance level.

### *9.8.3 Relationship between level of story sense and explicit ratings*

Chi-square results were  $\chi^2 (2) = 6.551, p < .05$ . This would seem to suggest revealed a statistically significant association between ratings of judges' explicit decisions and perceived level of speech coherence. The research hypothesis that differences in accurate explicit veracity decisions were related to different levels of perceived coherence in appellant's speech was supported by this analysis. Phi values indicate the strength of this relationship and chosen over Cramer's V as both variables used in the sample are dichotomous (Jones, 2009). The strength of association between the variables was small (.156), demonstrating the presence of a weak relationship (Real Statistics Using Excel, 2017).

### *9.8.4 Innocence bias*

Table 12 distinguishes guilt ratings from innocence ratings. Results do not seem to suggest the presence of a consistent innocence bias. While seemingly present in the High Sense condition, it was absent in the Low Sense condition. Ratings for detecting innocence in the Low Sense condition was unusually low at 22.70%, whereas guilt detection for this condition was much higher in comparison at 76.85%. Of course, caution must be implemented when interpreting these results based on low number of video clips shown to judges.

## **Table 12**

*Judgment rates for innocent and guilty appellants for High and Low Sense condition*

<b>Appeals</b>	<b>High Sense</b>	<b>Low Sense</b>
<b>False</b>	<b>50.70%</b>	<b>76.85%</b>
<b>True</b>	<b>72.85%</b>	<b>22.70%</b>
<b>Total</b>	<b>61.78%</b>	<b>49.78%</b>

**Table 13***Mann Whitney U item scores between High and Low Sense conditions*

	High Sense means (SD)	Low Sense means (SD)	<i>z</i>	<i>r</i>
<i>Innocence score</i>				
<b>Credibility</b>	<b>2.48 (1.27)</b>	<b>1.39 (0.98)</b>	<b>-2.036</b>	<b>-0.12</b>
<i>Cognition of Source content scores</i>				
<b>Plausibility</b>	<b>2.79 (2.84)</b>	<b>1.58 (1.49)</b>	<b>-10.513</b>	<b>-0.50</b>
<b>Story Sense</b>	<b>3.28 (0.45)</b>	<b>1.58 (0.49)</b>	<b>-10.880</b>	<b>-0.62</b>
Practice	2.17 (1.17)	1.90 (1.33)	-2.174	NA
<b>Well-thought-out</b>	<b>2.45 (1.21)</b>	<b>2.35 (1.18)</b>	<b>-1.272</b>	<b>-0.07</b>
<b>Organised</b>	<b>2.59 (1.23)</b>	<b>1.11 (1.44)</b>	<b>-5.256</b>	<b>-0.30</b>
<b>Wording</b>	<b>2.17 (1.29)</b>	<b>1.59 (1.19)</b>	<b>-3.047</b>	<b>-0.17</b>
<i>Emotion scores</i>				
Less Emotions	1.95 (1.07)	2.59 (1.33)	-2.501	NA
<b>Emotions Match</b>	<b>2.32 (1.25)</b>	<b>1.79 (1.21)</b>	<b>-6.327</b>	<b>-0.36</b>
<i>Appearance scores</i>				
<b>Voice attractiveness</b>	<b>2.12 (1.34)</b>	<b>1.62 (1.27)</b>	<b>-3.141</b>	<b>-0.18</b>
Face attractiveness	1.72 (1.42)	1.59 (1.19)	-0.350	NA
<b>Meta-emotion score</b>	<b>15.06 (4.88)</b>	<b>10.27 (3.44)</b>	<b>-6.615</b>	<b>-0.38</b>

Note: Significant *p* values are in bold; *p* < 0.01. Standard deviations appear in parentheses.

### 9.8.5 Implicit veracity judgments

Results for the Kolmogorov-Smirnov test for normality (Massey, 1951; Smirnov, 1948) indicated that the score distributions deviated significantly from a normal distribution ( $D = .193$ ,  $p < .005$ ). Due to the between-subjects design of the study and the normality not being met, the most appropriate analyses to run was the Mann Whitney U test (Field, 2000a). All results can be found in Table 13. Cronbach's alpha for each scale items can be found in Table 14.

**Table 14**

*Cronbach's alpha coefficients for Study 3*

<b>Meta-emotion</b>	<b>Cognition</b>	<b>Appearance</b>	<b>Emotions</b>
0.74	0.62	0.49	0.35

#### 9.8.5.1 Honest versus false appeals

As the present study utilised two true and two false appeals, judges may have utilised different implicit cues for truthful and false appeals. To distinguish between judgments made for true appeals and comparing this with those made for false appeals and whether this difference was significant between both conditions, further separate Mann-Whitney Us were conducted.

For truthful appeals, as seen in Table 15, items which were significantly different for both group levels are the innocence score, appeal wording, voice attractiveness, story plausibility, story sense, overall credibility, emotions-match and meta-emotion. For false appeals, as seen in Table 16 significant differences were found for the items of whether the appeal sounded organised or not, overall innocence score, matched-ness of emotional display and speech content, story plausibility, story sense, level of emotional display and level of meta-emotion. Again, effect sizes were mostly small. Implications of results are discussed below.

**Table 15***Mann Whitney U item scores for true appeals between both conditions*

	True appeals		<i>U</i>	<i>p</i>	<i>r</i>
	High Sense means (SD)	Low Sense means (SD)			
<i>Innocence score</i>					
Credibility	2.720 (1.187)	2.154 (1.214)	682.5	0.056	NA
<i>Cognition of Source content scores</i>					
<b>Plausibility</b>	<b>2.940 (0.985)</b>	<b>1.333 (0.485)</b>	<b>231.0</b>	<b>0.000</b>	<b>-0.38</b>
<b>Story Sense</b>	<b>3.340 (0.475)</b>	<b>1.556 (1.096)</b>	<b>174.0</b>	<b>0.000</b>	<b>-0.45</b>
Practice	2.208 (1.198)	2.000 (1.414)	893.0	0.629	NA
Well-thought-out	2.600 (1.209)	2.692 (1.032)	964.5	0.947	NA
Organised	2.820 (1.153)	1.923 (1.605)	682.5	0.059	NA
Wording	2.267 (1.294)	1.692 (1.032)	964.5	0.947	NA
<i>Emotion scores</i>					
Less Emotions	1.812 (1.049)	1.615 (1.387)	840.5	0.409	NA
<b>Emotions match</b>	<b>2.363 (1.414)</b>	<b>4.944 (1.349)</b>	<b>192.0</b>	<b>0.000</b>	<b>-0.33</b>
<i>Appearance scores</i>					
<b>Voice attractiveness</b>	<b>2.127 (1.362)</b>	<b>0.889 (1.023)</b>	<b>653.5</b>	<b>0.000</b>	<b>-0.22</b>
Face attractiveness	1.440 (1.556)	1.278 (1.074)	1265.5	0.644	NA
<b>Meta-emotion score</b>	<b>16.073 (4.629)</b>	<b>9.769 (3.854)</b>	<b>287.0</b>	<b>0.000</b>	<b>-0.24</b>

Note: Significant *p* values are in bold; *p* < 0.01. Standard deviations appear in parentheses.

**Table 16***Mann Whitney U item scores for false appeals between both conditions*

	False appeals		<i>U</i>	<i>p</i>	<i>r</i>
	High Sense means (SD)	Low Sense means (SD)			
<i>Innocence score</i>					
Credibility	2.154 (1.214)	2.094 (1.329)	1042.5	0.645	NA
<i>Cognition of Source content scores</i>					
<b>Plausibility</b>	<b>2.493 (1.346)</b>	<b>1.281 (0.457)</b>	<b>468.0</b>	<b>0.000</b>	<b>-0.22</b>
<b>Story Sense</b>	<b>3.145 (0.355)</b>	<b>1.375 (1.129)</b>	<b>177.0</b>	<b>0.000</b>	<b>-0.21</b>
Practice	2.087 (1.121)	1.594 (1.388)	851.5	0.057	NA
Well-thought-out	2.116 (1.145)	2.000 (1.270)	1044.0	0.651	NA
<b>Organised</b>	<b>2.088 (1.267)</b>	<b>1.625 (0.793)</b>	<b>676.0</b>	<b>0.002</b>	<b>-0.18</b>
Wording	1.959 (1.254)	1.625 (1.185)	903.0	0.129	NA
<i>Emotion scores</i>					
<b>Less Emotions</b>	<b>2.232 (1.073)</b>	<b>1.469 (1.218)</b>	<b>703.0</b>	<b>0.003</b>	<b>-0.17</b>
<b>Emotions match</b>	<b>2.229 (1.216)</b>	<b>3.781 (1.879)</b>	<b>519.5</b>	<b>0.000</b>	<b>-0.22</b>
<i>Appearance scores</i>					
Voice attractiveness	2.116 (1.323)	1.875 (1.289)	975.0	0.334	NA
Face attractiveness	2.333 (1.159)	1.774 (1.185)	872.5	0.081	NA
<b>Meta-emotion score</b>	<b>12.838 (4.711)</b>	<b>9.688 (3.649)</b>	<b>600.5</b>	<b>0.000</b>	<b>-0.21</b>

Note: Significant *p* values are in bold; *p* < 0.01. Standard deviations appear in parentheses.

### 9.8.6 Logistic regressions for within-group analyses

Mann Whitney U tests were useful in indicating where differences lie in judgments and whether this was for truthful or false appeals. While differences were not found for certain implicit cues, judges in both conditions may still have used those cues to come to their explicit judgments. Again, predictor variables included in regression analyses were derived from an extensive review of the literature expounded in the opening chapters of the present thesis. In this study, tolerance was greater than .10 for all items, and the variance inflation factor was also less than 10 for all items. This suggested that multicollinearity was not an issue in this study (meta-emotion, Tolerance = 0.67, VIF = 1.75; credibility, Tolerance = 0.75, VIF = 1.34; plausibility, Tolerance = .63, VIF = 1.20; practice, Tolerance = 0.86, VIF = 1.16; voice attractiveness, Tolerance = 0.73; VIF = 1.37; organised speech, Tolerance = 0.66; VIF = 1.53; face attractiveness, Tolerance = 0.83; VIF = 1.20; less emotions, Tolerance = 0.81; VIF = 1.24; wording, Tolerance = 0.80; VIF = 1.25; well-thought-out, Tolerance = 0.72, VIF = 1.40; emotions match, Tolerance = 0.92, VIF = 1.10; practice, Tolerance = 0.86, VIF = 1.16). Again, here a cut-off point of  $p < 0.01$  was used for implicit responses.

First, implicit veracity judgments used by the High Sense condition judges which predicted their explicit veracity judgment, with the full model indicating that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2 (12, N = 308) = 70.02, p < .005$ . The model as a whole explained between 33% (Cox and Snell  $R^2$ ) and 47% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 80% of cases. As shown in Table 17, two independent variables made a unique statistically significant contribution to the model for this condition. One of these predictors was the implicit cue of the appellant's overall credibility, recording an odds ratio (*OR*) of 2.49 which indicated that respondents who reported appellants were overall credible were 2.49 times more likely to give a judgment of Innocence to the appellants, controlling for all other factors in the model. The other predictor was how well-practised the appeal was perceived by judges (*OR* = .48).

Next, consideration as to which implicit veracity judgments measured in the present study contributed to the accuracy of their explicit judgment was also

made. Another binary logistic regression with the actual accuracy of judges' explicit veracity judgments as the dependent measure in this condition was conducted, with the full model indicating that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2(12, N = 308) = 23.73, p < .05$ . The model as a whole explained between 12% (Cox and Snell  $R^2$ ) and 17% (Nagelkerke  $R^2$ ) of the variance in accuracy, and correctly classified 70% of cases. As shown in Table 18, none of the independent variables made a unique statistically significant contribution to the model for this condition.

Implicit veracity judgments used by the Low Sense condition judges which predicted their explicit veracity judgment with the full model of the binary log regression indicating that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2(8, N = 308) = 24.32, p < 0.05$ . The model as a whole explained between 42% (Cox and Snell  $R^2$ ) and 64% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 84% of cases. As shown in Table 17, none of the independent variables made a unique statistically significant contribution to the model for this condition.

Consideration was then given as to which implicit veracity judgments measured in the present study contributed to the accuracy of their explicit judgment in the Low Sense condition. A final binary logistic regression with the actual accuracy of judges' explicit veracity judgments as the dependent measure in this group was ran, with the full model indicating that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2(8, N = 308) = 15.44, p < .05$ . The model as a whole explained between 30% (Cox and Snell  $R^2$ ) and 40% (Nagelkerke  $R^2$ ) of the variance in accuracy, and correctly classified 75.6% of cases. As shown in Table 18, no independent variable made a unique and statistically significant contribution to this model in this condition.

**Table 17**

*Bootstrapped binary logistic regression models for explicit judgments in High and Low Sense conditions*

Variables	Model 1: High Sense				Model 2: Low Sense			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	<b>.914 (.217)</b>	<b>2.495</b>	<b>1.632, 3.816</b>	<b>0.000</b>	-.025 (.726)	0.980	.235, 1.050	0.973
Plausibility	.427 (.202)	1.532	1.032, 2.275	0.034	4.388 (1.264)	2.450	.689, 2.980	0.071
Story Sense	.771 (.642)	2.162	1.614, 2.610	0.230	.291 (1.189)	1.338	.130, 3.730	0.806
Practice	<b>-.717 (.246)</b>	<b>1.532</b>	<b>.301, 1.791</b>	<b>0.004</b>	-.434 (.924)	0.648	.106, 3.961	0.639
Wording	-.271 (.198)	2.612	1.517, 2.924	0.171	-1.404 (1.022)	0.246	.033, 1.819	0.169
Organised	.215 (3.790)	1.240	.813, 1.891	0.318	-.189 (.788)	0.828	.177, 3.882	0.811
Well-thought-out	-.112 (.263)	0.894	.534, 1.498	0.671	-1.028 (.788)	0.358	.076, 1.678	0.192
Voice	.005 (.201)	1.005	.678, 1.489	0.980	-.114 (.729)	0.892	.214, 3.722	0.875
Face	.181 (.176)	1.198	.849, 1.692	0.304	.499 (.690)	1.647	.426, 4.373	0.470
Emotions Match	.187 (.179)	1.206	.739, 1.373	0.296	-.494 (.828)	0.610	.120, 3.094	0.551
Less Emotions	-.432 (.235)	0.649	.410, 1.028	0.065	.381 (.701)	1.464	.371, 4.782	0.586
Meta-emotion	-.035 (.065)	0.965	.849, 1.097	0.588	.422, (.318)	1.525	.818, 4.843	0.185
<b>Model description</b>	Model is statistically significant: $\chi^2 = 70.02$ ; $df = 12$ ; Nagelkerke's $R^2 = .47$ ; Hosmer-Lemeshow test: $\chi^2 = 12.27$ ; $df = 8$ ; $p = .14$				Model is statistically significant: $\chi^2 = 24.32$ ; $df = 12$ ; Nagelkerke's $R^2 = .64$ ; Hosmer-Lemeshow test: $\chi^2 = .918$ ; $df = 7$ ; $p = .996$			

Note: Significant  $p$  values are in bold;  $p < 0.01$ . Standard errors appear in parentheses.

**Table 18**

*Bootstrapped binary logistic regression models for accuracy of explicit judgments in High and Low Sense conditions*

Variables	Model 1: High Sense				Model 2: Low Sense			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	.027 (.162)	0.973	.709, 1.337	0.867	.651 (.431)	1.918	.824, 4.466	0.131
Plausibility	.085 (.175)	1.089	.772, 1.536	0.627	1.534 (.996)	2.637	.659, 3.649	0.123
Story Sense	-.464 (.488)	0.629	.241, 1.637	0.342	-.394 (.664)	0.675	.184, 2.477	0.553
Practice	.060 (.162)	1.062	.773, 1.458	0.711	-.535 (.399)	0.586	.268, 1.281	0.180
Wording	.125 (.155)	1.133	.835, 1.536	0.423	.812 (.468)	2.251	.900, 5.632	0.083
Organised	-.116 (.170)	0.891	.638, 1.243	0.496	.790 (.437)	2.204	.936, 5.192	0.071
Well-thought-out	.109 (.190)	1.115	.768, 1.620	0.567	-.352 (.355)	0.703	.351, 1.411	0.322
Voice	-.230 (.157)	0.795	.584, 1.082	0.145	-.873 (.400)	0.418	.191, .914	0.029
Face	.012 (.137)	1.012	.774, 1.323	0.930	.053 (.391)	1.054	.490, 2.269	0.893
Emotions Match	-.053 (.142)	0.948	.717, 1.253	0.708	-.078 (.398)	0.925	.424, 2.018	0.845
Less Emotions	.364 (.186)	1.439	1.000, 2.071	0.050	-.732 (.477)	0.481	.189, 1.226	0.125
Meta-emotion	-.094 (.053)	0.911	.821, 1.010	0.077	.247 (.167)	1.281	.923, 1.776	0.138
<b>Model description</b>	Model is statistically significant: $\chi^2 = 23.72$ ; df = 12; Nagelkerke's $R^2 = .17$ ; Hosmer-Lemeshow test: $\chi^2 = 12.10$ ; df = 8; p = .15				Model is statistically significant: $\chi^2 = 15.435$ ; df = 12; Nagelkerke's $R^2 = .39$ ; Hosmer-Lemeshow test: $\chi^2 = 12.09$ ; df = 7; p = .098			

Note: Significant  $p$  values are in bold;  $p < 0.01$ . Standard errors appear in parentheses.

### 9.8.7 *Not-sure answers frequencies*

Judges who reported 'not sure' for the question item pertaining to story sense were already removed to ensure only those who were sure were included in this study. For the remaining implicit cues, only percentage differences between both conditions that are higher than 10% are discussed. This table can be found in Appendix K. The High Sense condition found it easier to rate appeal practice and perceived plausibility however found it harder to rate the appellers' facial attractiveness.

## 9.9 Discussion

Results of the present study would seem to suggest that when the sense of a story was low, judges found it more difficult to accurately judge whether someone is innocence or otherwise. Judges displayed were better at telling apart an innocent appellant from one who is not when the appeals were presented in a highly sensible order, implying that their subjective assessments of how orderly the appeal sounds impacted decision processes. This would suggest that judges expected (or at least favoured) appellers to sound sensible in their appeal, even while holding the knowledge that these appellers were going through an ordeal of missing their loved ones.

Explicit judgments showed judges in the High Sense condition recorded an overall rating of 61.78%, much higher than the Low Sense condition at 49.78%, which was around chance level. Judges would seem to regard 'non-fluent' appellers as less believable and were more drawn to believe 'articulate' speakers to be credible.

The present study seemed to indicate what seems like an innocence bias for the High Sense conditions, with a 73.30% rating for true appeals and 50.70% for false appeals, totalling to an overall of a 69.20% rating. There was no such bias however in the Low Sense condition; instead, what appears to be a guilt bias seemed to be present. Essentially, judges performed better and more accurately for false appeals than truthful ones. The explicit judgments in the Low Sense condition's innocence detection yielded an unusually low rating. There are a few speculative reasons offered for this rating; firstly, this very low rating could be

because before this point no research has been carried out on high stakes stimulus material manipulated to have extremely low sense in content. Secondly, judges in this condition could have likely held a mental shortcut that the statements made by innocent appellants would make sense and that speech non-fluency was a cue of a false appeal. This heuristic could then have led to the occurrence of a guilt bias. This, if could be further replication and verified, would then infer an occurrence of an adaptive bias (in the face of limited information of the appellants and the emphasised non-diagnostic cue of non-fluent speech) that cannot conceivably be explained by other implicit cues. This could be why there were very few cues that had an association with explicit veracity decisions in both conditions across the board, in the present study. However, all inferences remain tentative and provisional at this stage.

Consideration was then given to these implicit judgments and their relationship with judges' explicit judgments. While the implicit judgments of overall credibility and perceived practice contributed to their explicit judgments, none of the implicit judgments used by judges in the High Sense condition contributed to accurate judgments. For judges in the High Sense condition, appellants were perceived to be more credible if they were innocent, and they were perceived to be more practiced if they were guilty. For judges in the Low Sense condition, not a single implicit judgment captured by the questionnaire significantly contributed to their explicit judgments, neither did any cue contribute to their accuracy in their explicit judgments. Here, several questions are raised; firstly, how then did judges in the High Sense condition achieve a high rating for innocent appellants and how did those in the Low Sense condition achieve a high rating for the guilty appellants? Secondly, how did those in the High Sense condition achieve a chance level ratings for false appeals and how did those in the Low Sense condition attain a low rating for truthful appeals?

Several points are discussed. To start off with, the two implicit judgments of 'overall credibility' and 'practice' predicted explicit decisions of judges in the High Sense condition. To explain why neither of these implicit cues afforded accuracy in their explicit decisions, as far back as the 1930's, Jackson (1932) proposed the term 'inferior speech', which was where hesitation pauses do not

occur and could mean a rehearsed speech. Unprepared speech is generally produced in haste where the speaker has no time to check every little detail in their sentences (Aitchison, 2008), whereas a prepared speech would tend to have more story flow and coherence with less errors. Aitchison (2008) contends that glib and fluent speakers are overvalued, whereas faltering ones are less liked. Judges in the High Sense condition perhaps erroneously used the implicit judgment of how 'practiced' the interviewees appeared for their explicit decisions.

Alternatively, no implicit judgments contributed to the predictive value of explicit judgments in the High Sense condition. While regression analyses showed relationships between certain implicit cues and the dichotomous outcome variable of 'is the appellant guilty or not' (thereby implying that judges were consciously or otherwise detecting some implicit cues and the questionnaire was capturing at least some of the cues they are using), none of the cues added to accuracy in their explicit veracity decisions. It is possible that while several implicit cues were captured, others that judges may have used were not. It is also possible that explicit judgments may be arrived at independently of conscious implicit judgments, as substantiated by Granhag (2006) and results in Anderson et al.'s (1999) study. As a caveat, the regression model for the High Sense condition has no predictors contributing to explicit decisions and yet could classify 70% of the responses correctly. This potentially demonstrates that, while valuable in allowing a speculation of which cues contributes to the guilt-innocence judgments, the model also has its limitation. One of the drawbacks is that the model is potentially overfit. An overfit model is expected to have a poorer fit with a new set of data (Frost, 2015). As discussed in the Methods section of Chapter 7, nevertheless, the variables included in the model were all based on theoretical background and checked for presence of multicollinearity (Dart, 2017). Due to the possible issue of overfitting, a researcher should be careful to put a heavy emphasis on this set of results as  $p$  values and  $R^2$  values could be misinforming (Frost, 2015). That said, the results would of course still necessitate a discussion.

Mann Whitney U results also showed significant differences for several variables. Judges thought story sense and story plausibility as significantly more

coherent in the High Sense condition when compared with their counterparts, both with medium effect sizes. An explanation as to why these results emerged is perhaps best explained by not only a reliance on different cues when a story is truthful and when it is not, but again also dependent on different cues when story sense is high versus when it is low. Perhaps individuals in the Low Sense condition turned to cognitive heuristics of voice attractiveness because the verbal cue of story sense and plausibility could not be deduced from; if they were seeking to use this cue they could not, thus having to adapt and resort to other available implicit cue short cuts. Judges in the Low Sense condition also thought that emotions displayed by the appellant did not match what the appellant was saying more so than those in the High Sense condition with a small effect size, although the emotions displayed by appellant was not different from those shown to the High Sense condition and was not tampered with in accordance to how they matched the speech of the appellant.

Judges displayed significantly lower levels of meta-emotion in the Low Sense condition, with a medium effect size. This seems to indicate that perceived story sense not only impacted the level of perceived emotional displays, it also seemed to have altered how coordinated or synchronised the perceived emotional displays seemed to be to a Judge. It also appears to have altered level of meta-emotion experienced by judges. Then again, it could also be said that judges displayed less meta-emotion with Low Sense appeals purely because they recognised that the appeals had been edited and thus were simply less genuine.

Interestingly, judges in the Low Sense condition also found the appellants' voices significantly less attractive than those in the High Sense condition, with a very small effect size however. Although not significant, facial attractiveness was perceived to be lower in the Low Sense condition. Literature in this area is scarce, mostly dominated by studies on how looks affects personality inferences and not vice versa, although Hui and Yam (2011) found in their study using forty-eight Chinese participants that their language proficiency displayed significant influences on how they were perceived, as well as how attractive they were seen to be although the latter comparison was weak. Lange, Zaretsky, Schwarz and Euler (2014) applied the sexual selection theory to language, finding that high

verbal proficiency increases perceptions of attractiveness. The authors divided participants into three groups listening to the same story content but with three differing levels of verbal proficiency in terms of lexical, grammatical, and fluency structures.

When examining not-sure responses in the present study, the percentage of not-sure answers for appealers facial attractiveness was 33.8% for the High Sense condition whereas it was only 22% for the Low Sense condition; this was perhaps because judges in the High Sense condition were not paying much attention to the facial attractiveness of the appealers, as they were busy cognitively encoding and processing other implicit cues of the verbal (and possibly non-verbal, not captured by the present questionnaire used) cues, thus eliciting a higher not-sure responses.

In summary, judges evidently found it more challenging detecting innocent appealers in the Low Sense condition compared to those who were exposed to the High Sense condition. Chi square results would also suggest that there was a meaningful relationship between both conditions and subsequent explicit veracity assessment. While a fuller discussion as to the reasons of why the chi square relationship was weak can be found in Sections 11.6 and 11.7 in Chapter 11, a second conjecture as to why this relationship was present but was weak is discussed here. This weak relationship is perhaps because some judges were aware of the crude and unpolished editing that took place for video clips in the Low Sense condition. Judges may have simply realised that an attempt to manipulate these video clips have taken place and this may have affected their explicit ratings. Editing of the video clips could have been more elegant; the weak relationship between conditions and explicit veracity ratings may have reflected this, rather than due to the actual variables in and of itself. Other possible limitations include the preservation and usage of only judges who agreed with the pre-set operational definition of speech sense of this study. This was done by asking participants what they thought of the speech sense in the videos. These steps were essential for this study to ensure that the selected independent variable is really measuring what it was set out to. There is a possibility that some of the variables in this study will be correlated and interrelate with each other, however, regression analysis cannot reveal much information of this possibility (Boduszek,

2017). In future studies, approaches such as path analysis (see Section 11.7.3) in combination with factor analysis could be used to explore correlations between variables and their interrelationships. This considers any likely mediatory or moderator relationships that may exist between the variables. Distinct from regression models but comparable to structural equation models, independent variables using these approaches can be both causes and effects of other variables (i.e., mediation) (Boduszek, 2017).

Evidently, replication of results is desired with a higher number of videos. Nonetheless, these initial results would infer that level of perceived story sense does indeed have an association with how accurate judges were in assessing whether the appellants were innocent or guilty. Overall, if the findings from this study can be replicated, retested and further explored using different or improved methods, they will be extremely relevant to the domain of televised press conferences, as most televised appeals are edited by the media before aired. News sites typically edit portions of appeals and/or interviews, not necessarily following a chronological order even. As revealed in the current chapter, editing the content of appeals can affect how these appeals are perceived implicitly, how appellants are explicitly judged and may impact how investigations are preceded. Differences in processing and thinking has not been studied when different levels of coherence are presented to participants in high-stakes situations; this study affords a level of insight in examining them in such mediums.

## CHAPTER 10: STUDY 4

---

### PERCEPTION OF PAIRED VERSUS SOLO APPEALERS IN TELEVISION APPEALS

From the previous chapters, it can now be said that the manner of Source presentation does have some level of impact upon observer's implicit and explicit judgments of appellant veracity. In the eyewitness literature, it is known that manner of presentation and police conduct during line-ups is crucial in the reliability of accurate identifications (Wells, Steblay & Dysart, 2011). Similarly, in the context of television appeals, having a member of the police force present or having two perpetrators collaboratively appealing together may perhaps increase story plausibility and hence the outcome of the veracity judgement by an observer. Additionally, statements solicited by police officers may be corrupted by both the interviewer and environment of the interview (Shuy, 1998).

Some television appeals are made with a police agent in the same room, which can possibly encourage viewers to adopt an innocence bias mentality and thus can hinder police investigation. Having the endorsement of the law enforcement agency during the appeal could increase an observer's confidence that the appellant is innocent. Observers may instinctively assume that police handling the investigation have eliminated the appellant as a suspect. This begs the question of whether the way television appeals and interviews are set up affects the way observers judge the innocence of the appellant. Specifically, there are no studies to date which investigate if two collaborating false television appellants are more successful at deception, as compared to when they appear solo.

Lie detection literature in this area is scarce. This present chapter examines a scenario where observers assess whether appellants give the impression of being more credible appealing alone, or in the presence of a co-appellant. To understand how a solo liar may be believed differently from paired liars, it is essential to review the literature on cognitive loading and possible mental shortcuts. But first the impetus behind the need to study perception of paired liars is discussed.

### **10.1 The need for studying perception of paired liars**

The study of co-appealing is germane to missing or murdered case investigation officers and can aid in the arrest of false appellants, as co-appealing may reveal much about the pair (Canter, 2000). For example, in a study of plagiarism by Carter (1999), some collaborating plagiarisers were more vulnerable

than the others in the group, and this was reflected in the complexity of their delegated task.

In fact, a fair number of television appeals to date feature more than one appellant appearing at a time. Table 19 shows the number of appellants who were guilty and innocent cross-tabulated with their appeal styles. These 61 appeals were collated from several different countries, which include United Kingdom, United States, Australia, New Zealand and Canada. The total number of cases was more than the number of appellants appearing with co-appellants or others and solo appellants as they appear in both styles in certain cases. These appeals are far more in number than the selected few used in the present thesis due to limitations in obtaining permission from news sites to use some of them for research purposes.

Additional support for studying paired liars was advocated by Vrij et al. (2010) who proposed that it is crucial to study liars and truth tellers in pairs as it reflects reality in the criminal world, where criminals usually work in groups or networks and not by themselves. Many crimes are planned and committed by several offenders working together rather than one person acting alone (Weerman, 2001). Bruinsma and Bernasco (2004) reported that criminals operating in a large-scale illegal market are usually close-knit, cohesive and ethnically homogenous with a high level of trust between the collaborators. However, the majority who co-offend do not do so in affiliation with big organised groups, but with one or two co-offenders (Reiss & Farrington, 1991).

Research in collective interviewing of lying and truth telling groups have recently been of interest. Vrij et al. (2010) emphasised the advantages of interviewing liars and truth tellers in pairs. There are several studies outlining the benefits of collective interviewing in drawing indicators of deception and generating discrepancies between truth-telling and lying pairs. Truth-telling pairs were found to interrupt and correct each other at a higher rate (Vrij, Jundi, Hope, Hillman, Gahr, Leal, Warmelink, Mann, Vernham & Granhag, 2012), exhibited a higher rate of behavioural synchrony in their transaction and were more interactive in their verbal transitions (Driskell, Salas & Driskell, 2012). Truth-telling pairs also made more eye contact with each other and made more eye contact with the interviewer than their counterparts (Jundi, Mann, Hope,

Hillman, Warmelink & Gahr, 2013). In their review, Vernham and Vrij (2015) established that cues to deception included a shortage of interaction and communication between lying individuals. Strömwall, Granhag and Jonsson (2003) found that accuracy in detecting pairs of liars and truth-tellers was moderate and indicators of truthfulness and lying as presented by these pairs were not very high in their accurately distinguishing the two. Von Hippel, Baker, Wilson, Brin and Page's (2015) study consisted of 54 groups of four to six individuals working together to arrive at a group verdict, when one individual in each group held the role of impairing this decision-making. The authors found that while there was a higher accuracy in finding out who this individual was once participants were given the information retrospectively that deception had taken place, interrogation did little to improve lie detection accuracy. Lastly, Granhag, Rangmar and Strömwall (2015) found the strategic use of evidence technique was effective in prompting cues to deception for small groups of suspects, more so than when evidence is disclosed early. This technique includes evidence disclosed later during an interrogation, or when evidence is disclosed late with an increased strength and precision during an interrogation.

So far, the aforementioned studies pertain to research in collaborative lying and collective interviewing. A vital missing piece in these studies are observer responses and their accuracy in judging co-offending versus solo liars – a gap which the current study aims to fill.

**Table 19**

*Number of appealers across UK, USA, Canada, New Zealand and Australia cross-tabulated with their appeal styles*

<b>Appeal type</b>	<b>Guilty</b>	<b>Innocent</b>	<b>Total</b>
<b>Paired</b>	3	16	19
<b>Solo</b>	30	15	45
<b>Total</b>	32	29	61

## **10.2 Lack of studies in perception of paired liars**

Most, if not all, prevailing studies in lie detection accuracy focus on observer accuracy when the liar is appearing solo, not in a pair or as a group (Bond & DePaulo, 2008; DePaulo & Pfeifer, 1986; Vrij & Mann, 2001; Fay & Middleton, 1941; Mann, Vrij & Bull, 2004; Kraut, 1980; Vrij, 2000; Fiedler & Walka, 1993; Ekman & Friesen, 1974; Meissner & Kassin, 2002; Vrij, Edward & Bull, 2001; Ekman, O'Sullivan & Frank, 1999; Miller & Stiff, 1993; Zuckerman, DePaulo & Rosenthal, 1981; Burgoon, Buller, Ebesu & Rockwell, 1994).

While it has been acknowledged that collective interviewing poses a strong potential to become a successful method to detect deceit, observers' affective responses and their judgments, in other words veracity judgment accuracy rates, of pairs versus solo appellers has previously been an unexplored process. While there are research investigating whether groups are better at detecting deception (Culhane et al., 2015), no study has looked at whether individuals are better at detecting lies when it is presented by a pair of liars or the same pair presented solo.

The only known study to date that has verged upon studying the perception of paired liars was conducted by Strömwall, Granhag and Reiman (2002). The authors let their participants watch videotaped statements of pairs of suspect individuals. These pairs are thought to be complicit and thus both associates were interrogated separately, twice. Their study indicated that observers focused more on the consistency between both suspects when making their veracity decisions, rather than the consistency within each suspect's statements. To be precise, they found that their participants used the cue 'consistency within pairs of suspects' twice as often as the cue 'consistency within single suspects'. In the next section, we consider the possibility of cognitive loading occurring during perception of paired liars.

### **10.3 Cognitive loading**

When viewing a video comprising of two appellers co-appealing, it is possible the participant may undergo an increase in mental processing when perceiving both appellers at once, and deciding whether one, or the other, or both, are innocent or guilty. Explained in Chapter 8, an increase in cognitive load

implies either an increase in the amount of information elements or sources to be processed or that a higher demand for processing is consigned for the same amount of information coming from the same source (Lavie, 2005; Sweller, 2010). Accuracy in task performance diminishes, and memory is impaired when an observer is cognitively loaded either visually or through auditory channels (Vredevelde et al., 2011; Weaver & Stewart, 2012). In knowledge of this, it is possible that observers assign incoming information into categories that are easier and quicker to process mentally in favour of rational and objective evaluation of them knowingly or unconsciously, which brings us to the next section.

#### **10.4 Pair/contagion bias**

One of the biases observers may subscribe to is the pair/contagion bias, introduced in Chapter 2. As introduced in Section 2.2.2, the contagion heuristic is defined as a mental shortcut whereby people identify someone as being of a certain trait or impression simply by close relationship with another person who exudes a similar trait or impression, and/or when both are perceived in combination with each other. A pair bias may arise as a consequence of this heuristic (Ng & Youngs, 2016). This bias has never been studied in the current context, although hints of it have been suggested in literature sporadically (Rozin & Royzman, 2001; Sutter, 2009; Bornstein et al., 2002). In the current study, it is hypothesised that an innocent appellant will be perceived as guiltier by means of observers using a contagion heuristic, resulting in a pair bias.

#### **10.5 Other biases in perception of paired liars**

It is unknown whether an innocence bias exists in a collaborative deception or guilt setting; all previous studies focus on deception at an individualistic level. Similarly, other biases discussed in previous chapters are also examined in the current study. Vrij et al. (2010) found that pairs of liars worked together to agree on a fabricated and plausible story before an anticipated interview. Therefore, it would be of interest to observe if paired liars are perceived to prepare appeals differently than if they were shown solo.

## 10.6 Key issues to explore

In light of this, the current study aims to explore questions such as ‘when a liar who presumably exudes negativity and a sense of ‘danger’ to beware of is paired with an innocent appealer who presumably emanates positivity, will negativity or positivity dominate when both entities combine? The general question then reflected is whether group or paired deception is more successful than solo deception. In co-offending, the inquiry is raised as to whether co-offending is more successful than solo offending (Weerman, 2001). The hypotheses of this study include: a) two appealers appealing together will be perceived to be more believable and their story more robust in plausibility than those appearing solo, b) a false appealer appealing with an honest appealer will appear more believable than when appearing solo, and c) the innocence bias will more likely be present in the paired versus solo setting.

## 10.7 Methods

### 10.7.1 Participants

The preliminary sample of participants totalled to 175 participants for both conditions, with 700 cases as each participant judged four cases of these appeals. None of the participants personally knew any of the featured appealers in this study but some reported knowledge of certain appeals from watching or hearing it on television. They were recruited using a mixture of snowball and opportunity sampling from the UK and Europe. They were from varying occupations ranging from prison officer to company director to farmer to psychologist to radio DJ and more; as well as nationalities/ethnicities (British, Asian, African, Scandinavian, Middle Eastern, European and American).

### 10.7.2 Material and Procedure

Two pairs of appeals will be chosen for the purposes of the present study. The first pair appeal will contain two guilty appealers, and the other a guilty appealer and an innocent appealer. The content of the videos are a mixture of press conference and interviews. One male and one female will be chosen for each level of emotionality so as to balance for any gender effects. For the Paired condition,

the videos will be shown to participants in their original form. For the Solo condition, the same videos will be edited so that the content of the appeal remains the same as the Paired condition but only showing one appellant at a time without altering any footage. Thus, both conditions will watch the exact same appeal content. Judges will be divided into two condition levels, where one level watched Paired appeals and the other Solo appeals. Video clips for the Paired condition ranged from 34.60 seconds to 37.64 seconds while video clips for the Solo condition ranged from 13.65 seconds to 19.93 seconds. It should be mentioned that it is difficult to find appeals consisting of two guilty appellants appearing together in the same video frame; consequently, this effect will be achieved through editing these appeals in order that the appellants appear sequentially. In contrast, the innocent and guilty pair will appear simultaneously and hence will not require extensive editing to achieve this effect.

This is a between-group design so all participants will experience the same four videos in a non-randomised order, with the predictor variable being Paired or Solo appellants and outcome variable being implicit assessments and explicit assignment of veracity. The advantage of using a between-subjects design is the ability to determine not only what implicit decisions were used by participants, but also what generated and predicted accurate explicit decisions. In addition, this design controls for appeal content, in that both conditions are exposed to the same appellants with the same appeal context, thus there is no issue of sender detectability here. 'Not-sure' responses for all implicit decisions were recorded for the remaining cases retained for analyses.

Analyses will be run for explicit decision to determine accuracy rates and if an innocence-guilt bias occurred. As for implicit responses, a Kolmogorov-Smirnov test for normality (Massey, 1951; Smirnov, 1948) will be conducted, and because the K-S test shows non-normal data a Mann Whitney U test for within-subjects cases will be conducted. The same methods of analysing results in previous chapters will be conducted in this study.

### *10.7.3 Ground truth*

Only cases where a substantial amount of evidence was obtained to place false appellers beyond reasonable doubt were used in this study, thus establishing ground truth. Appeler 2 was found guilty of perverting the course of justice by providing a false alibi for Appeler 1. Appeler 3 was convicted of murdering his stepdaughter with evidence of the body being found and a bloody t-shirt belonging to the victim within the property grounds of the appeler. Appeler 4 has been established innocent by the court of any crime in the murder of her stepdaughter.

## 10.8 Results

### 10.8.1 Participants

After removing cases where participants had knowledge of the appeals, 168 participants (557 cases) remain for analysis, with 71 participants (225 cases) for the Paired condition and 97 participants (331 cases) for the Solo condition. As some participants knew of more than one out of the four appeal cases, the total number of cases may be lower than the overall number of participants. The final pool of participants ranged from 16 to 72 years in age ( $M = 33.96$ ,  $SD = 12.62$ ).

### 10.8.2 Explicit veracity judgments

In the present study, rates of interest were between appellers appearing with a co-appeler (Paired condition) as compared to when appear by themselves (Solo condition). As seen in Table 20, the overall rating reported by judges in the Solo condition was 59.80%, higher than chance level. The judges who watched Paired appellers performed below chance level at 44.20%.

**Table 20**

*Judgment rates of explicit decisions for Paired and Solo conditions*

Appeler	Co-appeler	Guilt Ratings	Guilt Ratings
		Paired	Solo
Guilty male	Guilty female	36.40%	<b>66.30%</b>
Guilty female	Guilty male	29.50%	<b>55.70%</b>
Guilty male	Innocent female	<b>51.50%</b>	48.60%

		Innocence Ratings Paired	Innocence Ratings Solo
Innocent female	Guilty male	<b>59.40%</b>	<b>65.60%</b>
<b>Total</b>		44.20%	<b>59.80%</b>

### 10.8.3 Relationship between explicit rates and source presentation (Paired or Solo)

Chi-square results were  $\chi^2(1) = 10.430$ ,  $p < .05$ . This revealed that there was a statistically significant association between ratings of judges' explicit decisions and whether appellants are presented as pairs or solo. The research hypothesis that differences in accurate explicit veracity decisions were related to whether appellants were presented with a co-appellant or by themselves was supported by this analysis. Phi values indicate the strength of this relationship and chosen over Cramer's V as both variables used in the sample are dichotomous (Jones, 2009). The strength of association between the variables was .137, indicating a weak relationship (Davis, 1971).

### 10.8.4 Innocence bias

Analysing the breakdown by each appellant again, Table 20 distinguishes guilt accuracies from innocence accuracies. Un-bolded percentages represent accuracies below chance level and the bolded values represent accuracies above chance level. An innocence bias was not reliably evident in either condition. While it appeared to be more evident for the Paired condition, the results did not indicate a consistent bias for the guilty male appealing with an innocent female. An innocence bias was not apparent in the Solo condition, as some guilt accuracies were well beyond chance level. Results would suggest the innocence bias is not consistent in the present study.

### 10.8.5 Implicit veracity judgments

Results for the Kolmogorov-Smirnov test for normality (Massey, 1951; Smirnov, 1948) indicated that the score distributions deviated significantly from a normal distribution ( $D = .102$ ,  $p < .005$ ). Thus, non-parametric tests were applied

to assess observer judgments other than just dichotomous judgments of ‘guilty’ and ‘innocent’. Mann Whitney U test scores can be found in Table 22.

Cronbach’s alpha for each scale items can be found in Table 21. Implications are considered in the Discussion section and reviewed in depth in Chapter 11 under Sections 11.6 and 11.7.

**Table 21**

*Cronbach's alpha coefficients for Study 4*

Meta-emotion	Cognition	Appearance	Emotions
0.63	0.58	0.39	0.37

The implicit cue of story plausibility was perceived to be significantly higher in the Paired condition ( $M = 2.10$ ,  $SD = 1.31$ ) compared to the Solo condition ( $M = 1.82$ ,  $SD = 1.36$ ). The same pattern can be seen for the implicit cue of story sense with the Paired condition recording a mean of 1.92 ( $SD = 1.39$ ) and the Solo condition recording a mean of 1.69 ( $SD = 1.34$ ). Judges who viewed the appealers as a Pair felt that the combined appeals were significantly higher in plausibility that the accounts of what happened the day the victim went missing and the plea

**Table 22**

*Mann Whitney U item scores for Paired and Solo conditions*

	Paired means	Solo means	Z	R
<i>Innocence score</i>				
Credibility	1.85 (1.32)	1.93 (1.20)	-0.536	NA
<i>Cognition of Source content scores</i>				
Plausibility	2.10 (1.31)	1.82 (1.36)	-2.222	NA
<b>Story Sense</b>	<b>1.92 (1.39)</b>	<b>1.69 (1.34)</b>	<b>-1.990</b>	<b>-0.15</b>
<b>Practice</b>	<b>1.57 (1.10)</b>	<b>1.87 (1.08)</b>	<b>-3.102</b>	<b>-0.23</b>
Well-thought-out	1.81 (1.20)	2.02 (1.20)	-2.145	NA

Organised	1.88 (1.19)	1.83 (1.19)	-0.203	NA
Wording	1.63 (1.15)	1.65 (1.15)	-0.507	NA
<i>Emotion scores</i>				
Less Emotion	1.79 (1.25)	1.89 (1.23)	-0.997	NA
Emotions Match	2.10 (1.36)	2.06 (1.31)	-0.456	NA
<i>Appearance scores</i>				
<b>Face attractiveness</b>	<b>1.14 (1.04)</b>	<b>1.62 (1.14)</b>	<b>-4.351</b>	<b>-0.18</b>
Voice attractiveness	1.64 (1.25)	1.80 (1.20)	-1.568	NA
<i>Meta-emotion score</i>	9.74 (4.50)	9.25 (3.76)	-1.658	NA

Note: Significant *p* values are in bold; *p* <0.01. Standard deviations appear in parentheses.

for the return of the victim had taken place. Additionally, judges in this condition felt that the accounts given made significantly more sense. As for preparation scores, judges viewed the appellers presented as a couple as less practiced and less well-thought-out implying that they seemed less rehearsed and therefore possibly more ‘genuine’. Similarly, overall, judges viewed the appellers presented as a couple as less credible than those who viewed them presented individually. Remarkably, both conditions did not record any significant differences in either the perception of emotive implicit cues or the meta-emotion bias. Thus, both groups felt that the emotions displayed by the appellers were alike whether appellers were shown together or singularly. There were no significant differences to how sympathetic as they felt when shown the appellers as a Pair or as Solo members.

#### 10.8.6 Logistic regression analyses

In this study, tolerance was greater than .10 for all items, and the variance inflation factor was also less than 10 for all items. This suggested that multicollinearity was not an issue in this study (meta-emotion, Tolerance = 0.73, VIF = 1.37; credibility, Tolerance = 0.82, VIF = 1.22; plausibility, Tolerance = 0.76, VIF = 1.31; practice, Tolerance = 0.85, VIF = 1.18; voice attractiveness,

Tolerance = 0.82; VIF = 1.22; organised speech, Tolerance = 0.82; VIF = 1.21; face attractiveness, Tolerance = 0.89; VIF = 1.12; less emotions, Tolerance = 0.87; VIF = 1.15; wording, Tolerance = 0.88; VIF = 1.13; well-thought-out, Tolerance = 0.81, VIF = 1.23; emotions match, Tolerance = 0.83, VIF = 1.21; practice, Tolerance = 0.85, VIF = 1.18). To reduce Type 1 error and increase level of confidence in results, the more stringent cut-off level of ' $p < 0.01$ ' will be used.

First, consideration was given to the Paired condition as to which implicit veracity judgments they may have consciously or otherwise used towards to their explicit one. The binary logistic regression model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2 (12, N = 168) = 63.51, p < .005$ . The model as a whole explained between 32% (Cox and Snell  $R^2$ ) and 43% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 78% of cases. As shown in Table 23, two of the independent variables made a unique statistically significant contribution to the model for the Paired condition. These were judgments perceived voice attractiveness ( $OR = 1.99$ ) and matched-ness of emotions ( $OR = 1.74$ ).

For the Solo condition, the full model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2 (8, N = 168) = 68.465, p < .005$ . The model as a whole explained between 19% (Cox and Snell  $R^2$ ) and 25% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 69% of cases. As shown in Table 23, two independent variables made a unique statistically significant contribution to the model for this condition. The stronger predictor was perceived overall credibility ( $OR = 1.49$ ), followed by how well-thought-out they judged the appeals to be ( $OR = 0.69$ ).

Consideration was then given to the Paired condition as to which implicit veracity judgments accurately supported their explicit one. Another binary logistic regression with the actual accuracy of judges' explicit veracity judgments as the dependent measure in this group was conducted. The model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2 (8, N = 168) = 15.63, p <$

.005. The model as a whole explained between 7% (Cox and Snell  $R^2$ ) and 9% (Nagelkerke  $R^2$ ) of the variance in accuracy, and correctly classified 63% of cases. As shown in Table 24, no independent variables made a unique statistically significant contribution to the model for this condition with a cut-off point of  $p < 0.01$ .

As for which implicit veracity judgments measured in the present study may have contributed to the accuracy of their explicit judgment in the Solo condition, the full model indicates that it was able to distinguish between respondents who explicitly reported Guilty from the ones who reported Innocent, with results reported as  $\chi^2 (12, N = 168) = 22.795, p < .005$ . The model as a whole explained between 7% (Cox and Snell  $R^2$ ) and 9% (Nagelkerke  $R^2$ ) of the variance in veracity judgments, and correctly classified 65% of cases. As shown in Table 24, one independent variable made a unique statistically significant contribution to the model for this condition. This implicit variable was perceived story plausibility ( $OR = 1.29$ ).

**Table 23**

*Bootstrapped binary logistic regression models for explicit judgments between Paired and Solo conditions*

Variables	Model 1: Paired				Model 2: Solo			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	.392 (.167)	1.480	1.066, 2.055	0.019	<b>.403 (.113)</b>	<b>1.496</b>	<b>1.198, 1.868</b>	<b>0.000</b>
Plausibility	.198 (.177)	1.219	.861, 1.726	0.265	.057 (.102)	0.944	.774, 1.130	0.574
Story Sense	.306 (.165)	1.358	.984, 1.875	0.063	.077 (.103)	1.080	.882, 1.323	0.456
Practice	.210 (.208)	1.234	.821, 1.853	0.312	-.296 (.124)	0.743	.583, .948	0.017
Wording	.100 (.195)	1.106	.755, 1.619	0.606	-.197 (.115)	0.821	.655, 1.029	0.087
Well-thought-out	-.404 (.215)	0.668	.438, 1.018	0.061	<b>-.362 (.113)</b>	<b>0.695</b>	<b>.556, .868</b>	<b>0.001</b>
Organised	-.120 (.211)	0.887	.586, 1.334	0.569	.044 (.114)	1.045	.835, 1.307	0.700
Face	-.167 (.200)	0.404	.572, 1.252	0.404	.066 (.118)	1.068	.848, 1.344	0.576
Voice	<b>.691 (.192)</b>	<b>1.996</b>	<b>1.370, 2.908</b>	<b>0.000</b>	.212 (.115)	1.236	.987, 1.547	0.065
Less Emotions	-.015 (.184)	0.985	.687, 1.414	0.936	0.109 (.108)	0.897	.725, 1.109	0.314
Emotions Match	<b>.556 (.166)</b>	<b>1.744</b>	<b>1.258, 2.416</b>	<b>0.001</b>	.180 (.101)	1.197	.982, 1.459	0.075
Meta-emotion	-.054 (.057)	0.947	.847, 1.060	0.345	.070 (.039)	1.072	.994, 1.157	0.071
<b>Model description</b>	Model is statistically significant: $\chi^2 = 63.51$ ; df = 12; Nagelkerke's $R^2 = .43$ ; Hosmer-Lemeshow test: $\chi^2 = 1.159$ ; df = 8; p = .997				Model is statistically significant: $\chi^2 = 68.465$ ; df = 12; Nagelkerke's $R^2 = .26$ ; Hosmer-Lemeshow test: $\chi^2 = 6.750$ ; df = 8; p = .564			

Note: Significant  $p$  values are in bold;  $p < 0.01$ . Standard errors appear in parentheses.

**Table 24**

*Bootstrapped binary logistic regression models for Accurate explicit judgments between Paired and Solo conditions*

Variables	Model 1: Paired				Model 2: Solo			
	B (SE)	OR	95% C.I. (Lower, Upper)	p	B (SE)	OR	95% C.I. (Lower, Upper)	p
Credibility	.039 (.145)	1.039	.782, 1.382	0.790	-.038 (.109)	0.962	.777, 1.192	0.725
Plausibility	.273 (.164)	1.314	.952, 1.813	0.097	<b>.257 (.098)</b>	<b>1.293</b>	<b>1.067, 1.567</b>	<b>0.009</b>
Story Sense	.045 (.151)	1.046	.777, 1.407	0.768	.076 (.100)	1.079	.886, 1.313	0.451
Practice	-.146 (.179)	0.864	.608, 1.228	0.416	-.023 (.119)	0.978	.774, 1.235	0.850
Wording	.048 (.170)	1.049	.752, 1.463	0.778	-.077 (.108)	1.003	.748, 1.145	0.476
Well-thought-out	.028 (.169)	1.028	.738, 1.432	0.870	.055 (.108)	1.057	.855, 1.307	0.609
Organised	-.217 (.182)	0.805	.563, 1.151	0.234	-.074 (.106)	0.929	.755, 1.143	0.486
Face	-.287 (.174)	0.751	.534, 1.055	0.099	.003 (.111)	1.003	.807, 1.246	0.979
Voice	.326 (.158)	1.386	1.018, 1.887	0.380	.245 (.109)	1.277	1.031, 1.582	0.025
Less Emotions	-.040 (.152)	0.961	.713, 1.295	0.793	-.133 (.103)	0.875	.715, 1.071	0.197
Emotions Match	.076 (.140)	1.079	.820, 1.420	0.857	.164 (.097)	1.178	.975, 1.424	0.090
Meta-emotion	-.099 (.049)	0.906	.823, .997	0.044	-.013 (.036)	0.987	.920, 1.059	0.987
<b>Model</b>	Model is statistically significant: $\chi^2 = 14.72$ ; df = 12;				Model is statistically significant: $\chi^2 = 22.795$ ; df = 12;			
<b>description</b>	Nagelkerke's $R^2 = .12$ ; Hosmer-Lemeshow test: $\chi^2 = 8.065$ ; df = 8; p = .427				Nagelkerke's $R^2 = .092$ ; Hosmer-Lemeshow test: $\chi^2 = 6.754$ ; df = 8; p = .563			

Significant  $p$  values are in bold;  $p < 0.01$ . Standard errors appear in parentheses.

### *10.8.7 Not-sure answers percentages*

Only percentage differences between both conditions that are higher than 10% are discussed. This table can be found in Appendix L. The Paired condition found it harder to rate perceived appeal practice and perceived face attractiveness of the appellant than did the Solo condition.

## **10.9 Discussion**

Most existing lie detection research has focused on subjective veracity assessments of individual liars and truth tellers, when in fact deception can and often does occur with more than one liar in a situation. Therefore, the present chapter pursued to fill this gap with the expectation to reveal differences in underlying cognitive processes for both conditions with innocence-guilt verdicts. Two appellants appealing together were judged to be more believable and their story more robust in plausibility than those appearing solo. A false appellant appealing with an honest appellant appeared more believable than when appearing solo, and the innocence bias seems to be more pronounced in the paired versus solo setting (thus not consistently evident in both conditions).

Underlying differences in mental processing were implied via regression analyses. Main results of the present study would suggest that co-appealing duped observers more successfully than when the same appellant was presented solo, in this sample. Judges were more accurate in judging the sender's innocence or guilt when the appellant appeared alone without a co-appellant. When a guilty appellant was paired with a fellow guilty co-appellant, judges tended to assign more ratings of innocence to both appellants than when they were both presented alone. This indicates that innocence can be falsified when two liars corroborate and validate each other, as Vrij et al. (2010) speculated. Liars will work together to come up with a plausible story. This strengthens the lie, duping an observer into believing a fictitious story that seems to be truthful. When a story is plausible and is backed up by another and not just one of them, it appears that the plausibility of the story increases, at least in the context of the appeals chosen in the present study.

However, this situation was reversed when an innocent appellant is paired with a guilty co-appellant. Judges in this study tend to assign slightly lower ratings

of innocence to the innocent appellant than when this appellant appears alone, inferring a negativity bias (Rozin & Royzman, 2001). The results revealed a difference in the interplay between the guilty pair and that between guilty and innocent pair. For the guilty pair, there was a large difference in ratings whereas for the pair containing one guilty and one innocent appellant, the innocent appellant was more accurately judged when appearing alone and the guilty one is more accurately judged as 'guilty' appearing with his innocent wife (though the magnitude of this percentage difference was only around 3.00%). It may be beneficial for a guilty appellant to co-appeal with an innocent appellant as the present study found, although the percentage of this difference is very small (see Table 20).

A few explanations are provided for these results. Judges in the Paired condition may have performed less well in their explicit judgments due to cognitive loading, when they had to pay attention to two appellants at once in making their assessments. In the Paired condition, judges may have perhaps been inundated with extra auditory and visual stimuli that the Solo condition did not experience. As introduced in the literature review, the limited ability for our working memory to hold chunks of information and let alone process them conceivably decreased participant accuracy of judging coupled high stakes lies (Cowan, 2001).

In reality, the guilty pair in the present study presumably rehearsed their story before being interviewed by the media. As introduced in Section 6.2.2.3, Vrij et al. (2010) contended that pairs of liars tend to rehearse their stories before an interview. The authors found that while giving more vague answers compared to truth tellers, pairs of liars anticipated and prepared answers for possible questions they would be asked about. Well-prepared liars are harder to detect than those who are not (Bond & DePaulo, 2008). Interestingly, results in Table 23 would suggest that for judges in the Solo condition, appellants were perceived to have higher credibility when they were innocent and their appeals were perceived to be not thought-out very well if they were guilty. For those in the Paired condition, appellants were perceived to have a more appealing voice and a higher matched-ness of emotions with the appeal content if the appellants were innocent. Additionally, Table 24 would suggest that for judges in the Solo condition, innocent

appealers were perceived to have more plausible stories and this would appear to be an 'accurate' cue judges have used for this condition.

Although vastly conjecture at this stage, a reasonable explanation is that appealers appearing as a pair presumably reinforced judges' beliefs that this pairing itself was perhaps an indicator that they were more likely to be innocent. Alternately, when judges were not exposed to both appealers appealing together, this 'pair bias' was ostensibly not made available to them to then use as an indicator of guilt and innocence, compelling them to look for other available cues such as verbal cues (i.e. how credible overall the appealers appeared and how well-thought-out the appeal seemed to be) in making their veracity decisions.

Mann Whitney U results would show several significant differences. A significant result was discovered for appeal sense between the Paired and Solo condition, with a small effect, but was not significant for overall appealers' credibility. Judges found that the stories made more sense when pairs of appealers were presented together rather than separately. A remark here is that the Paired condition's appeals may have made more sense simply because there is more information available on the particular case. Judges performed less accurately when the appeals presumably made more sense (when more information was present) as they were presented as a joint effort. This seems to suggest the occurrence of an innocence bias in the presence of another induced bias – the pair bias. The innocence bias was not reflected when the pair bias was not present. Additionally, results did not seem to suggest a meta-emotion bias in the current sample. This implies that judges did not rely on the cues of emotional display and their own sympathetic reactions to come to their veracity decisions.

As a tentative conclusion, results from the present study would tend to suggest that how appeals are presented (i.e. whether appealers were presented solo or whether the same ones were presented with another) can alter how they are perceived. This inference was suggested by explicit veracity decisions, regression analyses and Mann Whitney U tests between Solo and Paired conditions. The manner of presentation seems to impact what implicit cues are consciously or unconsciously detected, and this also seems to initiate differences in underlying cognitive processing. Chi square results also suggest that a

relationship exists between how the appeals were presented and judges' assessments of whether appellants were innocent or guilty. A full discussion as to why this relationship may be significant but weak can be found in sections 11.6 and 11.7 in Chapter 11. Limitations include using a mixture of edited videos that appeared simultaneously and sequentially, which could have confounded results. These difficulties are inherent in working with real world stimulus material. Additionally, the usage of only a pair of liars and a pair consisting of a truth teller and a liar (low number of stimulus videos shown to Judges) connotes that the results (and subsequent interpretation in this study) could be idiosyncratic and only relevant to the videos used here.

Nevertheless, subject to further replication with a higher number of videos and dependent on further clarification as to the differences in cognitive processes between the two conditions, these findings are particularly germane to the area of televised press conferences. It raises questions of the potential importance in paying attention to the number of appellants pleading for a case. On a broader spectrum, clear guidelines as to how to carry out these press conferences in partnership with the media are not yet available to most law enforcement agencies around the world. The present study reiterated that guidelines and protocols are much needed to moderate biases while ensuring effective investigation in missing and murdered relatives' cases. Moreover, this chapter establishes a foundation for further studies to be conducted. For example, the finding that judges in the Paired condition performed less well in their explicit judgments with the suggestion that this may be due to cognitive loading could perhaps be empirically tested by presenting these videos to judges in a sequential order rather than a simultaneous order.

# CHAPTER 11

---

## SUMMARY OF RESULTS FROM EXPERIMENTAL STUDIES

## FINDINGS

### 11.1 Observers' explicit judgments

Issues considered in the present thesis were 1) whether judges were able to assess innocence and guilt in these appeals with a higher accuracy in one condition than the other, 2) whether variances in underlying cognitive processes lie across different experimentally manipulated conditions, 3) whether biases were present in these different conditions and 4) if so, which conditions are these biases more likely to surface.

Study 1 presented only truthful appeals to judges in a within-subjects experimental design. A difference is seen between both conditions. Highly Emotional appeals reported a veracity rating of 56.00%, whereas Low Emotional appeals reported a rating of 40.60%. Study 1 highlighted the limitations of using only appealers who are innocent in the sample. In addition to not being able to study a guilt bias, by using only truthful appeals it could not be determined from logistic regression whether results are predicting accuracy, or simply what implicit judgments the judges used for the appeal videos. A between-subjects design was employed for Study 2 not only so that a guilt bias can also be investigated, but also to control for the issue of sender detectability (to moderate for outcomes resulting from content differences).

In Study 2, a difference was seen in explicit judgments for different Source presentations, with 65.70% rating for false appeals the Audio-only condition compared to 54.90% in the Audio-visual condition. Innocence detection was 72.60% (much higher than chance level) in the Audio-only condition compared to 52.00% in Audio-visual condition (approximately at chance level). Study 2 inferred a visual bias, whereby being able to see the appearance of an appealers influenced veracity judgment assessments.

Results from Study 3 would suggest a story-sense bias. The High Sense condition recorded a total accuracy rating of 62.00% whereas the Low Sense condition yielded a total accuracy rating of 44.9%. Judges in general considered non-sensible appeals to be less convincing and tended to favour 'articulate' speakers more in judging them to be more innocent. The explicit judgments in the Low Sense condition innocence detection yielded a low accuracy rating. Judges

here perhaps held the (mistaken) mental short cut that because the appeal was so nonsensical and conceptually incoherent that the appellars most likely fabricated the appeal. A strong guilt bias was evident in the Low Sense condition for both innocence and guilt explicit accuracy ratings. This is further expounded in Section 11.4 below.

Results from Study 4 would suggest a pair bias. This study found that co-appellars were more successful in duping observers in contrast to when the same appellant is presented solo in this sample. Judges were biased into believing two guilty appellars when they co-appealed together (with an accuracy rating of 44.20%), whereas they were less biased and hence more accurate in judging the same appellars when they were presented solo with an accuracy rate above chance level at nearly 60.00%.

Across all four studies differences in accuracies were found between both conditions judges were assigned to, suggesting different underlying cognitive processing transpired when observers are shown similar appeals that were carefully manipulated. The highest rating of 'accuracy' was yielded for judges in the High Sense group, at 72.85%. When compared with truth-lie detection, these ratings are unusual in lie detection research in general (though similar to the percentages recorded in higher stakes research in lie detection such in Vrij et al.'s (2006) study with police officers. In fact, Vrij and Mann (2001) showed police officers television appeals and their participants only yielded an overall accuracy rating of 50%.

The lowest 'accuracy' rating was yielded for the group of judges who watched appeals with Low Sense, at 22.70%. What seems like a guilt bias was 76.85% accuracy for the Low Sense condition as well. The lowest total accuracy was recorded by judges who watched appeals with low levels of emotionality in Study 1, and the highest was recorded by the Audio-only condition in Study 2. A summary of explicit veracity ratings is presented in Table 25 below.

## **11.2 Explicit veracity judgments across Studies 1 to 4**

### **Table 25**

*Overall explicit accuracy rates by conditions across Studies 1 to 4*

Study	Conditions	Innocence rates	Guilt rates	Total rates
1	High Emotionality	<b>56.00%</b>	NA	<b>56.00%</b>
	Low Emotionality	40.60%	NA	40.60%
2	Audio-only	<b>72.60%</b>	<b>65.70%</b>	<b>69.20%</b>
	Audio-visual	<b>52.00%</b>	<b>54.90%</b>	<b>53.45%</b>
3	High Sense	<b>72.85%</b>	<b>50.70%</b>	<b>61.78%</b>
	Low Sense	22.70%	<b>76.85%</b>	49.78%
4	Paired	<b>59.40%</b>	39.13%	49.27%
	Solo	<b>65.60%</b>	<b>56.87%</b>	<b>61.24%</b>

\*Percentages higher than chance level appear in bold

Study 1 was not indicative of an innocence bias, as all appeals were honest and yet not all the accuracy ratings yielded were above 50.00%. Study 2 also reported no such bias for the Audio-only condition, with 72.60% accuracy for true appeals and 65.70% for false appeals. However, the Audio-visual condition reported ratings that are similar to the general research with accuracies close to chance level. This group would seem to display an innocence bias. Again, Study 3's results were not indicative of a consistent innocence bias; it was not present in the Low Sense condition. Finally, in Study 4, while it appears to be more evident for the Paired condition, the results did not indicate a consistent innocence bias for the guilty male appealing with an innocent female. The innocence bias claim was not present in the Solo condition, as some guilt accuracies were well beyond chance level.

Results from Studies 1 to 4 would suggest that the innocence bias is potentially merely a methodological construct of how experiments are set up and what cues become perceptible or important to judges. The inconsistency of the innocence-guilt bias across all four studies (more absent than present) would tend to infer that the occurrence of this bias is possibly only a by-product of and/or is naturally contingent upon the existence of certain other biases and/or when other biases are not as salient. The possibility of an innocence bias has not been studied using this high-stakes scenarios, let alone in controlled experimental designs

where the stimulus material was manipulated. For example, the likelihood that it occurs has never been studied in a collaborative deception setting (i.e. Paired versus Solo). Therefore, it must be stated that any conclusions should remain tentative at this stage as further examination into the processes of how people make innocence-guilt decisions and how this interrelates with the concept of truth-lie judgments needs to be made. What is interesting is that, for example, Study 3 would suggest that judges are noting different cues for guilty and innocent appellants and awarding different ratings of innocence and guilt for each. This would seem to infer that thinking processes are different for when a sender is guilty or innocent.

While it is recognised that the explicit question asked here pertains to a judgment of guilt or innocence of the appellant and not the truthfulness or deception of the appellant and both measures are not identical, there are conceivably overlaps between the two assessments in the present thesis. While a judgment of innocence-guilt and a judgment of truth-lie are not entirely identical, in the context of judging a brief video such as in the present thesis it appears that the two are linked. In other words, it would be difficult for a person watching a video in which the subject implies their innocence to make their decision whether this subject is guilty without thinking that the subject was lying. Guilt in this context refers to the appellant being culpable of being involved in the murder or disappearance of the relative they are appealing for, and the assumption of guilt lies on the belief that they have in fact been culpable and therefore are lying in their appeal.

Because the explicit question in this thesis asked was 'does the appellant seem guilty or not guilty of being culpable in the murder/disappearance of their relative' a judge must then use what they are given to make a decision, in this situation being the appeal video itself. In other cases, such as in an interrogation or in a jury trial, other forms of evidence or information about the case can be used to determine judgments of guilt or innocence apart from what the suspect says and how they behave. Therefore, while of course conjecture at this stage, in the present context it is reasonable to assume that the judgment of innocence/guilt interrelates with a truth/lie judgment. To believe that the

appealer is guilty in this context indicates that the appealer has in some way or form been involved in foul play, and therefore they are lying when giving an account during the press conference. They are guilty of holding the knowledge of what happened to their relatives (i.e. if they murdered or kidnapped them), and therefore are insincere in their pleas when asking for the relative to turn themselves in or seeking help from the public for tips. While judges were not being asked to assess deception *per se*, they may very likely incorporate it during decision-making.

### 11.3 Guilt bias

In Study 3, judges displayed what seems a guilt bias. The possibility of this bias has never been compared when different levels of sense are presented in high stake situations; this study affords a level of insight into the likelihood of its presence in such mediums. As can be seen across the four experimental studies where certain cues are manipulated and accented, the innocence or guilt bias either ‘surfaced’ or was ‘non-existent’ within different manipulations in conditions. The resulting innocence and guilt biases in the present thesis can perhaps be understood as a manifestation of an educated and rational deduction that satiates the deficiency of low or no diagnostic cues. While initially implied as an error, for example as suggested in Chapter 2, and/or that they highlight the wrong cues as mentioned in Chapter 4, the findings from this thesis led to an alternative conclusion. Although these were originally thought of as biases, as they are often referred to in some literature (Levine, 2014; McCornack & Levine, 1990; Levine & McCornack, 2001; Tversky & Kahneman, 1974; Peace & Sinclair, 2012; Paunonen, 2006; Ekman et al., 1999; Forgas, 1995; Doss, 2002; Operario & Fiske, 1999) it became clear that these are not errors in any strict sense but rather, most likely, decisions that are adaptive (Gigerenzer & Selten, 2001; Street, 2015).

Borrowing from lie detection literature, the ALIED account argues that individuals will depend more heavily on context-general information when individuating cues have low diagnostic value or are not diagnostic at all. The solitary use of context-general information may result in biases in decision making. This account proposes that individuals are adaptive in making

judgments that are informed. When uncertain, context-general information helps people to make decisions (Gigerenzer & Gaissmaier, 2011; Simon, 1990). Street (2015) proposes that as there is less diagnostic individuating information available, context influences decision making more. They adapt what they use to form their decisions, but this is not based on a 'predisposition' to using their knowledge of context either (Brunswik, 1952). The ALIED account argues that individuals will depend more heavily on context-general information when individuating cues have low diagnostic value or are not diagnostic at all. The solitary use of context-general information may result in biases in decision making.

The ALIED approach may help shed light as to why judges sometimes displayed a supposed 'innocence bias' and at times a 'guilt bias' just as readily. Street (2015) maintains that the ease at which the truth and lie biases can be employed is at odds with their very claim. The ALIED approach claims that the presence and direction of these biases are all dependent on context. Just like the truth bias, the tendency to disbelieve a statement or sender can be observed just as rapidly and effortlessly, because the lie bias is also not a cognitive default (Street, 2015). It claims that biased responses, rather than being ubiquitous, are adaptive and amenable to reflect an understanding of the given context. If there was cognitive bias in believing others are always honest, a quick and effective default towards disbelieving others cannot occur just as quickly. While this is still very much conjecture at this stage, the hypothesised innocence and guilt 'biases' are most likely either mere results of methodological constructs, or are remnants from an adaptive decision-making mode in the face of low or unhelpful information about the source.

#### **11.4 Implicit judgments across all studies**

The second general objective of the current thesis was to consider the role of implicit veracity judgments in the high-stakes scenarios of television appeals and how they relate to explicit judgments. The aim of this section is to discuss these results from Studies 1 to 4.

### 11.4.1 Overall credibility

**Table 26**

*Differences between conditions for Overall Credibility across Studies 1 to 4*

Study	Conditions	Means (SD)	Z	R	Test
1	<b>High Emotionality</b>	<b>2.22 (1.39)</b>			<b>Wilcoxon's Rank</b>
	<b>Low Emotionality</b>	<b>1.40 (1.14)</b>	<b>-2.482</b>	<b>-0.24</b>	
2	Audio-only	2.05 (1.21)			Mann Whitney U
	Audio-visual	1.78 (1.31)	-2.118	-0.08	
3	<b>High Sense</b>	<b>2.48 (1.27)</b>			<b>Mann Whitney U</b>
	<b>Low Sense</b>	<b>1.39 (0.98)</b>	<b>-2.036</b>	<b>-0.12</b>	
4	Paired	1.85 (1.32)			Mann Whitney U
	Solo	1.93 (1.20)	-5.36	NA	

Note: Significant  $p$  values are in bold ( $p < 0.01$ ).

First, a review of the responses given to the question “Is the appealer credible overall?” is considered. It can be seen from Table 26 that two out of four studies recorded a significant difference in responses for the question of appealer credibility across conditions. Table 29 shows judgments predicting judges’ explicit decisions and Table 30 shows the prediction as to how accurate they were in their explicit judgments across all experimental conditions. The following paragraphs discuss each study and the biases they pertain to in detail.

### 11.4.2 Emotional bias error

In Study 1, the results would tend to suggest that judges did not use any of the cues provided in the questionnaire to make their explicit decisions. While their explicit decisions would seemingly reveal that they valued highly salient emotional displays, this was not reflected in their thinking patterns using the implicit cues in the questionnaire. This potentially highlights the division between explicit and implicit judgment processing, where on occasions judges cannot or do not identify implicit cues, however, can make a correct explicit judgment. Alternatively, it is possible that explicit judgments may be arrived at

independently of conscious implicit judgments, as put forward by Granhag (2006) and as suggested by the results in Anderson et al.'s (1999) study.

This study provided preliminary insight into an observer's expectations of how an appellant should display an appropriate amount of emotional intensity and quality. At this stage, this study was unable to uncover what differences in thinking patterns might exist when observers watch these videos, at least with the current questionnaire used. Explicit decisions would suggest with highly emotional appeals, judges seemingly used the level of emotion displayed and gauged whether this was an appropriate level for the situation of a missing or murdered relatives' case, to judge whether they are explicitly guilty or not. Both highly emotional appeals contain spoken emotions and mental states such as 'fear', 'love', 'miss you terribly', whereas appeals with low emotional content chosen for this study did not contain verbalised emotions. Participants in this study perhaps registered the availability of these spoken emotions as a verbal cue, nonetheless this inference remains to be tested. While tentative at this point and necessitates further research, literature would also suggest that the presence of emotions reduces cognitive attention and reduces the capability for deep information processing (Wilder, Simon & Faith, 1996).

#### *11.4.3 Appearance bias error*

Study 2 was built on the premise of Mann et al. (2004) who found that people who were more accurate in truth and lie detection utilised verbal cues such as vague reply and contradictions in the story. Those who were poorer lie detectors mentioned more visual cues. In other words, people who paid closer attention to what the sender said performed better and more accurately. In this study, the expectation was that by eliminating a channel of source presentation where errors will be deterred from occurring, judges would be more accurate in their explicit decisions as they would pay more attention to diagnostic verbal implicit cues.

The results in this study would suggest that judges in the Audio-visual condition paid attention to more verbal cues in number, they were less accurate in their explicit decisions. They may have experienced cognitive overload in comparison to judges the Audio-only condition as they not only have audio data to

process but also visual data. In this study, a clearer indication of differences in thinking patterns was found compared to Study 1. Essentially, thinking was different when Source presentation was manipulated, and thinking was again different for deceptive and truthful stimuli (which shows a benefit of including both types of stimuli).

#### *11.4.4 Cognition in coherent speech bias error*

While Studies 1 and 2 concerned manipulations of visual cues, Study 3 examined manipulations of verbal cues. The results from this study would seem to suggest that judges valued and favoured a sensible speech over one that was not logical in terms of story progression. Taking into consideration that most appealers, whether true or false, would probably prepare before appearing on national television, judges in this study appeared to hold the (mistaken) belief that an appeal that is practiced was more believable.

Judges would seem to have resorted to the heuristic of a 'guilt bias' when unable to assimilate or understand speech easily. Judges ultimately mistrusted stories that did not make sense, contrary to the evidence in the literature which suggests that liars tend to practice their stories before interviews. Care in generalisation of results needs to be taken as only those who corresponded with the present author's definitions of high and low coherence were kept for analyses.

#### *11.4.5 Pair or contagion bias*

The number of appeal cases selected for the purposes of this study was notably small (as was for Studies 1 to 3). Difficulty remains in being able to find cases where only one guilty appealers is verbally appealing with another innocent appealers or both guilty appealers are featured, where participants have no prior knowledge of the case. For this reason, firm conclusions cannot be made at this stage. However, the results from this study indicated a difference in underlying cognitive processes when judges were shown two appealers in comparison to when they were shown the same ones appearing solo. Implicit judgments revealed different cues were paid attention to for both conditions, in theory showing differences in thought processes for each. If able to be replicated with a higher

number of videos, this has bearings for how appeals should be televised and how the police should conduct investigation for these cases. In future studies, perhaps a pair of innocent co-appealers should also be added and a larger number of these appeals should be considered if possible.

### **11.5 Effect sizes across Studies 1 to 4**

A summary of effect sizes,  $r$ , can be found in Table 27 for all four experimental studies. Cohen (1988) recommended that an effect size of 0.2 is a small one, 0.5 signifies a medium effect size and 0.8 denotes a large effect size. While in Study 1 (Chapter 7) and Study 3 (Chapter 9), effect sizes of Mann Whitney U tests were mostly substantial, effect sizes were low in Study 2 (Chapter 8) and 4 (Chapter 10).

Despite the fact studies two and four showed small effect sizes, they demonstrated a real effect if not only observed through careful study. Although it was hoped that by using high-stakes situations these cues will be easier to detect and while real effects were found it was not expected that effect sizes would be small. The small effect sizes could be due to the low internal validity of the questionnaire used, as the questionnaire may not be capturing what it set out to measure. If the appropriate questions were asked that actually measure the scale of what it was supposed to measure, perhaps bigger effect sizes would have been found. A discussion of how to further develop the questionnaire can be found in the next section.

The choice of statistical tests in Studies 1 to 4 could also have contributed sensitivity in detecting differences in effect sizes. While non-parametric tests maintain fewer assumptions than their parametric counterparts, they have less statistical power compared to their parametric counterparts. If the statistical power of a test is low, it is more unlikely that the test can detect a difference even if the different was really present. This increases Type II error where a true hypothesis is falsely rejected (Mumby, 2002). Thus, the Mann Whitney U's and Wilcoxon Rank tests conducted in Studies 1 to 4 failed to detect a real effect. As such, effect sizes in these experimental studies may have been under-represented by a non-parametric statistical procedure (Price, 2000).

**Table 27**

*Summary of effect sizes for significant implicit cues across Studies 1 to 4*

<b>Study</b>	<b>Test</b>	<b>Variables</b>	<b>Effect size</b>
<b>1</b>	Wilcoxon's Rank between High and Low Emotionality conditions	Credibility	-0.24
		Plausibility	-0.34
		Story Sense	-0.20
		Less Emotions	-0.54
		Emotions Match	-0.44
		Meta-emotion	-0.42
<b>2</b>	Mann Whitney U between Audio-only and Audio-visual conditions	Wording	-0.18
		Voice	
		Attractiveness	-0.16
		Meta-emotion	-0.15
<b>3</b>	Mann Whitney U between High and Low Sense conditions	Credibility	-0.12
		Plausibility	-0.50
		Story Sense	-0.62
		Well-thought-out	-0.07
		Emotions Match	-0.36
		Organised	-0.30
		Wording	-0.17
		Meta-emotion	-0.38
		Voice	
		Attractiveness	-0.18
<b>4</b>	Mann Whitney U between Paired and Solo conditions	tab	
		Story Sense	-0.15
		Practice	-0.23

Face	
attractiveness	-0.18

---

## 11.6 Validity across Studies 1 to 4

### 11.6.1 Internal validity

Internal validity is defined as how well the experiments in a study were conducted. This includes how well the experiments were designed, if variables were operationally defined well, the method of measurement of variables, what variables were measured and what were not. Internal validity can also refer to the level of confidence in arriving at the conclusion that the result of the outcome variable was accredited exclusively to the predictor variable and not attributable to confounding or extraneous ones (Huitt, Hummel & Kaeck, 1999).

To measure internal reliability and how meaningful the data gathering processes have been (the internal consistency of the current questionnaire), Cronbach's alpha coefficient for each scale was calculated for each study. An acceptable cut-off point for this coefficient is 0.7 (DeVellis, 2003). Table 28 shows the coefficients for each scale item as delineated in Chapter 6 where question items were combined together into separate facets. All items in each facet appear to be in the same direction.

The Cronbach's alpha for the facet of Appearance and Emotions were at an unacceptable level for all studies. This suggests a flaw in the choice of question items for these facets, drastically reducing scores and differences in implicit variables and helps explain low effect sizes mentioned above.

**Table 28**

*Cronbach's alpha coefficients for each scale items*

<b>Studies</b>	<b>Meta-emotion</b>	<b>Cognition</b>	<b>Appearance</b>	<b>Emotions</b>
<b>Study 1</b>	0.74	0.47	0.30	0.54
<b>Study 2</b>	0.56	0.35	NA	NA
<b>Study 3</b>	0.74	0.62	0.49	0.35
<b>Study 4</b>	0.63	0.58	0.39	0.37

These differences could have been better captured with questions that actually measure the scale of Appearance, Emotions and Cognition of source content. This also posits that differences actually due to treatment conditions (independent variables for each of the four studies) could be much larger than the levels found in the present thesis. It can also be seen from Table 28 that the alpha coefficients for the facet of Meta-emotion is acceptable for Studies 1 and 3, however questionable for Studies 2 and 4. Therefore, it appears that in Studies 1 and 3 the measure of judges' meta-emotions were valid and justifiable. Concurrently, it was previously discussed that Studies 1 and 3 showed relatively acceptable effect sizes that were higher than those displayed in Studies 2 and 4. Nonetheless, all results should be construed with caution rather than leniency, and err towards a more stringent interpretation.

### **11.7 Discussion**

One of the main challenges in constructing a questionnaire to measure implicit cues lies in the fact that it is a definition that is complex to capture, with a lack of previously identified perimeters and limits to the concept. It is acknowledged that to know where on the continuum of systematic to automatic these items are situated in terms of cognitive processing is an inherently difficult task. Fundamentally, implicit cues are inherently difficult to quantify. Researchers in the field of lie detection have acknowledged that certain variables or cues remain impossible to be captured objectively such as the variable of 'how forthcoming a sender seems to be' (Granhag, 2006). As established, it becomes problematic to claim high internal validity when it cannot be said with confidence that the difference that effects of outcome variables were due strictly to the causes (Validity in scientific investigation, 2016).

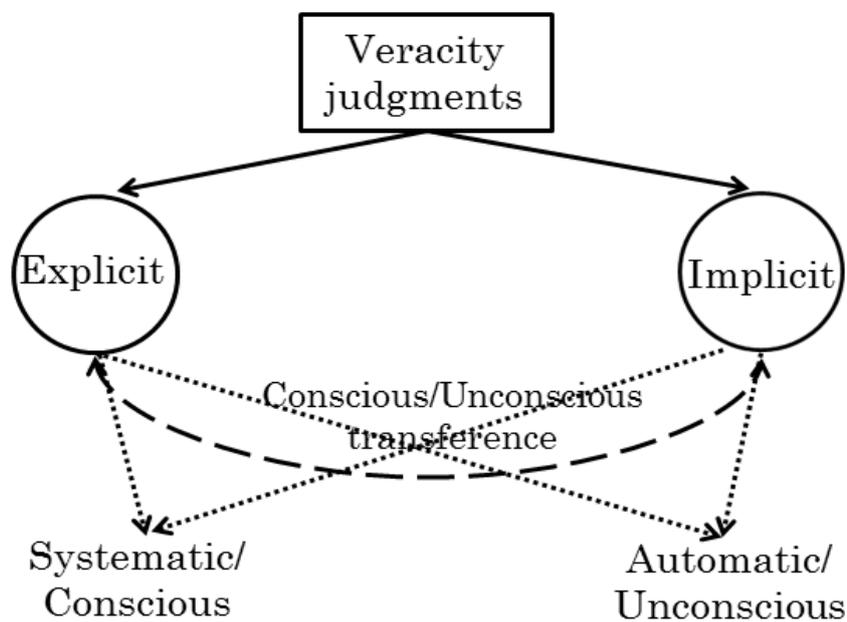
That said, a probability exists that sometimes, explicit judgments may be arrived at independently of conscious implicit judgments (i.e. Study 3), and as revealed in Anderson et al. 's (1999) study. There may also be a conscious or unconscious transference between explicit and implicit judgments but at times, or for certain senders/statements these two do not associate (and not because

judges cannot access this information even when asked directly, or that the questionnaire is not capturing cues that judges may have used). To elaborate on this, there exists a possibility where judges could be more accurate on a lot more implicit judgments, however, this does not transfer into their explicit decisions and the same can be said vice versa. They may achieve high accuracy in their guilt-innocence judgments but either 1) cannot provide an implicit reason or a cue as to how they arrived at this judgment, 2) not understand the meaning of their own assessments (Anderson et al., 1999), and/or 3) at times there may simply be no association between explicit and implicit decisions. Judges may detect an implicit cue but they may not have actually used it to make their explicit decision, or even reason to do so. Judges can detect a valid and diagnostic cue but not be convinced enough by it to expressly judge the sender accurately. This highlights the potential division between explicit and implicit judgment processing, where on particular occasions judges cannot or do not identify implicit cues, however, can make a correct explicit judgment effortlessly (or vice versa). At this point, we cannot know which of these reasons are true until we either 1) ask judges directly, as the current questionnaire may not be capturing their thinking patterns and process very well 2) rule out the unconscious transference between the explicit and implicit decisions *and* the unconscious decision-making process that potentially guides certain implicit decisions, or 3) allow judges to arrive at both types of decisions by themselves at their own pace. What could be tentatively concluded at this stage is that there exists a possibility of a relationship between the explicit and the implicit processes. Figure 2 presents a workable visual representation of the complex relationship between the variables of explicit and implicit judgments and the associated underlying mental processes from the current understanding and construal from Studies 1 to 4.

The results of the experimental studies in the present thesis would seem to indicate that response biases themselves show little indication of being pre-selected. While the present thesis started off with the presumption that response biases are often errors in judgment, it becomes clearer from the results obtained that these responses are adaptive decisions (Gigerenzer & Selten, 2001; Street,

2015). Results from Part 1 challenged the ubiquity of any default response biases (i.e. truth bias claim) (Levine, 2014; McCornack & Levine, 1990; Levine & McCornack, 2001) as well as the presence of an innocence or guilt bias.

It should also be acknowledged as a limitation that while participants were given instructions before they began watching the videos in all studies, a distinction between credibility and veracity would be useful to be explored further and delineated in future studies as participants could have admittedly confused one term for the other.



**Figure 2**

*Relationship between explicit and implicit judgments and the associated underlying mental processes*

### *11.7.1 Suggestions for improvements in questionnaire content*

A component that was clearly lacking in the current questionnaire was non-verbal cues. While a select few questions fell into this category, particularly the facet of Emotions (i.e. ‘The appellant’s emotions do not seem to match their appeal’ and ‘The appellant displays less emotions than appropriate’, more questions pertaining to non-verbal aspect must be included. Although the choice was made

to include only two question items relating to Emotions to reduce the number of variables subjected to regression analyses and to counter for maturation effects so as to shorten the overall length of the study, ultimately the number of items in this facet was perhaps too low. It is recommended that items with higher Cronbach's alpha coefficients be kept for future construction of an improved questionnaire and more questions be added pertaining to the facets of Emotions. A good basis is Ten Brinke and Porter's (2011) study where the authors found several emotional expressions to reliably distinguish truthful appealers from deceptive ones. Truthful appealers appeared 'distressed' and 'sad' whereas deceptive appealers showed more 'disgust'.

The same is said for the facet of Appearance. For example, tone of voice can be included in addition to voice attractiveness, as research has shown that it is a good source of observing leakages in comparison to the face as it is more difficult to control (Zuckerman et al., 1982). Reinhard and Sporer (2010) included the items '(The sender) is very attractive'/ 'beautiful'/ 'healthy-looking'/ and 'likable' in their lie detection study, with a Cronbach's alpha of .73. While the item 'attractive' is already in the current questionnaire, the items 'healthy-looking', 'likable' and 'good-looking' could be added to future questionnaire.

Initial reliability analysis and confirmatory factor analysis (CFA) are recommended in the early stages in future studies (Yale, 2013). In the present thesis, it was not feasible for Cronbach's alpha to be conducted after each study as data for all four studies were collected at the same time, and thus improvements to the questionnaire could not be made in time.

As seen in Table 28, Cognition of Source content displayed a varied range of alpha coefficients; all unacceptable values. The alpha coefficient was the highest for Study 3, presenting a potential for question items in this facet to be reconstructed. The items included in this facet were very similar to those included in Reinhard and Sporer's (2010) measure of statement content perception in their study. Their measure included five items: plausibility, consistency, coherency, structure and specificity, with a Cronbach's alpha of .85. Two out of five of these concepts were measured in Cognition of Source content, namely plausibility and structure (organisation)0

Future question items in this facet could include statement consistency and specificity.

A third category to be kept in mind is that certain variables that fall neither in the verbal nor non-verbal category such as ‘the sender seems forthcoming’ (Granhag, 2006). The focus of the present thesis weighed more towards verbal cues of lie detection, as research shows that verbal cues are more commonly acknowledged as the more diagnostic indicator of lie detection compared to nonverbal ones. There exists a stronger positive relationship between verbal cues and lie detection accuracy (Vrij, Granhag & Porter, 2010). However, in hindsight, non-verbal cues should have also been considered equally as much simply because judges may have relied on them in making their veracity decisions.

It is recognised that there are a much higher number of implicit cues compared to only a single explicit cue. Because including a higher number of implicit cues itself gives them a far higher likelihood for some to return significant values and therefore be more discriminative, this cautions against automatically accepting that these implicit cues are superior. One way to sidestep this issue would be to use a different method in attaining information of implicit cues, which also gives both explicit and implicit cues a more equivalent chance being discriminative. One such method is the think-aloud protocol, where the researcher can still access participants’ thought processes such as what they are thinking and feeling however this would not be predefined allowing participants to arrive at these cognitive processes on their own (Ericsson, 1996; Okada & Simon, 1997).

### *11.7.2 Predictive value of questionnaire*

Predictive validity is how much a score on a scale or instrument predicts and correlates with scores from another test score (Cronbach & Meehl, 1955). While the current thesis did not precisely measure predictive validity using the current questionnaire and compared it with another criterion measure, the predictive *value* of the implicit portion was investigated in relation to the explicit measure of appellant veracity. Thus, the results of interest were the predictive

value of implicit responses towards explicit ones within the same questionnaire (not with another criterion measure).

The conundrum of identifying ‘Which comes first - explicit judgments or implicit ones?’ remains. While the answer to this question tends to be found through rigorous literature review of the topic, very few existing research papers have actually explored the explicit/implicit connection. Granhag (2006) argued that certain intuitive decisions are usually difficult to verbally substantiate apart from expressing ‘It was a gut feeling’; ‘I went with my hunch’. It remains problematic, at least for some judges, to verbally explain certain explicit decisions to implicit detection methods. While efforts have been made to understand this relationship, it has been recognised that to answer this question is a difficult feat.

This topic is not exactly new, though existing research studies tend to isolate explicit and implicit components of lie detection and examine them separately rather than study the link between the two. For example, Vrij et al. (2001) separated explicit and implicit lie detection where one group of police officers were asked to assess video clips explicitly (whether the suspect is guilty or not guilty) and the other were asked to do so implicitly (whether the suspect seemed to ‘think hard’). Other studies that isolated explicit from implicit components in lie detection research include research by Landström et al. (2005), Hurd and Noller (1988) and Anderson et al. (1999).

In the present thesis, the likelihood is that explicit veracity judgments affected implicit ones, as the dichotomous ‘Is the appellant guilty or innocent’ was asked first. That said, judges were allowed to watch the video clips multiple times and changes in explicit answers were allowed (although the number of times appeals were watched and changes in explicit answers were both not recorded). It was observed that many judges did indeed change their explicit answers, some up to several times, indicating possible interplay between explicit and implicit components of processing.

Granhag (2006) stressed that sometimes it is beneficial for the researcher purely to be familiar with judges’ assessments of everything else apart from explicit veracity judgments. The aim of this thesis was never to investigate whether implicit or explicit assessments were the cause or effect, or whether

judges answered implicit cues in the questionnaire consciously or unconsciously, but to discern if a relationship exists between implicit and explicit veracity judgments. Still, to fully comprehend this relationship it is now clear that the conundrum of 'which comes first' (in which situation) cannot be eluded.

Since regression can be used as a prediction analysis, it was thus chosen as the method of analysis in Studies 1 to 4. Rather than regarding the independent variables as having a causal relationship with the dependent variable (herein lies the problem to claiming high internal validity), the aim of this prediction analysis was to develop models that make predictions of judges' explicit decisions based on the observed values of implicit cues.

Table 29 shows which implicit judgments help predict judges' explicit judgments, as well as when these explicit decisions are accurate (Table 30). Implicit judgments used differed between experimental conditions, accounting for individual variability and most likely due to distinctiveness of cue saliency in the appeals shown. The regression results from four consecutive studies showed that there exist correlations between the dichotomous dependent variable of explicit judgments and the continuous independent variables of implicit decisions, revealing different thought and encoding processes in different experimental conditions.

Many of the cues judges used across Studies 1 to 4 were *not* diagnostic of *accurate* explicit decisions. This implied that inaccuracies in cognitive processes occurred at a high rate. For example, there is a likelihood that judges used the salient cue of emotional display to make innocence-guilt judgments in Study 1, whereas a closely related cue with higher validity in predicting the innocence of an appellant would be spoken emotional words instead (see Chapter 12).

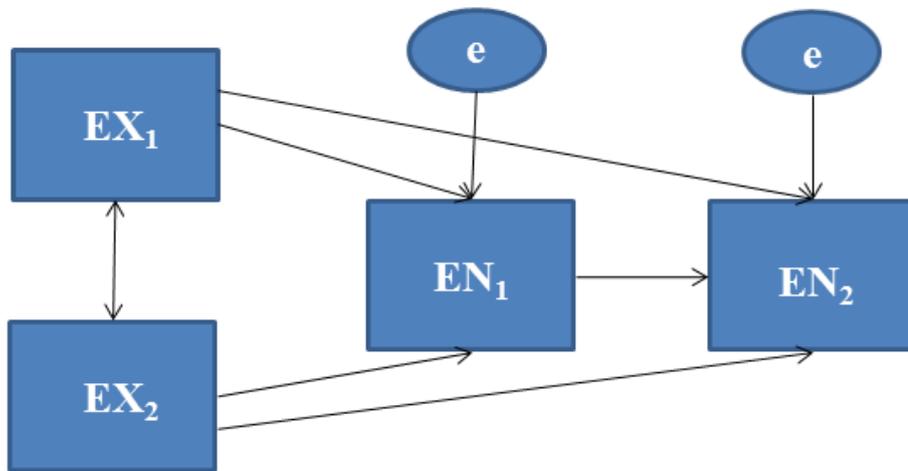
### *11.7.3 Suggestions for future analyses*

The present thesis postulates and indicates, to some level, this link between implicit and explicit components through the combination of regression analysis, Mann Whitney U tests, Wilcoxon Rank tests and chi square throughout Studies 1 to 4. The treatment conditions in each study in Part 2 showed differences between means and variances in veracity judgments. The next step in future research

would be to know the *extent* of this relationship (thereby subsequently solving the internal validity issue, or at least reducing its threat). Of course, there may also be other extraneous factors that contribute to change in the outcome variable. Confounding factors and alternative causes stand in contributing to this change (Huitt et al., 1999).

To account for these extraneous variables, different methodologies for analysis must be considered in future studies. To capture the complexities of extraneous variables and the intricacy of the explicit/implicit relationship, a statistical method that may be better able to capture this concept is path analysis. While regression analysis was useful in the present thesis to highlight the relationship between the two, path models will allow a stipulation of a model based estimates of the magnitude and significance of hypothesised causal connections between sets of variables (Suhr, 2008). Path analysis may also elucidate the magnitude of which explicit lie detection contains an implicit part, and the likewise the level of implicit lie detection comprising an explicit portion when examined in a case by case basis. In fact, this may not be the case at all. In some cases, perhaps only explicit processing takes place and in some others only implicit processing occurs. Path models will be useful in ascertaining the direct and indirect effects within this complex relationship (Stage, Carter & Nora, 2004). In the sample path diagram found below in Figure 3, exogenous variables  $EX_1$  and  $EX_2$  are shown as covariates. They both have a direct and indirect effect through  $EN_1$  on  $EN_2$  (both endogenous variables).

One of the benefits of path models is that alternative models are conceivable. For example, one variable may not have a direct effect and only an indirect one. The goodness of fit of these models can then be compared statistically (Wright 1921; Wright, 1934). Disturbance terms take into account random errors and measurement errors, which refers to the difference between an observed value and its true value (Dodge, 2003; Taylor, 1999).



**Figure 3**

*Example of a path model with exogenous variables, endogenous variables and disturbance terms*

#### 11.7.4 Advantages of the current questionnaire

While the current questionnaire used in the present thesis holds limited internal validity, there were also several advantages. One of these advantages has already been discussed at length in the section of ‘Predictive value’ that can be found above.

Another clear advantage of the current questionnaire was the inclusion of both object-level and meta-level dimensions in the measure. According to Granhag (2006) and Nelson (2002), on the object-level dimension implicit decisions are the product of a Judge’s perception of the sender, such as ‘the appellant sounds like they have had a lot of practice’. On the meta-level, these decisions are a product of their reflection on their own cognitions or emotions. DePaulo, Charlton, Cooper, Lindsay and Muhlenbruck (1997) have also included meta-level questions in their study, where they reported that participants have a higher confidence level when judging truthful statements compared to lies, even though the correlation between their confidence level and actual accuracy in judging both truths and lies were low.

Secondly, all operational definitions used in the current questionnaire were based on careful literature review on theories that go beyond lie detection research. Considering the suggestion that certain implicit cues are difficult to be linked with theoretical approaches available in lie detection research (Granhag, 2006), every effort was made to link each question to existing theories, all of which are elaborated at length in the introductory chapters.

A joint advantage with exercising a standardised questionnaire is the ability to control and curtail the parameters of inclusion (as well as preclusion of misapplications) of what comprises an 'implicit' measure. For example, a participant using the think-aloud method could name 'thinking hard' as one of the cues they used to arrive at their explicit veracity decision. Granhag (2006) recommended that the cue of 'thinking hard' should be excluded from the parameters of implicit deception detection as it categorically lies within the cognitive load framework. Careful selection of questions to be included in the current questionnaire precludes inclusion of criteria that misrepresent implicit mechanisms deception detection.

Thirdly, while a wealth of information was inaccessible as an open-choice questionnaire was not adopted (such as think aloud protocols), using a standardised questionnaire allowed standardisation of answers. The advantage of using a standardised questionnaire was that it was a relatively quick and practical way to gather a lot of information from a large sample in a short duration. Using a standardised questionnaire reduced subjectivity and data handling time during coding and analysing.

Using a questionnaire also mitigated the retrieval problem as Anderson et al. (1999) suggested. The authors advised that participants could have difficulty in either the retrieval of the memories of the cues used or in the reporting of these cues. For example, it is possible participants retrieve memories of a sender's verbal cues comparatively more so than recollections of the sender's visual cues, even if they actually utilised the visual cues more to make their ultimate veracity judgments of the sender. As follows, the predictive value of implicit questions was useful as it allowed a prediction of which cues related to judges' explicit decisions

whether it was used knowingly or unknowingly. Anderson et al. (1999) cautioned against discounting the likelihood and complications of difficulties in cue retrieval.

Another benefit of the current questionnaire was that a 'not sure' option was also allocated for participants to choose. This option allowed judges to select that they were undecided on their perception of a specific implicit question item. This can generate a higher degree of accurate data and increase the number of definite responses. By allowing this option, the researcher has the option of retaining only those responses by participants who were very certain of their responses to achieve more meaningful results and the subsequent interpretation of results.

Yet another advantage of the current questionnaire's brief length was that this possibly countered against maturation effects that could have further jeopardises the experimental studies' internal validity. Internal validity is threatened when participants perform better or worse as a result of maturation and not because of the predictor variable (Isaac & Michael, 1971). Maturation effects refer to any naturally occurring circumstances resulting from time elapsing by, such as boredom, tiredness and lack of motivation to complete the questionnaire. The brief nature of the questionnaire ensured that judges did not take more than 45 minutes to complete it.

## **11.8 Conclusion**

The absence of a consistent innocence bias in Studies 1 to 4 showed more support for the ALIED theory over the Spinozan rationale (Gilbert et al., 1990), with the former inferring that judges make informed judgments based on experience and make educated guesses in the face of low or un-useful information. This informed deduction was a source of response biases that surfaced (Street, 2015).

It is acknowledged that a truth bias is not synonymous with an innocence bias and the inter-relationship between the two biases needs further study. Further investigation into how people process, encode and manage information and how that translates to them making a guilt-innocence and/or a truth-lie judgment can either support these initial suggestions or challenge them. As

reiterated throughout this thesis, it is very likely people will consider impressions of truth-lie in some form en route to making an innocence-guilt judgment of a person, especially in the current context where judges receive very little of other evidence of the case and information about the appellants. However, at this point in time this possible interrelationship and their nuances are yet to be researched and further developed.

What is interesting is that, for example, Study 3 points to people picking up different cues for guilty and innocent appellants and awarding different ratings of innocence and guilt for each. Thinking processes seem to be different for when a sender is guilty or innocent. Results from Study 2, while needing further trials and replication, also seem to infer that there are differences in the thinking process depending on how the Source is presented (i.e. audio-only versus audio-visually), and this could also affect explicit judgments. If this conjecture is true, this has important bearings on how television appeals are presented and aired, as how they are presented can elicit or hinder these biases which result from 'rules of thumb' that people possibly can rely on.

While the implicit portion of the questionnaire may lack internal validity, explicit veracity judgments were not affected by the questionnaire as it was a standalone question. Additionally, chi square results infer that there are associations between manipulated conditions and explicit veracity assessment accuracy across the board from Studies 1 to 4. A possibility for moderate ratings of innocence-guilt judgments is that editing of video clips in certain conditions may not have been very elegant and polished, and judges may have detected this. For example, it could have been noticeable to judges that video clips in the Low Sense condition in Study 3 were modified due to a rearrangement of video frames. Their responses may have been influenced accordingly, again contributing to a weak effect size in the relationship between conditions and explicit accuracy ratings. It is also possible that there could be simply other factors that would have contributed to larger effect sizes other than the ones manipulated and tested across Studies 1 to 4.

An alternative likelihood is that these factors tested across Studies 1 to 4 would have always shown a weak significance when considered on their own.

DePaulo et al. (2003) found that out of 158 cues they examined most cues, when considered independently, showed either no association or only weak ones to deception. Of recent the need for a multiple cue approach to lie detection has been recognised (Whelan, 2014; Hartwig & Bond, 2014; Burgoon, personal communication, 25<sup>th</sup> August, 2015; ten Brinke & Porter, 2012; Vrij & Mann, 2004). Hence, the possibility exists that considering multiple factors or cues tested across Studies 1 to 4 collectively would yield higher accuracies than recorded in the present thesis.

Reliable verbal indicators in the specific context of television appeals are still yet unknown. In light of this, the next chapter sought to investigate multivariate and reliable indicators of innocence and guilt in television appeals.

**Table 29***Significant implicit judgments used by judges across Studies 1 to 4*

<b>Condition</b>	<b>Sense</b>	<b>Credibility</b>	<b>Plausibility</b>	<b>Practice</b>	<b>Well-thought-out</b>	<b>Less Emotions</b>	<b>Emotions Match</b>	<b>Face</b>	<b>Voice</b>	<b>Organised</b>	<b>Wording</b>	<b>Meta-emotion</b>
<b>High Emotionality</b>												
<b>Low Emotionality</b>												
<b>Audio-only</b>					<b>0.001</b>							
<b>Audio-visual</b>					<b>0.000</b>				<b>0.000</b>		<b>0.004</b>	
<b>High Sense</b>		<b>0.000</b>			<b>0.004</b>							
<b>Low Sense</b>												
<b>Paired</b>							<b>0.001</b>		<b>0.000</b>			
<b>Solo</b>		<b>0.000</b>			<b>0.001</b>							



## PART 3

---

### OBJECTIVE VERBAL CUE ANALYSIS

## CHAPTER 12

---

# OBJECTIVE VERBAL INDICATORS OF TRUTHFUL AND DECEPTIVE TELEVISION APPEALS

As expounded in Chapter 3 verbal content analysis shows much potential in discerning lies from truthful statements. The current study tests the hypothesis that verbal indicators differentiate high-stakes true from false statements. This study also draws from the wider grief literature as the basis for a content analysis of the television statements. Hypotheses concerning different linguistic markers for false from true appeals are drawn from literature on the underlying psychosocial and emotional changes which accompany it and its manifestation in language. Furthermore, as expounded in Chapter 3, the examination of verbal cues in relation to each other is also lacking in existing lie detection research; the present study seeks to employ a methodological approach that will include even cues with very low frequency counts in analyses.

### **12.1 Verbal cues to deception in high stakes conditions**

As Vrij et al. (2010) have pointed out, whilst training manuals emphasise non-verbal indicators for detecting deception, a growing body of research indicates that verbal cues can distinguish liars from truth tellers (Harpster et al., 2009; ten Brinke et al., 2012; deTurck & Miller, 1985). Indeed, participants who rely solely on nonverbal cues in making their veracity judgments have been shown to achieve a lower accuracy rate than those who use verbal cues (Burgoon et al., 2008; DePaulo & Morris, 2004).

### **12.2 Language aspects of grief and related emotional states**

The crucial emotional state of television appeals relates to the distress and grief the appellant is (apparently) feeling when making the appeal. This therefore offers a context for considering potential cues to deception.

Pennebaker et al. (2003) demonstrated that the content of speech is contingent upon the emotional state of the speaker; when an individual is grieving, there is a tendency to communicate in a particular manner. Stress, trauma and psychological disruptions can affect how one speaks, and the language chosen, whether consciously or unconsciously. Words hold the capability to reveal and expose cognitive and social constructs of a speaker (Pennebaker et al., 2003; McAdams, 2006). Tausczik and Pennebaker (2010) report that the language a

person uses is linked to cognitive, personality and social constructs within an individual at that point in time.

Content analysis of linguistic markers is not a new method and has been in development for decades, as in the study of Gottschalk and Gleser (1969) where samples were transcribed and sentences segregated into phrases. These phrases were assessed to reveal thematic content reflecting the speaker's emotional, psychological and cognitive conditions. The story an appellant chooses to verbalise, or omit to verbalise, may point to their psychosocial and emotional preoccupations at that point in time. The manifestation of these preoccupations after the murder or disappearance of a significant other in language, of whether finding the victim is at the foremost importance to the appellant or the cover-up of the crime, gives rise to hypothesised linguistic markers of true and false appeals. Thus, it is hypothesised that what is said in television appeals will indicate whether finding the victim is at the foremost importance to the appellant (with associated grief and emotional states), or the cover-up of the crime. The present study investigates the underlying preoccupations of an appellant. It is this which may provide the clue as to who is guilty and who is innocent.

### *12.2.1 Language in assessing psychosocial and emotional changes post-loss*

Grief produces psychological, emotional and social change within an individual (Vance et al., 1994) which manifests in their language (Pennebaker et al., 2003). These changes include altered relationships with society and changes in roles within the family of the victim. A study by Sprang, McNeill and Wright (1989) examining the mental health changes of those who experienced a violent death of a significant other reveals reactions that accompany grief include shock, denial, isolation, emotional release, guilt, anger, resentment, depression, acceptance, resolution and adaptation. The authors report that socially, the violent death puts forward changes in religious beliefs and social support systems they depend on post-loss. In a murder or missing person case, the relative who has recently suffered a loss due to sudden, violent circumstances will indeed undergo changes in their emotions, whether it is anger towards the killer or abductor, pain due to the inability to accept that the victim is no longer present or depressive

episodes. This in turn, will affect the language that they use to speak, which underlines their foremost anxieties and upheavals. Literature on reactions post-loss has never been paralleled and referenced to cases of television appealers in available studies, which hold vast potential for comparison and association.

### **12.3 Language areas in assessing changes post-loss**

#### *12.3.1 Changes in psychological intimacy*

Studies indicate that a typical response to grief is a preoccupation with the loss or separation from the victim. For instance, Maciejewski, Zhang, Block and Prigerson (2007) report that focused yearning for the dead person was the most frequent negative psychological response. Klass's (2006) theory of continuing bonds with the bereaved, and Field and Filanosky's (2010) externalized expressions, such as delusions that the deceased is alive, also both point to the griever still being attached and wanting to be connected to the bereaved. Shuchter and Zisook (1988) write that griever sometimes experience visual and sometimes auditory hallucination such as dreaming of the deceased, looking out for them in a crowd, and continuing to speak with them, striving not to sever relationship with the bereaved. Consequently, in the current study it is possible that if a televised appealers is truly grief-stricken, thoughts of the victim's existence will dominate in what they say even to the point of envisaging the victim's presence.

In addition to (or in place of, if victim is no longer known to be alive) thoughts of the victim's presence, reminiscence about the deceased's death is common (Rosenblatt & Elde, 1990). This involves talking about personal memories of the griever. Keeping count of anniversaries such as birthdays, death days, divorce date, or the day they went through a loss is also frequently part of the grieving process as Engel (1975) noted. A genuine appealers would thus potentially incorporate into the appeal a count of days since the victim disappeared or was murdered.

#### *12.3.2 Changes in psychological distancing*

In contrast to wanting psychological intimacy, someone who must feign a loss is hypothesised to speak less of the victim they murdered or kidnapped may

instead focus either on pleading their supposed innocence or distancing themselves from the victims. Harpster et al. (2009) posit the cue of ‘minimising’, where fake 911 callers use words like ‘just’ to diminish their connection in and distance themselves from the circumstance. Davis et al. (2005) report that liars in their sample tended to utter ‘I don’t know’ more often than truth-tellers. Ten Brinke and Porter (2011) suggest that this is an effort by liars to distance themselves psychologically or reduce the need to develop invented accounts. They found that liars tend to use more tentative words compared to truth-tellers.

A reference to others is a potential cue for a dishonest appeal (Knapp, Hart & Denis, 1974). Stirman and Pennebaker (2001) report that suicidal and depressed individuals refer to others less and tend to self-focus. Harpster et al. (2009) indicate that in 911 calls a deceptive caller tends to shift blame to the victim or someone else. Additionally, a true applier may use less conditional words when talking about a missing victim coming home, because to them this is a direct wish. In contrast, a false applier would insert more conditional markers, including conjunctions, when referring to the victim coming home, because presumably they are aware that is not a possibility.

### *12.3.3 Changes in emotional release or reserve*

Besides psychological distancing or intimacy, changes in emotions can also be studied via language in high stakes lies. Evidence comes from various research efforts from lie detection and grief literature. CBCA’s Criterion 12, ‘subjective mental state’, is defined as when a person expresses what they are thinking and/or their emotions during the event, in their statement (Granhag, et al., 2015). The Reality Monitoring (RM; Sporer, 1997) criterion of ‘affect’ and SCAN (Sapir, 2005) criterion of ‘emotions’ are related to Criterion 12.

In the grief literature, Freud (1917) asserts that a bereaved person has to work through their grief by reviewing thoughts and memories of the deceased. Shear and Mulhare (2008) state that acute grief, the initial reaction to loss, manifests in emotional symptoms such as meaninglessness, emotional lability and apathy.

Whelan et al. (2014), using high stakes video footage of public appeals, find that honest appealers refer to norms of emotions or behaviour which includes violations, such as ‘How could anyone do this?’ Ten Brinke and Porter (2012) found that honest appealers are more emotionally positive than deceptive ones, however the authors did not include emotional lability or apathy in their classification. They find that honest appealers express more pain than deceptive ones. Whelan et al. (2014) also discovered in their study that honest appealers expressed more hope in finding their missing relatives in cases where the body has not been found yet, as unlike a false appealers who murdered the victim they do not yet know that whether the victim is alive or otherwise.

The mention of expletives should be considered as well. Sprang et al. (1989) state that individuals typically report anger and resentment towards the killer whose family member has been murdered. Kubler-Ross’s (2005) grieving model includes anger as one of its stages. Kitson and Zyzanski (1987) find that widowers in her sample reported both anger and grief. Those who experienced an unexpected death experienced higher levels of anger. Research has also identified that when an individual is angry they usually care less about being polite because their preoccupation is to seek the victim’s return or find the killer. Harpster et al.’s (2009) report that some 911 callers did not care to sound polite during the emergency call and swore on the phone (e.g., “What the f\*\*\* is taking them so long!”). Politeness is seen as a cue to hoax callers, because a genuine caller’s paramount objective is to seek immediate aid. However, Van Swol et al. (2011) find seemingly contradictory evidence that on average, liars used more swear words than did truth-tellers especially in cases where the recipients voiced suspicions. The seemingly different results could potentially be related to the high-pressure of the situation in Harpster et al.’s (2009) study versus the lower-stress situation in Van Swol’s et al. (2011). The former is an unexpected, high-anxiety situation that does not allow for much preparation. The latter is less so, no life being at stake. Because television appeals are typically high-pressure situations in which the respondent has some preparation it is proposed that it is false appealers who will swear more in an attempt to simulate emotion.

#### *12.3.4 Social changes*

The death or disappearance of a loved one will also produce changes socially, such as turning to a higher agency and going to prayer locations to buffer the negative event, and this in turn may possibly be captured in language. For example, it has been shown that high use of positive religious coping may safeguard the deleterious effects of the situation (Bjork & Thurman, 2007). While not all truthful appealers are religious or at least spiritual, deceptive appealers may find it hard to turn to a higher power or may not be aware of the behaviours of a truly grieving individual.

This, again, may possibly be observed in the language used. Stone and Pennebaker (2002) observed that during and immediately after a disaster, individuals gather together and cooperate. After the death of Princess Diana, the use of 'we' instead of 'I' increased by 100% in online chat groups for a period of a week. Cohn, Mehl and Pennebaker (2001) found that people were more socially engaged and communally involved in the immediate aftermath of the September 11 tragedy.

From lie detection literature, Bond et al.'s (2014) meta-analysis report the indirect measure of 'cooperation' to be the strongest determinant of accurate lie detection. Inbau et al. (2001) posited that liars are not as cooperative compared to truth-tellers. However, in Vrij's (2005) experiment, the author did not find that cooperation was related to lying or truth-telling. This may be because of cooperation was measured as participants' inclination to give an account of what transpired during the incident, which is rather different from the current study's interest.

In the present context, the family of the victim may rely on the local community for support in finding the body or in finding the abductor if the victim is missing. They may use the television appeal as a strategy for getting help from community to find their loved ones. While a false appealers may feign converging with the community to gain help, it either may not be to their advantage to do so, or it may not come naturally to them to do so. Social aspects of the language in terms of reliance on others by television appealers warrant being studied.

### *12.3.5 Movement-related changes*

Movement in speech is a potential cue as yet unexplored in television appeals. Verbal activity here can be seen as words that give a compelling sense of motivation, hypothesised as typical of a person who really wants a murder or missing case solved. Genuinely honest relatives would be expected to exude a sense of urgency in their plea (Harpster et al., 2009) in finding the missing person, the body or the killer or abductor. This drive for action may potentially be present in the genuine appellant's speech, meriting a study on movement-related words or phrases in the current context. Vrij and Mann (2004) state that liars tend to be more passive and less involved compared to truth-tellers. In the present study, an appellant for missing or murdered relative instinctively calls for either activation of a target to find the missing person, change the predicament, and find the body if the relative is murdered and/or find the killer or abductor of the relative.

### *12.3.6 Tense changes*

The tense used by appellants may also be an indicator of how truthful they are. A false appellant who has to feign that the victim might still be alive while knowing that he or she murdered them is hypothesised to speak less of the victim in the present tense. Kastor (1994) reports that Susan Smith, who killed her two sons, used the past tense when speaking of them, before it was known that they were murdered and not just missing. Harpster et al. (2009) posit that fake 911 callers tend to accept the death of their loved ones at a higher rate (i.e., pronounce that they are dead) and tend not to be in the denial stage even in the knowledge of the death, compared to true callers. A true caller is less likely to declare their relative's mortality when this relative's actual state is unknown. Even when their relative's mortality is known, the authors state that true callers often deny this knowledge, still wanting to 'save the lives' of their relative.

Ebesu and Miller (1994) who observed college students enacting mock-interview scenarios report that their participants acting as liars used less present tense words. However, in a rather different study Dulaney (1982) show contradictory results. Lying participants in this study utilised fewer past tense words. These differences in results may be a function of context and how the

experiment was constructed methodologically. In Dulaney's (1982) study, the nature of the lie was 'deceit by omission', while in Ebesu and Miller's (1994) study participants told a direct lie. Participants could either choose to inform the experimenters whether they cheated or not and those who did not were considered liars. This is a prime example of the importance in examining context-specific cues where while the cue itself may occur in general, the direction in which the cue emerges given a specific context may change. In the present high-stakes situation, deceptive appellars may be more likely to use the past tense regarding the victim, regardless of whether it is known that they are missing or murdered.

### *12.3.7 Grammatical changes*

The attempted control approach posits that liars may attempt to control their behaviour to prevent deceptive cue leakages (Vrij, 2004). Ironically, by its very nature this attempt can give rise to deceptive cues (DePaulo & Kirkendol, 1989). The four-factor theory (Zuckerman et al., 1981) and Interpersonal Deception Theory (IDT; Buller & Burgoon, 1994) postulate that liars may leak certain cues when strategically trying to overcontrol their behaviour, particularly in situations demanding higher cognitive load. These cues include increased speech errors in the form of grammatical mistakes (Zuckerman et al., 1981). An increase in cognitive load may cause more speech errors (Goldman-Eisler, 1968). Besides tense changes, grammatical errors may be expected of deceptive individuals. Davis, Markus, Walters, Vorus and Connors (2005), Vrij and Mann (2001) and Whelan et al. (2014) find liars tended to commit higher levels of error in speech, such as grammatical errors and words or sentences that were not complete.

### *12.3.8 Chronological changes*

The form and structure of a story told is another attribute of language of relevance. CBCA hypothesises that statements delivered in a non-chronological temporal order, where the production is unstructured, are inclined to be more truthful (Vrij, 2005a). Vrij et al. (2010) assert that liars tend to give accounts in a chronological and rehearsed order such as 'I did this, then this happened, then I

did this'. Whether or not appellants verbalise events in the form of 'I did this, and then this happened, after that this took place' could potentially suggest preparation of speech in a pre-arranged chronological order, indicating a false appeal.

In addition, DePaulo and Morris (2004) stressed that liars show a tendency towards repetition of words and phrases in the same sentence more so than their truthful counterparts. They concluded that the reason for repeating words and phrases were to stall for time as liars undergo a cognitive load when being questioned or interviewed. Harpster et al. (2009) also found deceptive 911 callers resorted to repetition significantly more so than truthful callers.

#### **12.4 The current study**

These possible changes in language discussed serves as areas to explore, with televised appeals presenting a suitable platform to do so. In the present study based on the theoretical frameworks discussed above, it is argued that distinctive linguistic markers, stemming from differences in the underlying preoccupations of a truthful versus deceptive appellant, can provide indications as to who may be truly grieving (a true appeal) from a false appellant. Psychological, social and emotional changes after the murder or disappearance of a significant other are theorised to affect the language of appeals, indicating their preoccupations of either finding the victim and pushing forward an investigation or hiding the truth to get away with murder. Dissimilarities between genuine from false narratives revealed in and by these spoken appeals are examined, alongside differences social and affective processes or preoccupations of the appellants. Based on previous research efforts, it is hypothesized that certain variables will gravitate towards genuine appellants, and likewise for false appellants.

#### **12.5 Methodology**

##### *12.5.1 Developing content variables*

To reiterate, eight language areas were reviewed in the literature as presented above. These pertained to speech content, form and structure to comprehensively investigate the underlying preoccupations of a television

appealer. To test and measure the hypotheses concerning appeal veracity with these language areas, it was necessary to generate criteria for classifying each appeal as true or false. Hence, a dictionary of words based on the literature, hypothesised to distinguish true from false appeals will be selected and each of them expounded in the following section. A summary of the dictionary can be found in Appendix M and empirical support for each area from existing literature on grief and lie detection can be found in Appendix N. Considerable justification directed the decision to select the following sixteen variables under the eight language areas drawing on studies of verbal aspects of grief and emotion. These variables were selected to specify a sample from the different language areas that may unearth the underlying preoccupations of appealers, and therefore provide clues in distinguishing a true appeal from a false one. Each language area contains distinct content categories to be simultaneously analysed. These variables serve as measurable outcomes of either a desire for intimacy with and finding the victim or distancing and the cover-up of the crime.

### ***Area 1: Psychological distancing***

As shown in literature, suicidal and depressed individuals were less prone to reference people other than themselves, again self-focusing rather than other-focused whereas lying individuals are slow to own any statement in an attempt to reduce any possibility of personal culpability (Harpster et al., 2009; Knapp et al., 1974; Stirman & Pennebaker; 2001). As such, references to others are hypothesised as a potential cue for a dishonest appeal.

To measure psychological distancing, the presence of words, phrases and/or any reference to 'somebody' or 'someone else' including the attempts to lead culpability from the self, such as 'I don't know' or 'I didn't do' will be included in the dictionary (Granhag et al., 2015; Davis et al., 2005). This is expected to capture self-distancing from any activity that can be done to find the victim. To avert committing to a lie, liars tend to use tentative language such as 'maybe' and 'perhaps' (ten Brinke & Porter, 2012; Newman, Pennebaker, Berry, & Richards, 2003; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Another way to self-distance language is through using conditional markers (Harpster et al., 2009).

Thus, the presence of words depicting conjunctions such as ‘if’ and ‘but’ will be recorded. Here, a true applier is hypothesised to use less conditional words specifically when talking about the victim coming home if they are missing, because to them, there are no conditions or repercussions regarding their desire for the victim to return home. In contrast, a false applier would probably insert more conditional markers when wanting the victim to come home.

### ***Area 2: Psychological intimacy***

Arguably, expressing phrases and/or any reference of wanting to be close to the victim is the measurable outcome of a desire for psychological intimacy. Based on the literature introduced earlier which includes Maciejewski et al. (2007) findings on focused yearning, Klass’s (2006) theory of continuing bonds with the bereaved, Field and Filanosky’s (2010) and Shuchter and Zisook’s (1988) findings on hallucinations connected to the deceased, the hypothesis here is that more truthful appliers will include such words and phrases as a manifestation of wanting to be closer to the victim, compared to deceptive appliers. These include variables such as imagining the victim, for example, which captures the attachment to the bereaved and fixation towards the victim, where the applier fantasises about what victim would be doing if they were with them, imagines what they are going through, or wonder about their condition. True appliers would probably also keep count of anniversaries and/or days since their relatives went missing or was murdered (Engel, 1975), hence the presence of words depicting ‘keeping count’ will be recorded.

### ***Area 3: Social changes***

From the literature reviewed it is known that going through grief may propel social changes in several aspects that lead to changes in speech content. These aspects include, but are not limited to, convergence towards religiosity and a greater reliance on communality. To measure these social changes, the presence of words, phrases and/or any reference to religion and reliance on the community are taken to characterise the manifestation. People use religion as a buffer to cope with a negative event (Bjork & Thurman, 2007). There is also increased

cooperation between people (Stone & Pennebaker, 2002) and increased social engagement (Cohn et al., 2001) in the aftermath of a negative event. In the lie detection literature, liars appear to not be as cooperative as truth-tellers (Vrij & Mann, 2004). Perceived cooperation seems to be one of the most promising implicit cue in detecting deception (Bond et al., 2014, DePaulo et al., 2003; Anderson et al., 1999).

Religiosity was chosen to depict any declaration of faith, the mention of prayer or the declaration of a higher power. Communality applies to the verbal declaration of cooperation with the community and hopes to measure the social change of relying on the local community in helping find the victim and/or killer, as well as the change in reliance on the community for support. While it is possible that false appellers may display a presence of these variables (as they may go through a change in their social activities), it is the hypothesis of this study that the presence of these variables would be indicative of a truthful appeal. As discussed in the literature review this may be due to a natural regression towards these changes post-loss compared to a false appellant, who may or may not have knowledge of how to behave deceitfully in the aftermath of a murder or disappearance.

#### ***Area 4: Emotional changes***

Evidence for references to emotion post-loss comes from various research efforts discussed above. True appellers are hypothesised to verbally declare certain emotions or the lack of it, as the death or disappearance of a loved one sparks either a release or reserve in their emotions and this runs its course in their appeal content, without them having to put on an act or remember to do so (Granhag, et al., 2015; Sporer, 1997; Sapir, 2005). It is expected that the same pattern will emerge when considering the language of television appellers as indicated by Whelan et al. (2014) who found that truthful appellers mentioned violation of emotional norms, uttered more hope and affection towards the victims and ten Brinke and Porter (2012) who found that honest appellers are more emotionally positive than deceptive ones. In the present study, rather than measuring violations of emotional norms or valence, the present researcher seeks

to quantify the presence of words, phrases and/or any reference of emotions, the lack of emotions, the utterance of a lack of emotion, states of mind and utterances of affection towards the victim as a measure of possible emotional changes. It is hypothesised that these include measuring the presence of spoken emotions, lack of emotions, hopefulness and missing the victim.

In contrast, it is hypothesised that more false appealers will include expletives in an attempt to simulate emotion (Harpster et al., 2009; Van Swol et al., 2011).

### ***Area 5: Changes in movement***

A drive for action or the lack of may potentially resonate in an appealers' speech. Vrij and Mann's (2004) paper suggested that liars appear to be more passive and Harpster et al.'s (2009) research reported that genuine 911 caller would exude a sense of urgency in their plea for help. Similarly, in this study it is expected that a genuine family member will try to actively and urgently move an investigation forward and this drive for action may potentially be present in the genuine appealers' speech. To quantify movement, 'activity' was chosen to measure any declaration of motion-related words or phrases suggesting activity, direction and a desire for change. It is hypothesised that more true appealers would mention more progressive phrases or words such as "If anybody knows anything, let us know", "forward", "call for search" and "looking", compared to false appealers.

### ***Area 6: Changes in tense***

Lie detection research found that liars tend to use different amounts of past tense words compared with truth-tellers, depending on the context of the study, inferring that this area is worth exploring in terms of the current context of this study particularly concerning the missing relative (Dulaney, 1982; Ebesu & Miller, 1994; Harpster et al., 2009). To measure this, the presence of present tense when talking about the victim hopes to quantify the tense used by the appealers in the phrase or sentence regarding the victim specifically. It is hypothesised that

more truthful appellers will use present tense words in this context compared to false appellers, regardless of whether the victim is still alive or otherwise.

### ***Area 7: Changes in grammar***

Because a change in grammar may also be expected (such as infelicities and errors) of deceptive individuals (Davis et al., 2005; Whelan et al., 2014; Zuckerman et al., 1981; Goldman-Eisler, 1968; Vrij & Mann, 2001; Whelan et al., 2014) perhaps due to increased cognitive load on maintaining the deception, the presence of grammatical error(s) is hypothesised to appear more often in deceptive appeals compared to truthful ones.

### ***Area 8: Changes in form and structure***

Statements delivered in a non-chronological temporal order, where the production is unstructured, are inclined to be more truthful (Vrij, 2005a; Vrij et al., 2010). The form and structure of an appeal may be another area of language that may indicate appeal veracity. The presence or absence of whether the appeller refers to the event which took place is predicted to quantify this pre-meditated rehearsal. This is to test the hypothesis that a higher number of false appeals will incorporate this type of appeal structure. In addition, following Harpster et al.'s (2009) definition of repetition, any words or phrases that are used thrice or more in succession in the same sentence will be considered as repetition, excluding 'uhms' and 'ahs'. False appellers are hypothesised to reveal more repetition in their appeals compared to their truthful counterparts (Harpster et al., 2009; DePaulo & Morris, 2004).

All that the appellers said in every appeal was transcribed. Only spoken words, synonyms or phrases that were clearly discernible and were not easily misconstrued were included for analyses. Each of the 17 content categories derived were coded dichotomously to generate variables that indicated their presence or absence in a particular appeal. Previous research using naturally occurring aspects of behaviour in a criminal context has indicated that any distinctions more refined than presence-absence dichotomies are likely to be unreliable (e.g. Canter & Heritage, 1990). While application of content variables was based on a reading

of the transcripts, this was verified by a second coder blind to the ground truth of the appeal videos who completed the variable coding of 10 videos (25%) to assess inter-rater reliability.

Content variables derived from literature, as shown in Appendix M, were subsequently applied to the transcripts. In doing so, none of the appeals contained any mention of the hypothesised cues ‘expletives’ (i.e., none of the transcripts contained any swear words) and ‘repetition’. Thus, both variables were excluded from further analyses. The final content dictionary contained 15 variables.

### *12.5.2 Video selection and ground truth*

The present study utilised 39 (23 innocent appealers and 16 guilty) video recordings of individuals appealing during television press conferences for the return of their relative, to search for what happened, and to find the killer or abductor of their relative. They consisted of 18 female pleaders and 21 male pleaders. They were drawn from the United Kingdom, United States, Canada, New Zealand and Australia. There were 19 direct and 20 indirect appeals used (see Appendix O) indicating the number of innocent appealers and guilty ones gave direct and indirect appeals. Cases were selected based on ground truth, where it is known that the appealers were either truly guilty or innocent on unequivocal evidence.

Appendix H provides a list of evidence used for deceptive and honest appeals drawing on ten Brinke and Porter’s (2012) list, enhanced by Whelan et al. (2014). In cases where the appealers were deceptive, all appealers were convicted of their relatives’ deaths. In cases where the appealers were truthful, someone else was sentenced for the death of the appealers’ relative for 20 of these cases, and the appellant was not implicated with evidence of foul play for the other three innocent cases.

## **12.6 A Partial Order of Genuine and False Appealers with Base Coordinates (POSAC)**

Due to the small sample size, the usual statistics normally applied for this kind of data in lie detection literature as in the case of Whelan et al. (2014) may overinflate p-values, running the risk of rendering it non-generalisable. This method also overlooks variables with low frequency counts. Parametric tests lack power where sample size is small. If used, the subsequent results hold questionable assumptions that may need additional review. While useful for larger samples, parametric tests relying on traditional statistical measures that cannot effectively discriminate outcome are perhaps less fitting with the current sample and the general aim of this study.

Shye (1978) provided this by calculating base co-ordinates in his Partially Ordered Scalogram Analysis with Co-ordinates (POSAC) which is now available in the HUDAP 8 statistical package. POSAC analysis works by ranking data in a two-dimensional space. This method works well when data is categorical and for variables that have been observed on a scale where only ranking is important (Metric and non-metric scaling, n.d.). Since POSAC is an ordinal scaling (nonmetric) procedure, it makes no assumptions of the data and can be applied to analyse language patterns of these television appealers. The advantage of this method is that it enables observations of language patterns and variables to be studied simultaneously, reducing the likelihood of Type I and II errors. POSAC was chosen as it makes the best use of the existing structure of multivariate data. It achieves this by searching for patterns and examining the fundamental internal organisation of the data while doing little to simplify and reduce assumptions of linearity and distribution normality. The latter runs the risk of neglecting important aspects that may arise to be significant.

Another clear advantage of using POSAC is the ability to study several language areas simultaneously. Proponents have begun to realise that a multi-cue approach is necessary in lie detection methodology rather than comparing the mean frequency of one cue at a time to distinguish liars from truth tellers (Ekman et al., 1991; Leal et al., 2010; ten Brinke & Porter, 2012; Vrij & Mann, 2004). Instead of analysing variables one-dimensionally and focusing only on singular cues at a time, the current study uses POSAC to observe several language areas simultaneously and maps how these variables interrelate with each other. This

form of multi-dimensional scaling allows the exploration of relationships between ordered variables to reveal any underlying structure. Further, this method of analysis encompasses even the variables with the lowest occurrences.

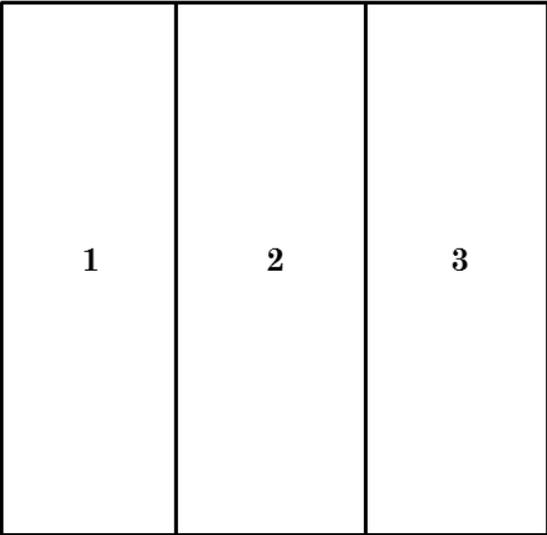
With many variables to consider and a small sample size, POSAC was conducted in the current study. POSAC uses a similar underlying calculation as principal component analysis (PCA). However, while PCA involves a weighted element of the variables using a procedure that is subjective, POSAC relies instead on locating variables by utilising only their partial order (Raveh & Landau, 1993). Consequently, this makes POSAC more systematic, less arbitrary and more robust to its data in comparison to PCA. Using POSAC as a method of analysis for the current sample increases our knowledge of collective language areas in its ability to distinguish truthful appellants from false ones in televised appeals.

For the present analysis eight language areas discussed above were created with 15 variables to quantify these areas under each category. POSAC was conducted to establish how the eight aspects when combined, could distinguish genuine from false appeals (Shye, 1978). If any of the variables assigned to an aspect occurred for a video then that aspect was assigned a value of 1, otherwise it is given 0. In the current study, the factor regulating the order and uniformity of these variables in direction and implication are verbal indicators of truthfulness. To ensure all aspects were consistent in their direction and implication, aspects that pertain to false appeals are inversely coded, i.e. their absence is given a score of 1. Here the language areas of self-centeredness, grammatical errors and chronology are reverse ordered. This means that all variables will measure the same hypothesis in the same direction (Griffel, 1999). Consequently, any video that contained all eight indicators of truthfulness would have a profile of 11111111; a video with no indicators would show a profile of 00000000. POSAC arranges profiles by positioning aspects that are greater than other profiles towards the right side of the solution. This positioning allows profiles to be compared to each other. For example, 00101101 is comparable to 00001101, but not to 11011000. 00101101 is quantitatively the same as 11011000 because the 1's add up to 4 in both cases, but they are qualitatively different. POSAC finds the best two-dimensional solution for all existing profiles. All

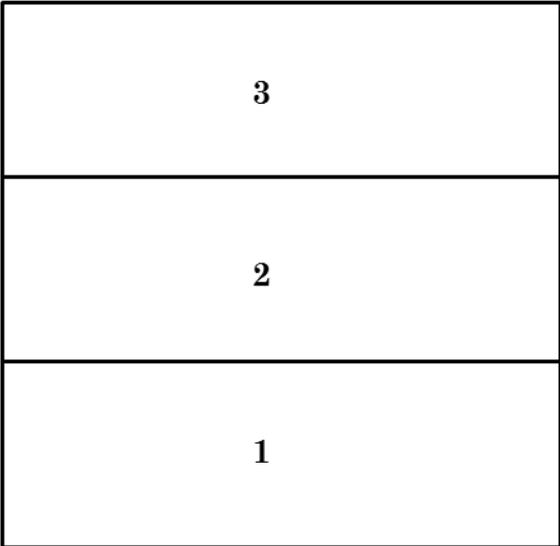
possible variations, i.e.,  $2^8$  which are 256, are considered though the assumption is that far fewer variations will be present and it will be possible to represent them in a two-dimensional configuration.

In this configuration, then, each unique profile will be represented as a point that can be identified and labelled. The location of the points indicates the partial order among the profiles. Integrated within the POSAC software are partitionings, which are built on what is known as ‘monotonicity coefficients’. These are computed for each variable (Shye et al., 1994) in terms of regions that are applied for analysis (HUDAP; Amar & Toledano, 2001). These coefficients determine how accurate a representation this partition or division is between appeal cases. While a coefficient of 1.0 represents a perfect division where all cases with the same variable score can be found on one side of this division, coefficients exceeding 0.8 is satisfactory (Shye et al., 1994). A lower coefficient indicates a lower chance this division truly discriminates between cases.

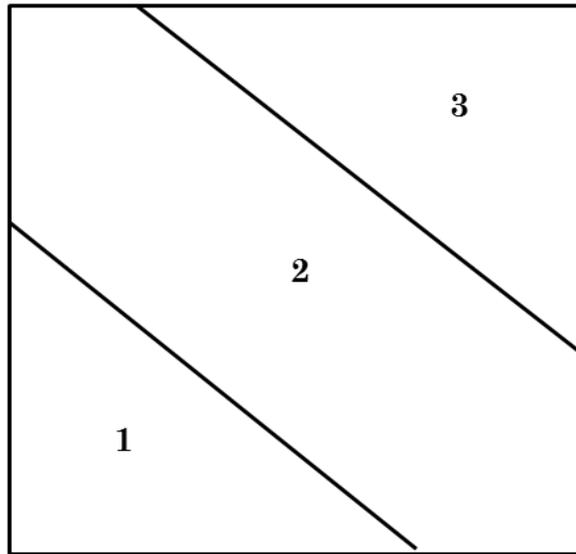
The  $X$  and  $Y$  division partitions the solution according to the  $X$  and  $Y$  axes. When multiple variables form a similar division, a pattern emerges. Another type of division that runs along the diagonal axis is the  $J$  division (Taylor, 2002). This is known as the Joint ( $J$ ) axis because it encapsulates the ‘joint’ score of the total number of 1’s in the profile. In this study, this would be the main quantitative axis determined by the joint combination of all eight language aspects. The most truthful profiles will be found in the north-east part of the plot, and the least truthful profiles will be found in the south west corner. Hence, it is hypothesised with a line drawn across the  $J$  axis that the north-east corner (region 3 on Figure 6) will contain most genuine appeals and that the south west corner (region 1 on Figure 6) will contain most false appeals. Profiles with the same score on a verbal variable are located nearer in space to each other than profiles with different scores on this variable.



**Figure 4**  
*X axis on POSAC solution*



**Figure 5**  
*Y axis on POSAC solution*



**Figure 6**

*J axis on POSAC solution*

An independent samples t-test will be conducted on  $J$  scores to compare genuine and false profiles to observe if score differences are statistically significant. These three different types of partitioning are shown in Figures 4, 5 and 6.

## 12.7 Results

### 12.7.1 Differences between genuine and false appeals

The dichotomously-coded presence or absence of each action unit was substantially reliable; (Kappa = .61-1.0,  $p < .001$ , 79- 86% agreement; Cohen, 1960; Landis & Koch, 1977). The frequency of each of the 15 variables for genuine and false appeals is reported in Appendix A. Although 39 videos are a far larger number than in previous studies, for statistical analysis a group of 23 compared with 16 provides only limited statistical power. It is therefore instructive at this stage to briefly consider the differences in frequencies of the variables between the two groups. Firstly, 65% of innocent appellers appeared alone in comparison to only 13% who appeared with another family member or relative.

Honest appeals had high frequencies for the use of the victim's name in the present tense. A higher number of truthful appellers kept count of time since the disappearance or death of their relative (27%) versus 6% for false appellers. Genuine appellers never use the conjunctions 'if' and 'or' regarding victim, and the phrase 'I didn't do it'. This may be because they were not accused.

The only variable that had a frequency higher than 50%, at 88%, was the variable Passive for deceptive appellers. A higher proportion of false appellers were subject to grammatical and syntax structure errors in their appeal (41%) compared to 18% of truthful appellers. Likewise, false appellers were more likely to mention 'I don't know' in their appeal. False appellers never mentioned 'religion', 'hope, misses and 'imagining the victim'.

### 12.7.2 Results of POSAC

As explained above, each profile is illustrated as a point in the POSAC solution. Points of profiles that are comparable will be found on a line that slopes positively in the *joint* direction, or  $x + y$  (Rebhun, 2012). POSAC then computes a coefficient of goodness-of-fit for the representation of the partial order identifying how well the program correctly represents the proportion of profile pairs. Figure 7 displays a two-dimensional POSAC solution with a coefficient of correct representation of 0.96. If a POSAC perfectly represents items in its plots then this measure of correct presentation will be 1. This means that 100% of the profiles are placed in the correct regions of the solution (Taylor, 2002), indicating the best goodness-of-fit (Shye, 1985). Shye (1985) stated that a coefficient of 0.80 and over is desirable for meaningful interpretation of results. In the current study, results indicate that 96% of profile pairs are correctly represented, and this goodness-of-fit measure is more than satisfactory (Borg & Shye, 1995). The interpretation of the underlying system that gives rise to these variations provides insight into the linguistic features used by appellers when pleading for their relatives' return. This interpretation relies on how each of the features partitions the overall POSAC solution. Figure 7 provides the POSAC diagram of the 39 appeals in this sample across the eight aspects. Here an increase in eight language aspects runs from the lower left (i.e., 00000000) to the upper right corner (i.e., 11111111). In

order, the aspects running from left to right are self-centred, tense, victim-centred, social, emotional, grammar, chronology and movement. Thus, a profile of 11111111 indicates no presence of self-centredness, using present tense when talking about the victim, is victim-centred, incorporates social and emotional aspects in the appeal, with no mention of chronology and no grammatical errors, and is active in moving the investigation forward.

In the present case, there are 25 unique profiles across the 39 different appeals. The strong association between all the eight language categories and the *J* axis indicates that the variables are actually ordered in the same direction with the same meaning, according to appeal veracity (Raveh & Landau, 1993).

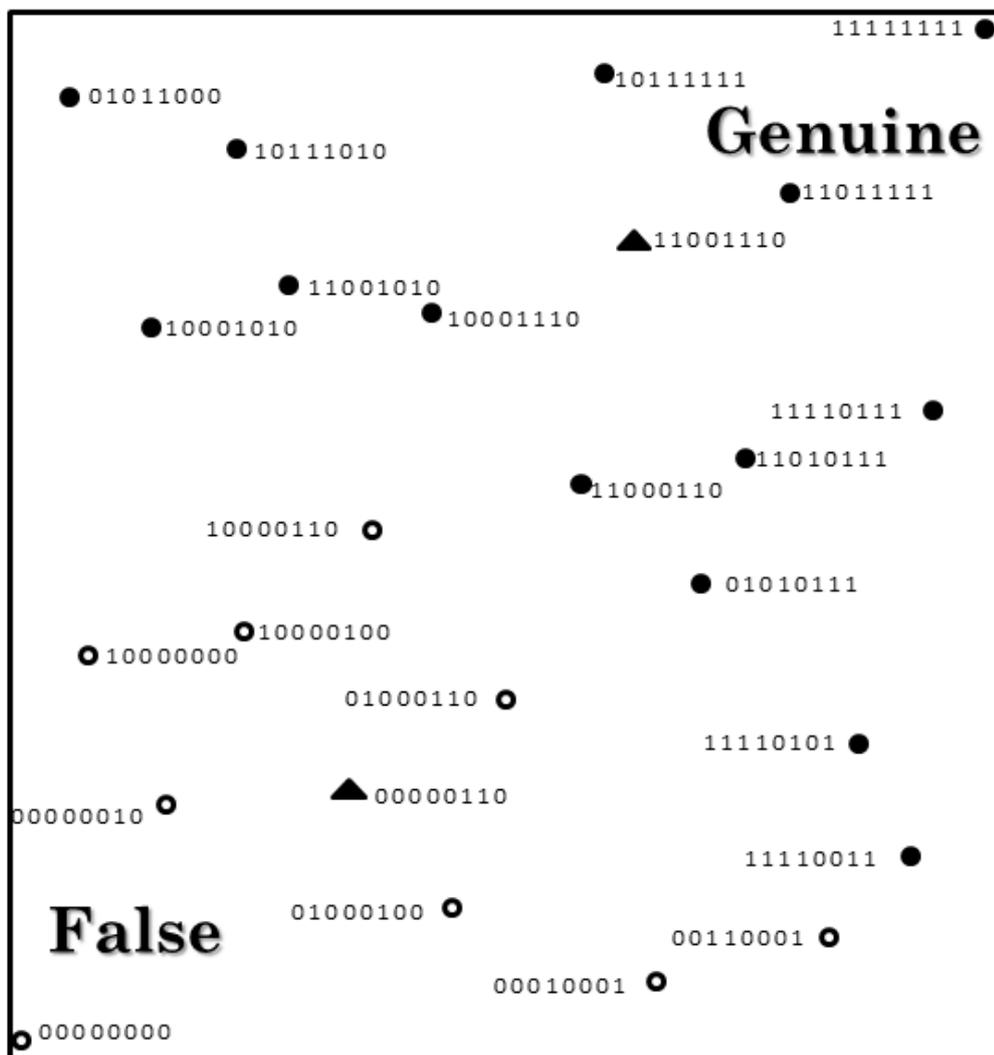
The POSAC program gives the coefficients of weak monotonicity coefficients of the 8 variables with the two main axes, *X* and *Y*, shown in Appendix P. Victim-centredness is correlated with the *X* axis monotonically (0.82), while self-centredness is correlated (0.91) with the *Y* axis, as seen by the factor loading of each variable. The emotional aspect is perfectly correlated with the *Y* axis (1.00), and chronology is correlated with this axis as well (0.86). The aspect of movement in language is almost perfectly correlated with the *X* axis (0.99). Present-tense (0.86) and social aspects of appeal language (0.80) are correlated with the *X* axis as well. The measures of monotonicity between each observed variable and the eight language categories indicate that appeals with relatively high (low) self-centredness, chronology and emotional aspects in appeals on the top (bottom) side of Figure 7, where a high monotonicity association is seen with the *Y* axis. Appeals that have a relatively high (low) movement-related words, social aspects, grammatical errors and victim-centred-ness are located at the right (left) side of the same figure. The right-hand side region will contain active and emotional appeals whereas the left will contain passive and less emotional appeals.

#### *12.7.2.1 The X-axis: Movement*

The underlying structure of POSAC is best ascertained by defining the *X* and *Y* axes (Borg & Shye, 1995). Appendix P shows that the aspect most associated with the *X* axis is that of 'movement', which has a loading of 0.99. Figure 7a shows this partitioning of the POSAC space. This shows that indication of

desired action is a major factor underlying all the other aspects. This fits with the central assumption that the false appellant knows such action is fruitless, or indeed potentially dangerous in revealing his/her own culpability.

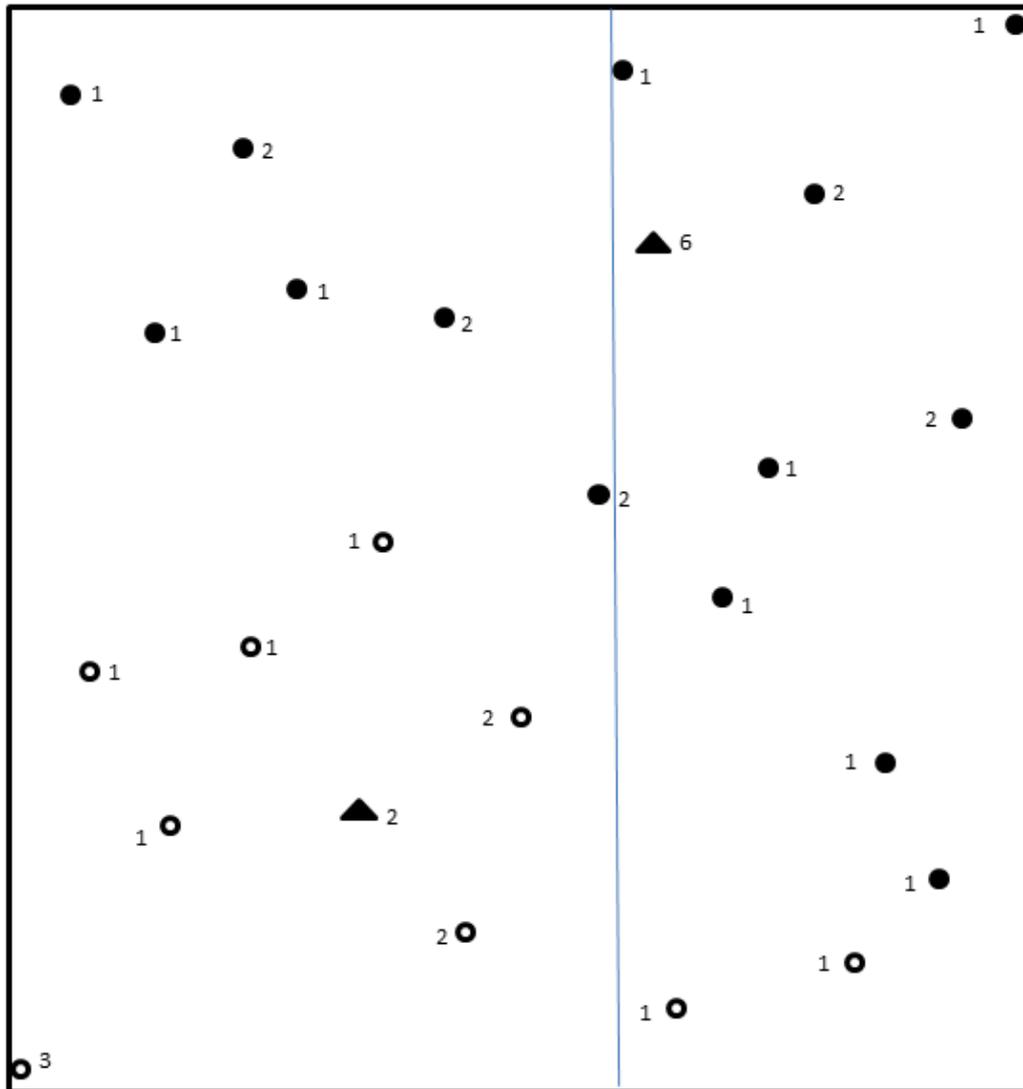
Exploring appeals with this axis in mind, a majority of appeals towards the right-hand side of Figure 7a contain the aspect of activity, and a majority of them are true appeals whereas on the left-hand side of the same figure the truthful appeal profiles feature mostly only on the top-left region. In the bottom-left region, there are two appeals that feature the same profile of 00000110, where minor grammatical errors and chronology are present in their appeal. These appeals are marked with a triangle on Figure 7a, where one appeal is true and the other false. Present in both appeals are any language aspects of self-centredness, and absent in these appeals is the usage the present tense when talking about the victim. The lack of present tense is justifiable in the truthful appeal as the victim is already



**Figure 7**

*Two-dimensional POSAC solution*

dead; however the use of past tense in the false appeal raises caution as the victim is still alive and only kidnapped. The appeals also lack of social and emotional aspects of language. This appeal is situated in this region as it lacks any movement in finding the victim's killer, as well as a lacking a clear social and emotional element in the appeal.

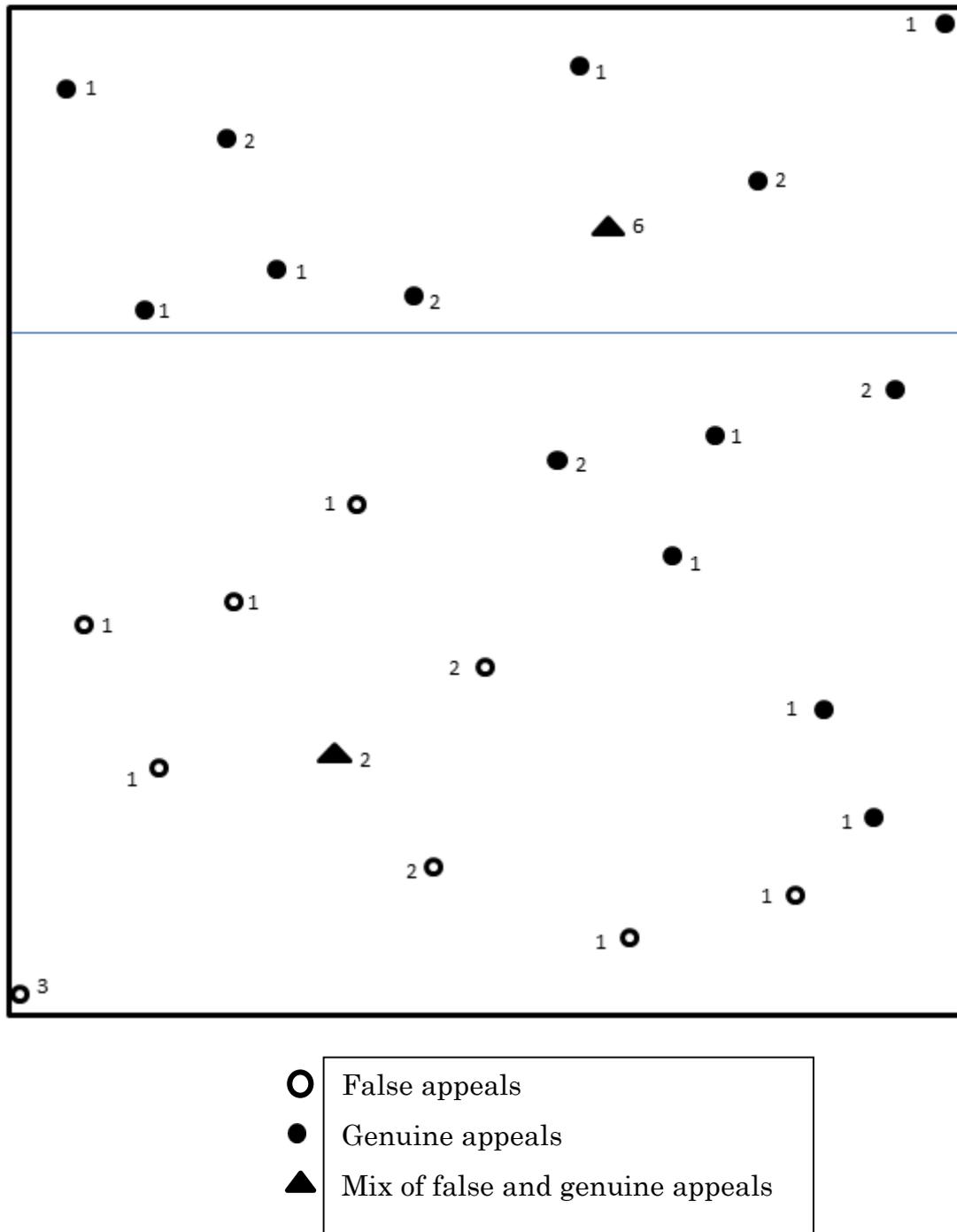


- |   |                                  |
|---|----------------------------------|
| ○ | False appeals                    |
| ● | Genuine appeals                  |
| ▲ | Mix of false and genuine appeals |

Note: X partitioning loading coefficient = 0.99

**Figure 7a**

*Partitioned item diagram for the aspect of movement with frequencies of profiles*



Note: Y partitioning loading coefficient = 1.00

**Figure 7b**

*Partitioned item diagram for the aspect of emotions with numbered frequencies of profiles*

*12.7.2.2 The Y-axis: Emotional*

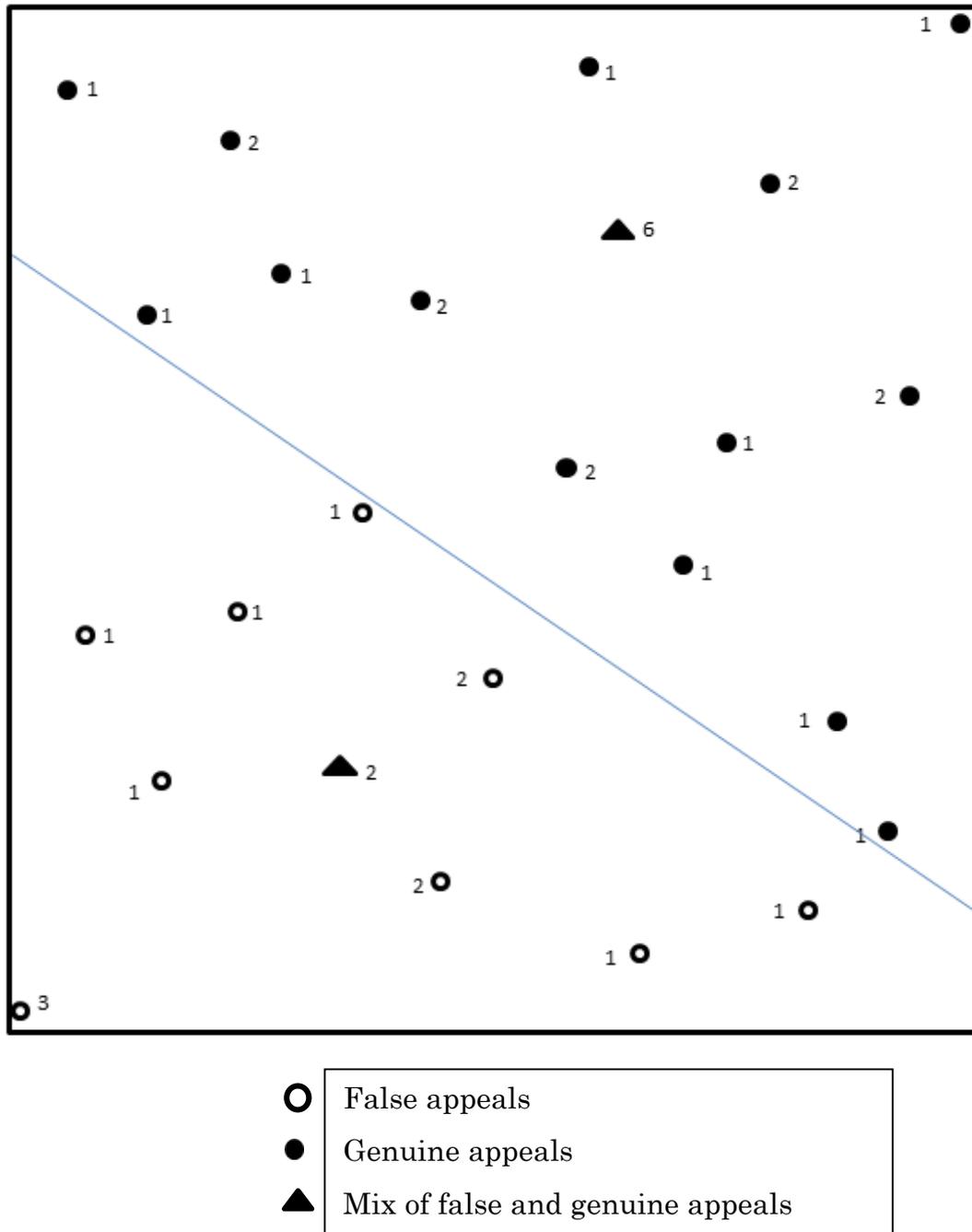
The vertical axis as shown in Figure 7b is created by the aspect of emotions, having a loading of 1.00 on the *Y* axis. Further towards the top side of the partitioning the appeal profiles are indicative of truthful ones. This therefore indicates that emotional expression is orthogonal to action in distinguishing between appeals. Focusing further on profiles of appeals with this axis in mind, the bottom-most left region of Figure 7b contains three appeals that are devoid of any emotional aspect, including hope and missing the victim. As hypothesised, these appeals also contain aspects of chronology and grammatical errors. In contrast, the appeal profile on the topmost-right region contains spoken emotions, as well as showing indicators of activity and social aspects as hypothesised.

In the mixed profile on the bottom-left region of Figure 7b the false appealer's statements was devoid of any mention of emotion and so was the true appealer's. Coupled with a lack of movement and social aspects of the appeal this appeal, the truthful appeal seemed to contain elements of a false appeal. Within the top-right region of Figure 7a, it can be seen that six profiles show a pattern of 11001110. These profiles contain a mix of truthful appeals and false appeals. Emotional aspects of language are present in all six appeals with this profile. Together with the self-centredness aspect absent, no grammatical errors and chronology present the false appeals with this profile show aspects of a truthful appeal by including spoken emotions. The false appealers also speak of the victim in present tense (all but two appeals with this profile contain cases where the victim is not yet known to be dead, and thus still considered alive) while two hold knowledge that the victim is indeed dead.

### *12.7.2.3 The J axis: Appeal veracity*

The POSAC analysis does not directly incorporate appeal veracity. However, because each of the appeals is represented in the output it is possible to identify which ones are genuine, and which not. These are indicated in Figure 7c. It can be seen there that the *J* axis, the diagonal from bottom left to top right, reflects appeal truthfulness. A line can be drawn across the *J* axis that has virtually all the genuine appeals in the top right of the line, and all except one of the false appeals to its bottom left.

Each video, as represented in the POSAC output, has a value on the J axis. The line that distinguishes nearly all the truthful from nearly all the false appeals



Note: J partitioning loading coefficient = 0.95

**Figure 7c**

*Partitioned item diagram for appeal veracity with frequencies of profiles*

can be drawn at  $J = 100$ . Appeals with a value greater than this tend to be true, those less than 100 tend to be false. An independent samples t-test was carried out on these  $J$  scores comparing genuine and false profiles. A highly significant difference was found;  $t(22) = 5.23$ ,  $p < .0001$ . Genuine appealers scored higher ( $M = 124.72$ ,  $SD = 32.56$ ) than false appealers ( $M = 56.94$ ,  $SD = 27.24$ ). The magnitude of the difference in means (67.78, 95% CI: 40.9 to 94.7) is large (eta squared = .53).

## 12.8 Discussion

A person who has never actually gone through bereavement has to fake grieving. Even if they successfully feign emotions in their non-verbal expressions, their speech has the capability of revealing aspects of their true emotional state which they may not be aware of, as shown by results of the current study. Furthermore, the genuineness of the desire to find the victim and/or the culprit is implicit in what the appealers says. As shown in Figures 7a and 7b, these two axes of emotion and active movement underlie the six other aspects of appeals derived from the existing literature and the consideration of what psychological processes are relevant to discriminating between genuine and false television appeals.

Genuine appealers are significantly more in touch with their emotions, even if they are experiencing emotional apathy, and are more significantly likely to publicly state them. They are more cognisant of their mental state, be it true grief regarding the situation, being at peace, or completely devoid of emotions. The deficiency of emotional words and even emotional lability revealed in false appealers is highlighted; complementing the behavioural cues reported by others (ten Brinke & Porter, 2011).

Significantly more true appealers talked about the victim in the present tense, whether they are missing or dead, reiterating hope for their return. While it is acknowledged that there may be appealers who are not spiritual, believing in a higher power or powers (i.e. they are atheist or agnostic), honest appealers are significantly more likely to rely on belief systems to sustain them and are significantly more hopeful, an aspect of genuine appeals not previously identified.

Although emotional aspects, such as those found in the present study, have been acknowledged as important in detecting deception, other context specific

aspects have also emerged as relevant. Of particular interest are victim-oriented comments especially, the emphasis on bringing the victim home, if there is a chance they might still be alive, or catching the killer. What were characterised here as ‘movement’ related words have not been explored in previous studies. Because of the focus of the television appeals studied, the urgency in wanting relatives back has a particular significance and has a much lower frequency for deceptive appealers.

Deceptive appeals are generally self-oriented, using distancing language, denial of involvement, blaming others and setting conditions to the return of the victim. This verbal distancing is echoed in more conditional clauses containing the conjunctions of ‘if’ and ‘or’ regarding the victim coming home, such as ‘If you can’t come home’. These conjunctions tellingly create a disruption in the sentence, breaking the flow of their narratives. True appealers in this sample never used these conditional clauses or conjunctions.

Distinct differences in an appeal statement is revealed between truth-tellers and liars, where truthful appealers show a higher tendency to imagine the victim. This implies that genuine appealers are more likely to be highly fixated with the missing or murdered victim. The verbalised imagining about the victim and yearning to be close to their murdered or missing relative is entirely absent in false appeals within this sample (see Appendix A).

False appealers are significantly more likely to refer back to the event, and talk about the situation during their press conference or interview statements, frequently in sequential order. This is an interesting attempt to reveal something they know to be true as part of creating a false narrative.

## **12.9 Conclusion and implications**

Distinctive differences in verbalisations separate genuine appeals from those that are not. Truthful appealers reveal various coping styles missing from lying appealers’ narratives. Taken together the eight aspects of what is said in television appeals provide a highly significant discrimination between genuine and false appeals. By utilising a novel multi-dimensional scaling analysis on dichotomous variables it was found that a score could be generated that could be

used to assign videos to genuine or false categories with an extremely high degree of accuracy.

The underlying psychological processes that give rise to the differences in verbalisations relate directly to the specific context of these appeals. Genuine appealers are preoccupied with the victim. They are using their presence at a televised press conference to generate action that will be of value. Emotionally they still think of the missing relative as being alive and yearn to continue the bond they had with them. False appeals are being used to exonerate the appealers by attributing agency away from themselves, denying committing the crime, retelling of the crime events (presumably fabricated), and a general passivity.

The results demonstrate the value of elaborating cues to deception that derive directly from the specific context in which the truth/deception is being generated. For whilst in the present study some of the significant differences, notably those relating to emotion, may have generic value across a range of situations, most of the differences take their meaning directly from the events being described.

The 39 videos used in the present study are still relatively few in number for statistical analysis. Further validation of the results found here with other examples is therefore of great importance.

As a verification of the results of the study, content analysis in future studies should be carried out in total ignorance of the truthfulness or otherwise. That was not possible in the present study, by necessity, because of the need to establish ground truth as the basis for selecting the videos. Although in the present case the second rater, who contributed to the inter-rater reliability assessment, was ignorant of the truthfulness of the videos.

Generating new cues derived from grief theories that relate directly to the context of the appeal, and producing profiles of appeals across a number of aspects analysed by POSAC all provide novel possibilities for future studies in a wide variety of different circumstances. They show that speech differences can be identified that distinguish between honest and deceptive appealers with a very high degree of reliability.

# CHAPTER 13

---

## GENERAL DISCUSSION AND CONCLUSIONS

The purpose of this final chapter is to summarise and present an overall discussion of the results found in the current thesis. While it is known that lie detection accuracy is poor, it is not known whether the same accuracy would persist if people are asked to make innocence-guilt judgments. Additionally, efforts in understanding the cognitive perceptions that underlie judges' accuracy rates are lacking in this type of veracity judgment. As expounded in Chapter 1, existing explanations as to why we perform so poorly in accurately detecting innocence-guilt have not been systematically reviewed and categorised into an operational framework. A visual framework of where errors may lie is introduced in Figure 1 (Chapter 1). Currently, it is inferred that the concepts of truth bias and innocence bias would overlap, however this remains to be further studied and extricated. In the present thesis, both truth-lie judgments and innocence-guilt judgments warranted a discussion as they have been previously interchangeably used (Kassin & Fong, 1999; Inbau et al., 1986). Logically, there would seem to be a relationship between the two in the current context, especially given that participants in Part 2 were not privy to any other forms of evidence or information about the appellars. While conjecture, it seems that they would most probably have to rely on whether they think the appellars are lying or not to judge their innocence or guilt.

As given emphasis throughout the present thesis, one of the foremost methodological validation in using a very specific context is that the saliency and production of cues are dependent upon a range of situational factors (Porter & ten Brinke, 2010) therefore cues that are salient in one situation may not be so in another. Not having a contextual focus high stakes research scenario runs the risk of precluding certain cues from appearing. Apart from the small pool of research that investigated high stakes scenarios, most lacked a context-specific approach (Harpster et al., 2009; Reynolds & Rendler, 2010). Several authors put forward the view that this lack of context-specific approach may have promoted to the seemingly inconsistent discoveries in high stakes lie detection literature (DePaulo et al., 2003; Ten Brinke & Porter, 2012; Mann et al., 2002; Whelan et al., 2014) as well as counteracted context-specific cues from being discovered. By utilising a methodological approach which allowed not only using frequency counts in

investigating the cues to deception but also in examining cues in relation with each other, the findings (expounded below) propose that using a context-specific approach facilitated the findings of novel cues that were formerly not capable of being identified.

One of the central objectives of the current thesis was to examine what effect using ecologically valid high-stakes stimulus materials has on the accuracy of judges in detecting innocence and guilt across several experimentally manipulated conditions, thereby observing different biases and heuristics that may surface in each condition. In order to elicit differences in responses across conditions (associated with differences in underlying mental processes) manipulations were implemented by accentuating (or minimising) different cues across conditions in Studies 1 to 4. These manipulations were grounded upon theoretical bases and substantiating arguments all expounded in the opening chapters of this thesis.

### **13.1 OVERARCHING THEORETICAL ARGUMENT AND DEVELOPMENT**

This section reviews current theories that account for all the issues this thesis has explored. Most existing theories discussed in the present thesis would best explain a truth-lie judgment and subsequently the concept of the truth/lie bias and truth/lie default. However, as aforementioned they warrant a discussion nevertheless as logically there would seem to be an overlap between these concepts and innocence-guilt decisions especially in the current context.

The Spinozan model asserts that individuals have no other option but to initially accept incoming information as truthful by default (Gilbert, 1991). While it can be re-examined later, this initial truth bias in this model is non-adaptive, inflexible and cannot be substituted with another way of thinking even after being alerted (Gilbert et al., 1990). The truth default theory (TDT; Levine, 2014) includes this Spinozan model in its tenets. It proposes that people either assume truthfulness after failing to actively bring to mind the likelihood that they may be lied to, or they assume truthfulness as a default thinking mode when failing to find *adequate* indication to confirm deceit (Levine, 2014). On the contrary, the

ALIED model states that it does not always require greater mental effort to consider the possibility of deceit, as TDT claims (Richter, Schroeder & Woehrman, 2009; Street & Richardson, 2014; Hasson et al., 2005).

Levine (2014) argued against his original depiction of the truth bias as a non-adaptive cognitive distortion (McCornack & Levine, 1990) and recently asserted that both the truth default and truth bias are practical and adaptive in helping one make accurate veracity judgments. While ALIED shares this adaptive perspective, it claims that there are no set cognitive defaults, rather, response biases stem from the same underlying process. That is, there is no truth default as there is no lie default. ALIED goes further to contend that in particular situations it is more cognitively taxing to assume truthfulness, for instance when made suspicious (Street, 2015; Levine & McCornack, 1991; McCornack & Levine, 1990; Stiff, Kim & Ramesh, 1992; Toris & DePaulo, 1985). Whether it takes more effort to trust or distrust depends on the congruency of the incoming information with the observer's beliefs, understanding and past experience (Ng & Youngs, 2016; Street, 2015; Lane & Harris, 2014; Gervais & Henrich, 2010; Woolley & Ghossainy, 2013).

The ALIED approach stems from Brunswik's (1952) theory that individuals adapt the type of information they utilise to make a decision instead of always relying on their understanding of the general context. ALIED advocates that individuals endeavour to make veracity decisions of a sender or statement by utilising information from or about the specific statement or sender itself, as opposed to general information about statements or senders (Street, 2015). However, it is not always that information about a statement or sender is available or if available, the diagnostic value of these cues may be low.

In the present thesis, the expectation was that by using material from high stakes Sources the diagnostic value of cues will be higher. However, because the stimulus materials were very short, information available to judges about the appealers and appeal circumstances were partial and limited. In the deficiency of information about the Source, there is a possibility that judges seemed to have relied on context-general information to come to their veracity decisions, even more so when certain cues were experimentally accented (or minimised) across

conditions. The reliance on context-general information was expected to, in turn, affect accuracy of veracity decisions. Because each experimental condition was manipulated it was also expected that differences in reliance on context-general information would be observed (through implicit and explicit lie detection measures) and this will also be reflected in accuracy rates. The ALIED model would provide the best explanation for the results achieved in the present thesis, at this stage.

ALIED acknowledges that one of its limitations is that it offers no claim on how individuals choose and incorporate these context-general cues. In fact, ALIED does little to elaborate on these cues themselves. The theory loosely mentions that these context-general cues include social norms, individual relationships and emotional aspects (Street, 2015). While Street (2015) discussed the truth and lie bias in depth, little is mentioned about the affective facet – neither did the author dispute its existence. The present thesis posits that meta-emotion biases can also arise in the face of limited information about the Source as a result of affective heuristics, even more so in a context where emotional outpour of appealers is highly anticipated such as that in television appeals. Response biases pertaining to emotions can be understood as a usage and elaboration of Street's (2015) ALIED model.

Granhag (2006) alleged that the dual process theory is less applicable in lie detection, as he stresses that the mutual exclusivity concerning the explicit and implicit is questionable. It should be noted that there are many classes of dual process models (Evans, 2007; Street & Masip, 2015). One class is known as the 'default-interventionist', where these models claim that heuristic processing is the default processing mode (Gilbert et al., 1990; Gilbert et al., 1993). However, this can be interrupted by analytic processing if and when enough time is offered (Evans, 2007). Another class, called the 'parallel-competition', is where both System 1 and 2 take effect concurrently. That said, if there is only limited time offered the resulting judgment is more prone to be built on heuristic processing. These two classes claim that analytic processing comes after heuristic processing (Street & Masip, 2015; Sloman, 1996; Sloman, 2002). Lastly, the 'pre-emptive conflict resolution' class claims that either System 1 or 2 is

selected from the beginning and analytical processes take place when more effort in processing is required (Evans, 2007).

In response to the premise of there being an interaction between the two systems, several researchers propose a singular-system model where the implicit and the explicit processes operate on a continuum (Cleeremans & Jiménez, 2002; Farah, 1994; O'Brien & Opie, 1999). Cleeremans (2006) contended that both conscious and unconscious information processing arises from the same underlying system, and that consciousness ought to be regarded as a gradient on a continuum. The dynamic gradient account (DGC) of consciousness contrasts dual processes by advocating a singular system framework (Cleeremans & Jiménez, 2002).

Another contention against the dual process model is that the dichotomy does not satisfactorily explain or encompass the breadth of cognitive routes found in research on mental reasoning. Moshman (2000) argued that the dichotomy of System 1 and System 2 overlooks an important distinction between automatic and implicit processing. The author contends that, for example, rule-based processing can be automatic and explicit processing can encompass the conscious use of heuristics. Hence, Moshman (2000) suggested two additional mental processes on top of System 1 (automatic heuristic processing) and System 2 (explicit rule-based processing), the automatic rule-based processing and explicit heuristic processing. These models highlight alternative theories that assume that the conscious and unconscious are aspects of more than only two systems or one interactive system.

Figure 2 exhibits a culmination of all theories reviewed in this section pertaining to this thesis. At this stage, the hypothesis underlying Figure 2 is that consciousness operates on a continuum and is interactive with the unconscious; the two '*sub*-systems' work in concert. The results from the experimental studies in the present thesis found little evidence for any default response that is pre-selected where people will always be innocence-biased or guilt-biased. This is akin to a truth default (Levine, 2014) in lie detection research. Further support for other abovementioned models (truth default, ALIED, dual-system, multi-system and single-system) can be further verified if the role of the unconscious is clarified.

In relation to the framework presented in Figure 1 (Chapter 1), while preliminary, results from the present thesis is most consistent with ALIED's claims that there are no set cognitive defaults. Rather, at present they seem to infer that response biases stem from the same underlying process. The results from Studies 1 to 4 at present challenged the initial assumption at the beginning of this thesis that judges will always draw on heuristics (which always leads to 'biases') to determine veracity. Evidently, it must be acknowledged that the number of videos shown in Studies 1 to 4 were low and thus any interpretation of the results must be made with caution. However, these results set valuable groundwork for further studies to either further support or challenge these initial suggestions, and to increase our understanding of the workable representation set out in Figure 2.

The novel aspect of the present thesis was that it included a grief paradigm, when previous research predominantly focused solely on the lying process. Before considering what subjective perceptions and expectations observers hold, the present thesis first explored what was the norm behaviour in a given situation. In the current context, truthful television appeals are made by individuals who are in distress and grieving either the momentary loss of a missing relative or the permanent passing of one murdered. This consequently proposed a framework for studying expected indicators of a deceptive appeal. While of course grief can be displayed through physiology and behavioural actions, verbal aspects were chosen to be focused on in the current thesis as speech content has been demonstrated to indicate the foremost importance and mental states of the applier; whether it is finding the victim urgently or fervently attempting to cover up a crime (Pennebaker et al., 2003). When norm behaviour of a specific context is known, deviations from the norm hypothetically serve as a red flag towards the occurrence of deception.

As shown in the literature review set out in the opening chapters, the lying process is ultimately context specific and previous research efforts have not been able to focus on a model of expected content before. Considering the normative behaviour and responses in a unique high stakes situation allowed the present thesis to build a framework based on the expected verbal content.

### 13.2 Grief theories as a context in studying deception

Models and theories of grief explain that grief is the outward presentation of inner feelings and mental states, sometimes indicating presence of clinical symptoms in accordance to bereavement (Rosenblatt, 1988; American Psychiatric Association, 2013). Overall 1) research over the past several decades implies that an archetypal reaction to grief and/or distress is a fixation upon the loss or separation from the victim and the detachment from a loved one. Grievers usually want to continue this bond and keep ties by being preoccupied in terms of thinking about them, writing about them, remembering them, make a meaning of the loss, imagining them, keeping count of the days since the victim left them (Lindemann, 1944; Hogan & DeSantis, 1992). 2) True grievers and individuals in distress experience an emotional release or reserve, where they either experience a parade of emotions in reaction to the loss as manifested in emotional symptoms, or emotional lability and apathy towards the situation (Moyle Wright & Hogan, 2008; Shear & Mulhare, 2008). 3) In times of distress, the need for others and social cooperation increases to alleviate the pain and/or distress of the loss. The loss of a loved one, whether temporary or permanent, will also create social changes. They may turn to a higher agency such as God or a higher power to assume responsibility of their outcome, as well as to serve as a buffer for the pain and loss (Bjork & Thurman, 2007; Stone & Pennebaker, 2002). They may also experience stress, anxiety and depression amongst other family members as a change within the family structure will take place, as members attempt to communally compensate for and balance the loss. 4) Because emotional energy is invested in the victim a person in distress and/or grief will immediately want to find and secure their loved ones again (initially evident by the fact that they want to publicly appeal for their return, although verbal aspects will also show this urgency) or to urgently find their killer to remedy the loss they feel, as only finding them will alleviate their grief, the sooner the better. They would want to execute firm plans and rally the community in helping their cause, driving a sense of action to speedily change the situation. 5) Grievers express hope and tend to speak in the present tense. From a grief theory perspective, this conveys hope in being able to find the missing relative again in cases where a body of the victim has not been

found yet. This also signifies not having the knowledge of what actually happened to the victim.

One of the main premises of Part 2 in the current thesis was that while grief can be simulated, a deceptive griever may not know truly what a true griever would go through and may even contra-indicate certain deceptive cues themselves. They may find it hard to turn to a higher power, for example, as they would not be aware of the behaviours of a truly grieving individual. A true griever would have invested emotional energy in the victim for many years, and when this loss is real the grief that overcomes them is often challenging to imitate, however keeping in mind the caveat that false appealers may also truly feel grief for the loss of their loved one, where they may have killed them in a spur of a moment and later come to regret their actions. Contraindications, for example, includes blatant denial in distancing themselves from involvement in the crime and assigning agency away from themselves (Davis et al., 2005; ten Brinke & Porter, 2011) and using conditional markers when talking about a missing victim coming home, perhaps aware that there is no likelihood for return.

### **13.3 Perceptions and expectations of grievers**

Research on perceptions of grief has shown that people view levels and types of grievers' emotional expression as important, and favour certain styles of grief coping. Existing research tells us that people tend to favour an unfounded grief stage model where a firm expression of grief such as crying is expected with an unsupported time period, despite the fact that observers also hold the knowledge that grievers may experience variations in emotional expressions and knowing that grief responses cannot be generalized (Breen & O'Connor, 2007; Costa & Stewart, 2007; Penman et al., 2013). While holding the assumption that responses departing from the norm are not necessarily pathological, observers still favour the 'norm'. Research also tells us that emotional display and level of emotional expressions are substantially expected by judges in grieving and/or distressed individuals (Costa & Stewart, 2007).

Burgoon et al. (1975) proposed the language expectancy theory, a subset of expectancy violation theory, where the authors claimed what a speaker chooses to

say can significantly predict whether an observer is successfully convinced. This theory contends that people hold expectations as to what they consider normal language in a given situation, grounded in psychological and cultural norms of their immediate society, and bring these stereotypes into social situations. This means that a sender's words (and non-verbal behaviour) may be inconsistent with the stereotypes judges carry. Due to this violation of expectancies the observer carries, a sender may be seen to be less liked and even believed as their language and/or behaviour is inherently less persuasive. In the context of the present thesis, as discussed the language of grief is governed by emotional, psychological and social changes. As such, expectancies in the language of a griever will follow. Judges could and do expect, to a certain degree, what a griever should and would say.

While theories and models of grief exist (yet never served as a theoretical basis in studying high stakes deception), little is also known of perceptions and expectations of television appealers should behave.

#### **13.4 Implicit and explicit decisions**

Assessing implicit statements in the present thesis was to assist in determining a Judge's rating of anything relating to the appealers apart from 'do you think he/she is guilty or innocent?' The notion is that lying can be judged by perceivers in different implicit ways by asking universal questions, such as of 'do you want to marry him?' whereby an implicit measurement of the explicit notion of 'trust' is taking place, rather than asking 'do you trust him?' (ten Brinke, personal communication, 26<sup>th</sup> August, 2015). There is nothing 'hidden' about implicit judgments, and one might even go as far as to say that there is no division between implicit judgments being superior to explicit ones either, at least some of the time. Furthermore, certain explicit decisions cannot be explained by implicit decisions and implicit decisions cannot be explained by explicit ones (Granhag, 2006).

As shown in the series of experimental studies in the present thesis, explicit decisions were proven useful and recorded accuracies better than chance level in some circumstances perhaps due to the manipulation (saliency) of certain Source

characteristics and not necessarily due to the high stakes nature of the Source. Measuring both implicit and explicit veracity judgments allowed the exploration of the relationship between which of these cues participants utilised to form their judgments, and how this relates to their explicit veracity decisions.

Further, it allowed the investigation as to which cues predicted accurate explicit decisions. The aim to do so revealed fundamental complexities between the two concepts. Throughout Studies 1 to 4, it was discovered judges can identify accurate implicit judgments that potentially leads to correct explicit judgments, but they may still not accurately judge an appellant to be innocent or otherwise in their explicit decisions. This showed that they may not know the meaning of their decisions, and that accurately identifying implicit cues does not necessarily translate towards accurate explicit decisions. Thus, while there seems to be a relationship, this is still vastly conjecture at this stage and undisclosed is the extent of implicit judgments influencing the explicit decisions (and under which circumstance is this more likely), and explicit judgments influencing implicit decisions (and under which circumstance); if they do consciously, and if so under which particular circumstance do they occur consciously or otherwise. To confidently state the predictive value of one decision for the other is premature at this stage. Nevertheless, based on the current understanding of these results, Figure 2 offers a workable representation that stands to be further developed, verified or challenged with more research in this area.

### **13.5 VERBAL ANALYSIS**

Part 2 of this thesis sought to examine diagnostic indicators of deception in television appeals. A supplementary aim was to not only utilise a methodological approach which objectively observed verbal cues that differentiate honest from deceptive appeals to a high degree of certainty, but to examine these verbal cues in relation to each other to generate a verbal profile that will prove valuable in practice. The current thesis not only uses frequency counts, in line with a traditional method in investigating the cues to deception, but also examines verbal cues in relation with each other. The latter has not yet been conducted in the current context. The importance of investigating variables in relation to each other

cannot be undervalued and was outlined in Chapter 12. POSAC analysis provides a robust consideration of how cues to innocence and guilty appeals work with each other in a three-dimensional space. By using this method of analysis, rare behaviours that occur infrequently were also able to be incorporated to provide a methodical and robust representation of deception cues in this context.

This technique enabled an assessment of the degree of similarity between all appealers in the sample, based on the commonality of their appeal features. This was done by partially ordering these aspects on a geometric representation. This then allowed Study 5 to explore the relationships between ordered variables to reveal the fundamental structure of aspects derived from grief theories. The present thesis was able to convey in a conceptual space the relationship between appealers as determined by eight carefully selected aspects. The resulting plots showed appealers located close together, and within the spatial representation appeal profiles are plotted in a way that the closest to the appealers has the most similar profile and furthest away from the profile is most disparate. In Study 5, the J axis (which was the dominant quantitative axis determined by the joint combination of all eight aspects) was able to significantly and reliably discriminate between genuine and false appeal. This means that the interpretation of the underlying structure which imparts variability offers much insight into the language characteristics appealers use when pleading for a loved one's return.

Put together, all axes discussed in Study 5 underlie the aspects of appeals derived from the pre-existing theories found in literature and their significance on the cognitive processes were construed to successfully discriminate a true appealers from a false one. Study 5 found that genuine appealers were significantly more likely to publicly verbalise their emotions, emotional state or emotional lability they were experiencing whereas false appealers were generally devoid of verbalising this despite expressing emotions in non-verbal ways. This was an important distinction and finding. Whether the victims were missing or dead, genuine appealers were also significantly more likely to talk about the victim in the present tense. This showed hope for their return, presenting a clearer indication of innocence in not having knowledge of what happened to the victims. In contrast while true appealers in this sample never used conditional clauses,

false appealers used conjunctions of 'if' and 'or' regarding the victim coming home (e.g. 'If you can't come home, just call', indicating that they were either reluctant for the victim to come home as they will suffer repercussions in a case of kidnapping, or they held the knowledge that the victim coming home is not a possibility in a murder case.

While false appealers generally sought to distance themselves from any involvement with the victim, true appealers wanted to continue the bond that they lost with the victim. As shown in previous research by ten Brinke and Porter (2012), Zuckerman et al. (1981) and DePaulo et al. (2003), liars tend to show behaviours that generally impede a direct communication, choosing instead to use equivocal language in an attempt to psychologically eschew possible feelings of shame and/or guilt. In this instance, perhaps false appealers used distancing language, blame shifting and conditions to reduce their conflict created by the incongruity between their inner knowledge of what happened to the victim and their outer public pleas. In contrast, true appealers attempted maintaining a connection with the victim by verbally imagining them, counting the days since they were gone, and generally were preoccupied with the victim in their appeals. They relied on a higher power to sustain them through the pain they were going through, whereas all these aspects were absent in false appeals. True appealers displayed a verbalised urgency in moving the investigation forward to find their relatives or to find the killer in cases where a body has been found, whereas most false appealers remained passive in their appeals. The aspects categorised in Study 5 as 'movement' words had not been investigated before in the present context. False appealers were significantly more likely to refer back to the event, and talked about the situation during their press conference or interview statements, frequently in sequential order. This was an interesting attempt to disclose something they knew to be true as part of fabricating a false appeal.

Taken as a whole, the strengths of the present thesis included using high stakes situations, generating diagnostic cues derived from a strong theoretical foundation using grief theories that relate directly to the context of the appeal, and producing profiles of appeals across a number of aspects analysed by POSAC. All these provided novel possibilities for future studies in a wide variety of

different circumstances. They showed that speech differences can be identified that distinguish between honest and deceptive appellants with a very high degree of reliability. In future, these diagnostic indicators will benefit from being included in the implicit portion of the questionnaire in Part 2.

### **13.6 METHODOLOGICAL LIMITATIONS**

First and foremost, it must be acknowledged that there was no room for studying the objective verbal aspects found with observer reactions to these implicit judgments in the present thesis. It would have been valuable indeed to include diagnostic indicators found in Study 5 in the questionnaire used in Part 2. This will be expounded in the Future Suggestions for Research section. In addition, although the questionnaire used provided interesting results it had only limited internal validity. Future studies with a more fully developed questionnaire would therefore be of value. A fuller discussion of improvements that can be made towards the current questionnaire can be found under Section 11.8.

The present thesis can only infer that there is a meaningful relationship between several implicit variables and explicit decisions. Probability estimations were calculated that a particular explicit decision was arrived at, determined by judges making certain implicit judgments. However, a methodological limitation here was not being able to objectively and directly measure to what degree explicit innocence and guilt detection comprises an implicit element, and to what extent implicit innocence and guilt detection comprises an explicit element. The placement of the explicit question of whether the appellant is guilty or otherwise at the start of the questionnaire could have further confounded results.

The use of regression was a rudimentary and a gateway method to understand the complexities of how judges' implicit decisions associate with explicit ones. The assumption behind this choice was that implicit decisions judges used will always be correlated with their explicit ones. However, it remains a possibility in Studies 1 to 4 that in certain circumstances judges can indeed detect accurate implicit cues, and yet this does not necessarily translate to higher correct explicit ratings. In other words, there was no way to tell (in certain conditions within the present thesis at least) if judges had actually used an implicit cue

knowingly or unknowingly, meaningfully or otherwise. The findings could also be taken to indicate that judges may not have recognised the association between implicit and explicit decisions in the first place. Moreover, regression does not take into account measurement error. The assumption with regression is that data is measuring perfectly the underlying construct. Path analysis is a promising avenue for exploring the extent to which implicit variable influence explicit ones for future research, as will be elaborated in the section below.

The third methodological limitation was that conceivably television appeals were too short a stimulus material to make any sort of valid judgment and may have further convoluted findings. Consequently, this made it problematic to credit findings to the usage of highly ecologically valid stimulus material, with absolute certainty. Care was taken so that only appeals with a longer content was chosen, but taking into consideration only a limited range of appeals were available this became difficult to always select longer videos. For example, for Study 1's high emotionality condition, there were only so many videos to select from that will fit the study's criteria and definitions of what constitutes high emotional display. Although very highly ecologically valid, televised appeals tend to be short clips with not much information provided on the case, which may further lead to cognitive heuristics processing. Thin-slicing has been found to improve accuracy in some instances and in others, decrease it (Ambady & Rosenthal, 1992; Albrechtsen, Meissner & Susa, 2009; Street & Masip, 2015). In the present thesis, since the video clips shown to judges were very short, indeed it was found in certain conditions innocence and guilt detection accuracy were better than others. Therefore, while using television appeals as stimulus material was beneficial in extracting the use of heuristics and expectations in the face of limited available information, it also presented a limitation by influencing veracity decisions by its very nature (that may not be evident in other types of high stakes sources). Caution must be exercised in generalising the current results to other high stakes situations.

Certain implicit cues were bound to require a more systematic and conscious processing than others, and in the current thesis there was no way of observing this unless participants were asked this directly. Even then,

participants may have found it a difficult task to express or remember their cognitive decisions let alone accurately. While there are benefits in implicit assessments of veracity, challenges also exist. For example, observers may not come to a full understanding in terms of how they came to the assessment or the meaning of it. A case in example, in Anderson et al.'s (1999) study some participants are not able to realise that by them refocusing from the verbal cue of story plausibility to a visual cue of gaze aversion may potentially be an indication of deception. Their participants found that when making implicit veracity assessments it was often difficult to make a parallel explicit decision. While using the questionnaire developed in the present thesis combatted some issues highlighted in the authors' studies (instead of depending on participants' own memory retrieval processes in cue self-reporting), the current questionnaire could benefit from encompassing a fuller range of items to include non-verbal cues as well, as will be expounded in the Suggestions for Future Research section.

Another methodological limitation in the present thesis was excluding participants who did not agree with the operational definitions set out in Studies 1 and 3. For example, in Study 1 those who did not agree that videos selected for the condition of Low Emotionality were indeed so, were excluded from further analysis. While vital that this step was taken, it highlighted subjectivity of perceptual impressions and individual differences, and overlooked a unique group of individuals. In future research, attention should be paid to this group of individuals and observe further into how and why they perceived the way they did. Rather than treating variables as absolutes and disregarding perception that is different from the chosen definition, studies could be designed to consider these variables as relative and/or on a continuum. This could also potentially reduce the artificiality of the experimental design and reduce guesswork in having to choose a veracity response in the face of an uncertain outcome.

In Study 5, it was acknowledged that appeals utilised in this study differed in status as to whether the relative was missing or dead. A mixture of direct appeals and indirect appeals were also used. In addition, the timing of the appeals was not controlled for either. It could be construed that what is said during the first week of an appeal would differ from an appeal made after several months of

the investigation being carried out. While controlling for these issues may or may not offer any differences in results from the ones found in Study 5, it would certainly be useful in regulating and monitoring these questions. In Whelan et al.'s (2014) paper the authors used a balance of 10 cases where the relative's body had been found, meaning the relative was missing and not publicly known to be dead, and also another six cases where the status of the relative was publicly known to be dead. In Study 5, 26 appeals were used for cases where the status of the relative is missing and another 13 for cases in which the body had been found and the relative known publicly to be dead.

As already acknowledged notwithstanding the many advantages of using high stakes data, using these comes with a price which presents limitation to their generalisability. One of these limitations was not being able to control for video quality, where some footage quality were better than others, which was why non-verbal behaviours were not chosen in the present study as a focal point. Judges may have a hard time straining to view behaviours such as 'pupil dilation' with such questionable video resolution. Furthermore, the range of appearance that can be seen in each video would be very much varied, where in some videos only a head can be seen and in others a full body. While the implicit variables of 'body attractiveness' and 'whether the appellant was perceived to be well-dressed' were initially measured in the questionnaire, it was not included in any analyses as in some appeals a body was not within the image range.

An additional limitation in using a high stakes dataset was not being able to control for where the appeals are being made from. All appeals used in the present thesis were collected from the United States, United Kingdom, Australia, Canada and New Zealand, where most appellants were from the Western culture. Thus, it was essential that all participants in the current thesis spoke and understood English. This also presents a possible methodological limitation for the generalisability of the set of results in the present thesis, where certain verbal aspects measured in Study 5 may not be present across other cultures, for example. Furthermore, due to the small number of videos that were presented to participants to counter for maturation effect, this could impact the generalisability of the findings in Studies 1 to 4. Results could also strictly be

idiosyncratic and attributable to the videos chosen, and warrants further replication with a higher number of videos. The difficulty also lies in not being able to find many video clips for certain studies. For example, there were only a small number of video clips showing an innocent appellant who was very emotional/unemotional. Likewise, there were only a small number of guilty appellants who were/were not.

Appeal veracities could not be ignored while selecting the videos for study analyses, and in order to establish ground truth as the basis for selecting these videos. While in Study 5 there was a second inter-rater present, who contributed to the reliability assessment and was ignorant to appeal veracities, in future studies this number of inter-raters could be higher. As such, future studies should endeavour to carry out content analysis in total ignorance of the appellants' veracities with a higher number of inter-rater reliability assessments to prevent the possibility of false positive cues, or cues overlooked.

A final limitation of the studies presented in the current thesis was that because data from all four studies were collected at same time, improvements to the methodology section could not be made succeeding each study. Randomisations of the question item order as well as video clips were also not employed. Although there has been widely cited research studies in lie detection that also did not employ randomisation of video clips such as in Kassin, Meissner and Norwick (2005), Gilbert et al. (1990) and Street and Kingstone's (2016) studies where these authors opted to present them in a constant order, a randomisation future questionnaires would benefit from this improvement to reduce question order effects as well as increase quality of overall data.

### **13.7 SUGGESTIONS FOR FUTURE RESEARCH**

While the present thesis' groundwork results should be cautiously considered, its contribution in understanding subjective perceptions and biases as well as understanding deception in this particular high-stakes context are invaluable. The studies presented in the present thesis grapples with many issues that lie detection research has so far not delved into yet, and warrants further

research. This section highlights suggestions for where certain area can benefit from further research.

Studies 1 to 4 provided some insight into the underlying cognitive processes of judges is now accessible. In the present thesis, these cognitive mechanisms were studied through implicit questions. Implicit judgments were shown to be affected by cues that were emphasised to judges as well as the amount or form of information presented. Judges adapted according to the available information and make decisions accordingly.

Asking participants to make an explicit judgment at the start of the questionnaire and then proceeding to implicit assessments, as well as not allowing for a free mental recall meant that this thesis was not able to study what they truly thought. Implicit judgments could have been influenced by explicit ones. Reverse-ordering explicit and implicit segments in future research is recommended to further explore this relationship. The number of times appeals are watched and changes in explicit answers are also suggested to be reported in forthcoming studies as this can reveal cognitive interplay between explicit and implicit components of lie detection. Further, allowing participants the opportunity to view video clips multiple times and change their answers may have neglected the issue of perceived consistency across multiple viewings. Multiple viewings may have also affected innocence-guilt judgments due to the judge's perceived familiarity with it (Weaver, Garcia, Schwarz & Miller, 2007; Schwarz, Sanna, Skurnik & Yoon, 2007). Schwarz (2015) proposed that people tend to make decisions based on their perceived feelings pertaining to how easy or difficult it is for them to process a stimulus. This ease or difficulty is influenced by factors such as fluency in processing, which then affects how truthful people find this stimulus to be. The association of fluency and repetition heightens perceived familiarity when the stimulus is watched again and again (Weaver et al., 2007; Schwarz et al., 2007). Consequently, this repetition increases perceived truthfulness (Schwarz et al., 2007).

Allowing multiple viewings may have also reduced the processing of new information when existing stereotypes has already been made, affecting consistency across multiple responses as it is more favourable to one's schema. A

schema is a mental shortcut that is used during the process of encoding and retrieving information from memory. The easier it is to access a schema, the more likely to be used as a heuristic. This is akin to a mere exposure effect (Zajonc, 2001). Certain schemas can hold a long-lasting effect because of the emotional arousal experienced when watching or listening to the story reinforces the neural networks that result from that experience (Freeman, 1991; McGaugh, 1989). This results in the activation of this schema later and increases with repeated exposure as well as the rejection of alternative understandings. This may especially be so in the present thesis as the stimulus material is emotional in nature.

Asking an explicit question at the very start of the questionnaire would mean judges would have had the least time to systematically process the very brief video or audio clip they would have just experienced. As such, judgment rates could have been affected by this. In the lie detection literature, Street and Masip (2015) contended that this decrease in truth bias is not attributable to the factor of time, rather, it is a product of simply making more than one veracity judgments. To capture the actual moment a judge arrives at a decision as to whether a sender is innocent or otherwise may have been important and warrants further study in future research. While it is known that the truth bias disappears over time (Masip et al., 2009), it is not known if the emotional bias or the possible innocence-guilt bias outlined in the present thesis will follow the same trajectory, requiring additional study as well as what account will best explain the results (Street & Masip, 2015).

From Part 2 of the present thesis it can be tentatively surmised that there is a relationship between explicit and implicit decisions, perhaps even an ‘unconscious’ segment yet to be methodically eliminated if it does not exist. Street and Richardson (2015) maintained that the unconscious cannot provide an explanation for indirect lie detection. Street and Vadillo (2016) further argued that there is not much existing empirical evidence for (fully) unconscious but accurate lie detection. While of course it is possible that the ‘unconscious’ accuracy increase is really the result of a conscious process (Street & Richardson, 2015), the question remains of whether an unconscious process exists. Even if

unconscious lie detection has come under attack (Franz & von Luxburg, 2015; Levine & Bond, 2014; ten Brinke, Stimson & Carney, 2014), Street and Richardson (2015) did not entirely dispute that the unconscious segment does not exist, rather, that conscious processes can explain accuracy rates in absence of a grounded theory and measurement of the unconscious process. In order to study unconscious lie detection, Street and Vadillo (2016) proposed that a theoretical model of unconscious lie detection is necessary to clarify (and measure) the role of the unconscious in influencing lie detection accuracy, developed from current knowledge.

To note, by directly asking judges a supposed ‘unconscious’ question could be said to be evoking what could have been embedded in the unconscious, into the conscious, so to speak. The thought process, if stemming from an unconscious source, fundamentally surfaces into the conscious mind. How much of this is retained as an automatic and instinctive judgment, and how much of each (the unconscious part and the now cognisant part) plays into the ultimate judgment of whether a sender is lying or not are aspects that need to be distilled in future research.

In addition to this, where the concept of ‘intuition’ or gut feelings fits into this relationship is also presently unclear. Granhag (2006) wrote that in the general lie detection literature, ‘implicit lie detection’ is broadly accepted in the most part due to the immeasurability of one of its component; ‘intuition’ or ‘gut feeling’. While there are insinuations and undertones of this intuitive, unconscious process, there is little empirical evidence of it captured. Intuition may certainly contribute in the process of making implicit judgments for some question items more than others; however, Granhag (2006) proposed that there may not be a strong relationship between the two concepts. While this may or may not be so, it is either important to eliminate it as an extraneous variable or find its place (and relationship with other variables) in Figure 2.

Some dual process theorists do not make a distinction between what is implicit and what is automatic, claiming that both forms of mental processes are under System 1 and thus both are not able to be measured (Stanovich & West, 2000). Conversely, DGC (single system framework) proposes that ‘automatic

reasoning' should be distinguished from 'implicit reasoning' because with automatic reasoning observers can demonstrate what Osman (2000) calls 'meta-knowledge' that can be reliably reproduced as the representations used are strong, stable and distinctive. In contrast, for implicit reasoning observers cannot broach into the meta-knowledge of their thought or conduct, and it cannot be consistently replicated. While it is not disputed that some levels of knowledge are currently inaccessible on the conscious dimension, this shows that there are different levels of awareness observers have into their own mental processes and more importantly there is a prospect of accessing and studying it (Osman, 2000; Siegler & Stern, 1998). Evidently, while this was beyond the scope of the present thesis additional research is required to extricate what distinguishes the implicit from the automatic system of reasoning.

As an alternative or add on, future questionnaires could also benefit from adopting an open-ended response option rather than the close-ended one employed in the present thesis. For the purposes of this thesis, the close-ended questionnaire option was useful and presented important insights as well as to avoid experimental maturity. In future, besides including cues found in Chapter 12, an open-ended response option gives judges a freer rein to accurately and wholly explore their implicit and explicit decisions. Sporer and Schwandt (2007) highlighted that another issue with using questionnaires is that participants may not have instinctively thought about the implicit assessments asked when assessing the veracity of the appellant had a prompt not been available. The think-aloud protocol can prevent this issue and be applied to future research endeavours (Ericsson, 1996; Okada & Simon, 1997). While the implicit judgment questionnaire used in the present thesis is no doubt valuable compared to only using explicit judgments, they may be less able to gather the subtler and refined mental processes.

Think-aloud protocols (TAP), which is used to reveal the strategic decisions an individual uses to come to a decision or when performing tasks, can be used in its place in future studies. It is the most direct and therefore one of the best implements to examine the continuum of cognitive processes when a judgment takes place. There are two types of measuring think-aloud procedures, which are

concurrent and retrospective think-aloud (Kuusela & Paul, 2000). The concurrent protocol is prompted during an experimental task. Here participants usually either state out verbally their thoughts, how they reason and how they feel during the task, or takes a break from the task and state their thoughts, reasoning and feelings at a prompt. In contrast, the retrospective protocol lets participants think-aloud their mental processes when the task is completed while their short-term memory is still fresh. Guan, Lee, Cuddihy and Ramey's (2006) study showed evidence for the validity and reliability of the retrospective think-aloud method. The authors compared their participants' verbalizations with eye movement trackings and accounted in a valid way what they attended to during their tasks. They concluded that the retrospective protocol introduces a very low possibility of fabricating a thought or mental process, and the validity of this method is unaffected by how difficult the task performance is. As with the concurrent protocol, retrospective think-aloud method provides the experimenter with useful evidence of an observer's interpretations, judgments and approaches when performing a task. Ericsson and Simon (1998) established that the think-aloud protocol can be used to study verbalised mental processes and processes that mediate it without altering their course while being able maintain uninterrupted focus on the completion of the experimental tasks. When participants are thinking aloud, their sequences of thoughts have not been found to be systematically altered by verbalisation.

However, Meissner, Brigham and Kelley (2001) and Schooler and Engstler-Schooler (1990) found evidence that thinking aloud can affect and impair the thought process, through a process called 'verbal overshadowing'. Meissner, Brigham and Kelley (2001) and Schooler and Engstler-Schooler (1990) found evidence that thinking aloud can affect and impair the thought process, a process called 'verbal overshadowing'. In the latter study, the authors divided their participants into a 'description group' and a control group. All participants watched a video showing a mock robbery, and only participants in the description group had to describe the suspect in the video. Both groups were then asked to perform a recognition task of the suspect. The authors found that participants who were in the control group performed considerably better in the

recognition task than their counterparts. A similar process to verbal overshadowing called 'articulatory suppression' which impedes the working memory was examined by Murray, Rowan and Smith (1988). Articulatory suppression in this study included participants having to articulate a simple 'the' and spelling a word out aloud while showing a list for them to memorise. The authors found that the more difficult the articulatory suppression was, the more it interfered with participants' encoding of visual information processes. Their memorisation accuracy decreased as the articulatory task became more difficult, because more effort is required for them to remember a task and thus less attention is paid to incoming visual information presented at the same time.

That said, the think-aloud protocol will still benefit future research in studying subjective perceptions in guilt and innocence detection. This method can be particularly informative, especially when judges may expose alternate tactics or cues not thought of by the researcher. One of the advantages of this method in lie detection research is rather than grossly assuming that judges are simply resorting to heuristics, for example, deeper enquiry could perhaps disclose that there was a belief or principle they followed and why is this so. Think aloud protocols offer a higher degree of certainty of discerning whether participants knowingly or unknowingly and meaningfully used a certain implicit cue and whether this fed into their explicit decision (or vice versa). This could potentially bypass the issue of the current questionnaire not being able to capture what implicit cues led to accurate decisions (see Table 30), or at least provide a clearer indication of whether participants arrived at explicit judgments independently of conscious implicit judgments, and whether they arrived at such judgments knowingly or unknowingly. At this stage, it is unclear whether the reason for very few significant cues as highlighted in Table 30 is a result of 1) the type of questions asked, or that 2) whether implicit assessments are in fact not related to explicit decisions at times.

Instead of only using participants' responses that corresponded with the author's definitions (as was done in Studies 1 and 3), a preliminary study involving a group of participants to rate the level of emotionality in these videos could be conducted. Only videos achieving very high consensus among the raters could be

retained, and consequently shown to judges. This could avoid the correlational issue between variables as put forward in Section 7.9 as well as promote a more meaningful interpretation of results. Another suggestion for future research is to compare responses of judges who did not agree with the author's definitions of a certain variable (i.e., level of emotionality of appeals) to those who did, to examine why these differences in thinking exist and how this interrelates with veracity decision-making.

While a judgment of innocence-guilt and a judgment of truth-lie are not entirely identical, in the context of judging a brief video such as in the present thesis it appears that the two are linked. Furthermore, with someone having no other extra information on the case (unlike most jury trials, where a juror would not have to only rely on an appeal video to make a judgment of guilt or innocence of the subject with no other evidence or information given to them) in the experiments set out here in the present thesis they must rely heavily on the appeal video itself. In the typical criminal trial setting, jurors typically receive a great deal of evidence and usually over a longer period of time, sometimes even for days and weeks (Pennington & Hastie, 1992). One model put forward is that jurors process this evidence as they receive it, assess this evidence and try to fit it into a story form that they can understand (Pennington & Hastie, 1990). Pennington and Hastie (1990) proposed that this explanations-based model is relevant when there is a considerable amount of complex, interconnected pieces of evidence which are pertinent to deciding verdict. In the current thesis, it was useful to observe how participants adapted and made decisions of the appellant's guilt or innocence with virtually no other information about the appellant or the case. However, it is recommended that in future studies, if the research interest is to be more reflective of a legal trial procedure, for more information and/or evidence to be given to a participant in order for them to make an innocent or guilty verdict towards a sender or an account in question.

Study 3 raises the inquiry of measuring cognitive loading, as cognitive load was surmised to have taken place and would imaginably be more robust if measured directly. A more direct method to objectively measure cognitive loading is the use of a dual-task analysis, which posits that when participants have to

perform two separate tasks at the same time, and particularly if requires the same resources in verbal or visual working memory then the available resources would have to be distributed between the two tasks (Brünken et al., 2003). The performance level in both tasks would be the outcome variable with the hypothesis being lower performance in the dual-task group compared to the single-task one. Yet another method is a self-reported stress level and/or self-reported difficulty of materials presented (Brünken et al., 2003). Suggestions for future studies will include a more direct method of objectively measuring cognitive load.

Study 4 indicates that perception and subjective impressions of when more than one person lying is different from when an individual person lies. Added to this, most crimes occur as a group and in a group. Vrij et al. (2010) proposed that it is crucial to study liars and truth-tellers in pairs as it reflects reality in the criminal world, where criminals usually work in groups or networks and not by themselves. Many crimes are planned and committed by several offenders working together rather than one person acting alone (Weerman, 2001). Further studies must be conducted with more than one liar present in high-stakes deception, as little is still understood of the processes that take place when corroborating a lie together, as well as the implicit assessments made by judges other than the questions asked in the present study. For example, group lying and the actual collaboration of a lie should be observed and with that a comparison with truthful collaborative recall procedures be observed.

Furthermore, Study 4 introduced a suggestion for development that is beyond the scope of the present thesis. Limitations of the current stimulus material include not being able to study a conversation between two liars for example, as television appeals are a unique stimulus material. Interactive and non-interactive deception is fundamentally different in important ways (Burgoon, 2015a). Burgoon and Buller (1994) provide several interactional characteristics where liars may display such as higher levels of uncertainty, vagueness, non-immediacy and reticence, negative affect, arousal and non-composure, formal and submissive behaviour and generally a poor impression. While television appeals cannot be used to investigate this interpersonal deception as they are simply recordings with no observer interaction, this calls for studies of group or dyadic

lying behaviour. In this context, a clear and actual interpersonal deception never took place as participants in the present study were not directly being lied to. Lying is, at the end of the day, a social behaviour; it happens mostly during an interaction and within the transaction between two individuals. As identified by the interpersonal deception theory, lying is a dyadic and interactive event (Burgoon & Buller, 1994). Rather than focusing on very short abstracts of behaviour or even observing responses after watching these short abstracts in artificial settings, IDT proposes changes over time in patterns of dynamic and unique within-pair interaction and judgments (rather than the broad between group differences such as that employed in the present thesis).

Due to the confines of the present thesis, studies presented here largely ignored non-verbal aspects of guilt and innocence detection in television appeals. While verbal cues are valuable to be studied as shown in the present thesis, the importance of non-verbal implicit cues (as well as those that are neither verbal nor non-verbal) cannot be overlooked. Ten Brinke and Porter (2011) who used television appeals could not find a single reliable indicator in body language to distinguish truthful from false appellers (although they did find reliable cues in the emotional expressions facet). Of note, certain implicit cues are not straightforwardly classifiable, as they are along both the verbal and non-verbal continuum. Granhag (2006) extended an example of this type of cue to be the magnitude of perceived sender cooperation. While many of the implicit cues examined in the present thesis yielded significant differences future questionnaire items still need to include other cues perhaps not captured in this questionnaire. Future studies' question items should be expanded to include non-verbal cues such as 'demeanour' (i.e. the sender was laid back when telling the story, the sender was not very comfortable when telling the story), 'gaze aversion', 'head shaking' and 'illustrators' to name a few, as well as cues that do not fit in a particular category such as 'they stated that from the beginning' (Anderson et al., 1999; Vrij & Mann, 2001a; Zuckerman et al., 1981; Vrij et al., 2008). The innocence score items discussed in the present thesis could be expanded to include other verbal aspects not currently incorporated such as story complexity, story flow, restatements and story repetitiveness (Anderson et al., 1999). This way, both

verbal, non-verbal implicit judgments and everything in between can be studied with some cues pertaining to a liar and truth-teller's verbal behaviour, some others pertaining more to their non-verbal behaviour, and still some others which lie in between the continuum.

As suggested in the conclusion section of Chapter 12, the consideration of a combination of cues increases accuracy in distinguishing lies from truths as it has been established that there are no universal cues to deceit (DePaulo et al., 2003; Whelan, 2014; Hartwig & Bond, 2014; Burgoon, personal communication, 25<sup>th</sup> August, 2015; ten Brinke & Porter, 2012; Vrij & Mann, 2004). Hence, the possibility exists that considering multiple factors or cues tested across Studies 1 to 4 collectively would yield higher accuracies than recorded in the present thesis. In keeping with this, studies such as that of Ekman, O'Sullivan, Friesen and Scherer (1991) and Leal et al. (2010) reported accuracy rates as high as 80% using a case by case system of individual cases as opposed to observing the variances between two conditions. Rather than observing the frequency of single cues between groups, these studies categorised single subjects as innocent or guilty based on the occurrence of multiple cues. In line with this, methods of analysing data such as generalised estimating equations (GEE) can be employed to model lie detection accuracy using a correction of estimated standard errors to account for within-cluster correlation such as in Sweeney and Ceci's (2014) study. Thus, for example, if a participant watched 10 video clips the data can be clustered as such that each participant is their own cluster (Norton, Bieler & Ennett, 1996).

As stated in the very beginning of the section above, there was no room in the present thesis for studying the objective verbal aspects found with observer reactions to these implicit judgments. Importantly, the findings from Chapter 12 highlights 'true implicit cues' that are diagnostic in identifying truthful appealers from their counterparts, therefore, these cues should be included in the development of future questionnaires. These items include 'the appealers seem to imagine the victim', 'the appealers distance themselves/herselves from the victim', as well as the one reliable indicator found in ten Brinke and Porter's (2011) study of 'tentative words'.

### 13.8 PRACTICAL APPLICATIONS AND IMPLICATIONS

As reiterated throughout the present thesis, one of the fundamental premises was using highly ecologically valid stimulus material in the hopes that results presented will implicate a higher practical application to real life police investigations. As of yet, there are no formal recommendations as to how these appeals should be conducted, how they should be televised and how ought these appeals help police officers with their investigation. While there are police manuals for lie detection (Inbau et al., 2001) and police interrogation manuals (Gudjonsson, 2003), a specific standardised procedure for conducting television appeals have yet to be put forward. In fact, existing police manuals on lie detection are wrought with erroneous information on how to detect deceit (Kassin & Fong, 1999; Mann, Vrij & Bull, 2004). Kassin et al. (2005) found that special 'training' in lie detection given to police investigators (with an average of nearly 11 years of experience) actually reduced the accuracy to accurately identify a truthful confession from one that is not. Conversely, Forest, Feldman and Tyler (2004) maintain that being knowledgeable about valid and diagnostic cues to deception makes an individual better and more accurate in discriminating truths from lies. The authors found that the higher the accuracy their participants were in answering questions pertaining to their beliefs and perceptions of cues to deception, the better they accomplished in distinguishing a truth from a lie.

The need for a general guideline specifically for televised appeals is important because as of yet, solvability of missing or murdered relatives' cases based on the airing of the appeal itself is unknown. The extent to which these appeals are of help to the investigation is unidentified; whether investigations take longer or shorter to solve compared to cases where public appeals are not made is yet to be revealed. As of yet, it is not known whether television appeals help police investigators in solving these cases of missing or murdered relatives, while holding the potential to do so.

Sporer and Schwandt (2007) asserted that before being able to be applied practically, beliefs about observed behaviours need to first be examined and understood. They posited that it is important to find out whether these cues are valid and diagnostic or otherwise. While more research in this area is necessary

to be distilled as these studies can only imply a correlation and not causation thus far, these initial results would seem to suggest that judges think differently when presented with different manipulated conditions. These studies were yet unable to encapsulate the progression of cognitive processes involved in explicit and implicit decisions but were able to present a snapshot of them. The studies in the present thesis seem to support an adaptive decision-making tendency (Street, 2015) rather than an uncontrollable mental default in believing others (Gilbert, 1989; Gilbert et al., 1990). Further investigation into how people process, encode and manage information and how that translates to them making a guilt-innocence and/or a truth-lie judgment can either support these initial suggestions or challenge them.

Although firm conclusions cannot yet be made at this stage, the results from the present thesis were valuable in providing an idea of the complex underlying process of decision making when presented with high-stakes stimulus materials. For example, the results from Study 2 indicated that better innocence-guilt detection can potentially be achieved by judges when no visual element is presented. This replicated Bond and DePaulo's (2006) findings in their meta-analyses where people were less accurate in judging lies they could visibly see, compared to audible-only lies. When a visual element was presented judges tended to use invalid and un-diagnostic cues for veracity assessments. A practical suggestion for future investigation and airings of such cases would include police officers instructing only one appellant to appeal rather than a pair or several at a time, with police officers not included in introducing the appeal (as this would generate biases of its own). By making aware these possible perceptions and biases, not only can these results be used to aid in how these appeals are presented but in also notifying police officers of cues to watch for when studying the appeal contents.

Another implication lies in how appellants present themselves, and instructions given to courtroom juries. Sporer and Schwandt (2007) contend that in many legal proceedings jurors are asked to or will observe the behaviours of a defendant or witness on trial to gather evidence for deception or truthfulness. When this observation is carried out with invalid expectancies and erroneous

beliefs, consequences can be grave. As exemplified in the case of Amanda Knox, and to some extent in Study 1, even honest grieverers can be mistakenly judged to be lying; these implications also extend beyond television appeals. If the conjecture in Study 1 is true, subject to replication with a higher number of videos and trials, a lack of crying or overtly grieving behaviours displayed when making a television appeal or during trial can be misinterpreted and seen as a defendant who is guilty. An awareness and/or instruction of how emotions displayed may present as an invalid cue of veracity judgments will aid juries to consciously attempt in disregarding this cue when making decisions of guilt or innocence, enabling juries to reach a fairer and more accurate outcome. The results in Study 3 seemed to suggest that biases persist in how appellars present themselves and in how they deliver their speech. An awareness and/or instruction to members of the jury as to be cautious in writing off an incoherent speech as a valid indicator of guilt and vice versa will enable juries in attaining a more objective verdict.

The eight aspects presented in Study 5 in accordance with the hypothesised role of emotion and indicators of grief, can hopefully aid police officers in guiding considerations to certain cues in an appellant's speech to predicting truthful from false ones. The present thesis proposes cues such as verbally declared emotions (or the lack of), the declaration of needing social and communal help, activity and movement in appeal, seeking religion, maintaining hope, speaking about the victim in present tense, denial of involvement, conditional markers in regard to the return of the missing relative, declaring presence near or at the time of murder and referencing the event to be helpful indicators in distinguishing a true from a deceptive appeal. This presents a simple way to guide police officers in charge of these investigations where little to no training is pre-required. These findings are potentially relevant also in similar research areas in other high stakes scenarios in terms of cues that emerged.

### **13.9 CODA**

The present thesis distinguishes itself from both ten Brinke and Porter's (2012) and Whelan et al.'s (2014) work in two ways. Ten Brinke and Porter's (2012) research, while using the same stimulus material as the present thesis',

focused more on investigating the behavioural consequences of these pleaders and how that relates to differences in verbal and non-verbal cues including facial expressions between truth-tellers and liars. Similarly, Whelan et al.'s (2014) research examined behavioural cues discriminating truth-tellers and liars. These works consider psychological theories, such as distancing and cognitive load, to help explain why their findings where liars' speech, behaviour and facial expressions differ from truth-tellers (ten Brinke & Porter, 2012). Study 5 in the present thesis supported these differences using a novel method in analysing a similar dataset. In addition, Studies 1 to 4 considered the underlying processes of making innocence-guilt judgment from a judge's point of view, rather than a sole focus on identifying diagnostic and discriminative cues between truthful appealers and deceptive ones. The present thesis also discussed the implications of cognitive models such as different classes of dual process theories (Gilbert et al., 1990; Gilbert et al., 1993; Sloman, 1996; Sloman, 2002; Reinhard & Sporer, 2008; Reinhard & Sporer, 2010; Evans, 2007) and the single process system (Cleeremans & Jiménez, 2002). While a firm conclusion as to which model best explains the mental development of making explicit and implicit decisions, the present thesis offers a step forward towards clarifying this process.

While Part 2 investigated subjective perceptions of these television appeals, Part 3 concerned developing theoretical frameworks linking the more objective aspects of verbal cues and how these cues can be studied in organisation with each other. This study not only found verbal cues discriminated between truthful and false television appealers but also generated verbal profiles from the relationship of these cues in accordance with each appealers, which is valuable in an applied context. Both objective findings and subjective assessments may operate as a guide to not only raising awareness in biases that may surface in a practical context, but also serve as a checklist to law enforcement agents when airing an upcoming television appeal in a future missing or murdered relatives' case.

The approach that best fits the results presented in the present thesis is the ALIED approach (Street, 2015), where the existence and trend of these biases seems to be conditional upon context. The results from Part 2 would appear to challenge the existence of any intrinsic tendency towards biases, as initially

anticipated in the earlier chapters of the present thesis. Rather, results would suggest that biases that surfaced are consequences of an adaptive decision-making mode when confronted with little or unconstructive information about the Source.

The extent to which implicit cues contain an explicit element and the extent to which explicit cues contain an implicit element remains a subject to explore. This thesis presents a first step in understanding this complex interaction which was largely missing in lie detection literature. With regression analyses, implicit judgments can be construed at this stage to have a certain degree of association with explicit decisions, although taking into consideration that they could have been influenced by explicit decisions in the first place. Standing theories in current lie detection research are subject to be further developed, supported or challenged based on whether certain cognitive processing is a default (therefore unconscious) or not, for example.

Additionally, this thesis presents the first steps in uncovering if an innocence or guilt bias persists using high-stakes scenarios. While results here would tend to suggest that such biases are not consistent (i.e., people seem to be also poor in telling apart innocence from guilt as they are in detecting truths from lies), more research is needed to uncover the overlap and distinctions between the concepts of truth-lie and innocence-guilt. The presence or absence of an innocence bias (and its link with other person perception judgments), if can be further developed, supported or challenged, would carry implications in the psycho-legal arena. Future studies would benefit more from adopting an experimental setting that simulates a criminal trial more closely (i.e., presenting judges with more evidence regarding the case before having to make an innocence-guilt judgment where the prosecution holds the burden of proof, for example) (Helgeson & Shaver, 1990; Pennington & Hastie, 1993).

## REFERENCES

---

- Adams, S. (2002). *Communication under stress: indicators of veracity and deception in written narratives* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University, Virginia.
- Adèr, H. J., Mellenbergh G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Huizen, Netherlands: Johannes van Kessel Publishing.
- Aitchison, J. (2008). *The Articulate Mammal*. New York, NY: Routledge.
- Akehurst, L., Köhnken, G., & Hofer, E. (2004). Content credibility of accounts derived from live and video presentations. *Legal and Criminological Psychology, 6*, 65–83.
- Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology, 10*, 461–471.
- Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance? *Journal of Experimental Social Psychology, 45*, 1052–1055.
- Alhakami, A. S., & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis, 14*(6), 1085–1096.
- Amar, R., & Toledano, S. (2001). *HUDAP manual*. Jerusalem: The Hebrew University of Jerusalem.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256–274. doi: 10.1037/0033-2909.111.2.256.
- Ambady, N., & Weisbuch, M. (2010). Nonverbal behaviour. In D. T. Gilbert, S. t. Fiske, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed). Hoboken, NJ: John Wiley and Sons.
- American Psychiatric Association. (1994). *Bereavement*. In: Diagnostic and Statistical Manual of Mental Disorders, *Fourth Edition*. Washington, DC: American Psychiatric Association.

- Anderson, N. H. (1965). Averaging versus Adding as a Stimulus-Combination Rule. Impression Formation. *Journal of Experimental Psychology*, 70(4), 394–400.
- Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom. *Journal of Nonverbal Behavior*, 23, 67–89.
- Armstrong, J. S. (2012). Illusions in Regression Analysis. *International Journal of Forecasting*, 28(3), 689–694. doi: 10.1016/j.ijforecast.2012.02.001ro.
- Aronson, E., Wilson, T. D., & Akert, R. M. (2010). *Social Psychology* (7th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Ask, K., & Landström, S. (2010). Why Emotions Matter: Expectancy Violation and Affective Response Mediate the Emotional Victim Effect. *Law and Human Behavior*, 34, 392–401. doi: 10.1007/s10979-009-9208-6.
- Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421.
- Bal, P. M., & Veltkamp, M. (2013). How Does Fiction Reading Influence Empathy? An Experimental Investigation on the Role of Emotional Transportation, *Plos One*, 8(1): e55341. doi: 10.1371/journal.pone.0055341.
- Baldry, A. C., Winkel, F. W., & Enthoven, D. S. (1997). Paralinguistic and nonverbal triggers of biased credibility assessments of rape victims in Dutch police officers: An experimental study of “nonevidentiary” bias. In S. Redondo, V. Garrido, J. Perze, & R. Barbaret (Eds.), *Advances in psychology and law* (pp. 163–174). Berlin: de Gruyter.
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, 20(2), 395–418.
- Baron, R. A., Branscombe, N. R., & Byrne, D. E. (2008). *Social psychology* (12th ed.). Boston, MA: Pearson.
- Baudet, S., Jhean-Larose, S., & Legros, D. (1994). Coherence and truth: A cognitive model of propositional truth attribution. *International Journal of Psychology*, 29, 219–350.
- Berry, D., & Dienes, Z. (1993). *Implicit learning*. Hove, Brighton: Erlbaum.

- Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–362). New York, NY: Wiley.
- Bjork, J. P., & Thurman, J. W. (2007). Negative life events, patterns of positive and negative religious coping, and psychological functioning. *Journal for the Scientific Study of Religion*, *46*(2), 159–167.
- Boduszek, D. (2017). *Introduction to structural equation modelling* [PowerPoint slides]. University of Huddersfield, UK.
- Boelen, P. A., van den Bout, P., & van den Hout, M. A. (2003). The role of negative interpretations of grief reactions in emotional problems after bereavement. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 225–238. doi: 10.1016/j.jbtep.2003.08.001.
- Bogaard, G., Meijer, E., Vrij, A., Broers, N. J., & Merckelbach, H. (2014). Contextual bias in verbal credibility assessment: Criteria-Based Content Analysis, Reality Monitoring and Scientific Content Analysis. *Applied Cognitive Psychology*, *28*, 79–90.
- Bollingmo, G. C., Wessel, E. O., Eilertsen, D. E., & Magnussen, S. (2008). Credibility of the emotional witness: A study of ratings by police investigators. *Psychology, Crime & Law*, *14*(1), 29–40.
- Bollingmo, G. C., Wessel, E. O., Sandvold, E., Eilertsen, D. E., & Magnussen, S. (2009). The effect of biased and non-biased information on judgments of witness credibility. *Psychology, Crime and Law*, *15*(1), 61–71.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. doi: 10.1207/s15327957pspr1003\_2.
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, *134*(4), 477–492.
- Bond, C. F., Levine, T. R., & Hartwig, M. (2015). New findings in nonverbal lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 37–58). Chichester: Wiley.
- Bond, C. F., & Rao, S. R. (2004). Lies travel: Mendacity in a mobile world. In P.

- A. Granhag and L.A. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 127–147). Cambridge: Cambridge University Press.
- Borg, I., & Shye, S. (1995). *Facet theory: Form and content*. Newbury Park, CA: Sage.
- Bornstein, G., Kocher, T., Kugler, T., & Sutter, M. (2002). Whom do you trust more in bargaining - Groups or individuals? *Working Paper*.
- Bowlby, J. (1961). Processes of mourning. *International Journal of Psychoanalysis*, *42*, 317–339.
- Breen, L. J., & O'Connor, M. (2007). The fundamental paradox in the grief literature: a critical reflection. *Omega (Westport)*, *55*(3), 199–218. doi: 10.2190/OM.55.3.c.
- Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, *20*(2), 116–120. doi: 10.1016/j.cub.2009.11.034.
- Bruinsma, G., & Bernasco, W. (2004). Criminal groups and transnational illegal markets. *Crime, Law and Social Change*, *41*, 79–94. doi: 10.1023/B:CRIS.0000015283.13923.aa.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53–61.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: The University of Chicago Press.
- Buller, D. B., & Burgoon J. K. (1996). Interpersonal Deception Theory. *Communication Theory*, *6*(3), 203–242.
- Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal deception: Accuracy in deception detection. *Communication Monographs*, *61*, 303–325.
- Burgoon, J. K., Blair, P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*(4), 572–599. doi:10.1111/j.1468-2958.2008.00333.x.
- Burgoon, J. K. (2015, August 25). Personal communication.
- Burgoon, J. K. (2015a). Rejoinder to Levine, Clare et al.'s Comparison

- of the Park–Levine Probability Model Versus Interpersonal Deception Theory: Application to Deception Detection. *Human Communication Research*, 41(3), 327–349. doi:10.1111/hcre.12065.
- Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception II: The inferiority of conversational participants as deception detectors. *Communication Monographs*, 58, 25–40.
- Canter, D. V. (2004) A partial order scalogram analysis of criminal network structures. *Behaviormetrika*, 31(2), 131–152.
- Canter, D. V., Grieve, N., Nicol, C., & Benneworth, K. (2003). Narrative plausibility: the impact of sequence and anchoring. *Behavioral Sciences & the Law*, 21(2), 251–267.
- Canter, D. V., & Heritage, R. (1990). A multivariate model of sexual offences behaviour: developments in offender profiling. *Journal of Forensic Psychiatry*, 1(2), 185–212.
- Canter, D. V. & Youngs, D. (2009). *Investigative Psychology: Offender Profiling and the Analysis of Criminal Action*. Hoboken, NJ: John Wiley & Sons.
- Carlucci, M. E., Compo, N., S., & Zimmerman, L. (2013) Lie detection during high-stakes truths and lies. *Legal and Criminological Psychology*, 18(2), 314–323.
- Carter, J. (1999). Collaboration or plagiarism: what happens when students work together. In ITiCSE '99: Proceedings of the 4th annual SIGCSE/SIGCUE ITiCSE 80 conference on Innovation and technology in computer science education (pp. 52–55). New York, NY, USA: ACM Press.
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and Systematic Information Processing within and beyond the Persuasion Context. In J. S. Uleman, & J. A. Bargh (Eds.), *Unintended Thought* (pp. 212–252). New York, NY: Guilford.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10(2), 151–170.

- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In Bad Taste: Evidence for the Oral Origins of Moral Disgust. *Science*, 323(5918), 1222–1226. doi: 10.1126/science.1165565.
- Chatterjee, S., & Hadi, A. S. (2012). *Regression Analysis by Example*. Hoboken, NJ: Wiley-Blackwell.
- Chungh, G. (2010). *The Correlates of Detecting Truthfulness and Deceit: An Exploratory study of False and Genuine Public Appeals for Information*. (Unpublished doctoral dissertation). University of Huddersfield, United Kingdom.
- Cleeremans, A. X. (2006). Conscious and unconscious cognition: A graded, dynamic perspective. In Q. Jing, M. Rosenzweig, G. d'Ydewalle, H. Zhang, H. C. Chen, & K. Zhang (Eds.), *Progress in psychological science around the world: Vol. 1. Neural, cognitive, and developmental issues* (pp. 401–418). Hove, England: Psychology Press.
- Cleeremans, A. X., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamical perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making?* (pp. 1–40). Hove, UK: Psychology Press.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. doi: 10.1177/001316446002000104.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cohen, D., Beattie, G., & Shovelton, H. (2010). Nonverbal indicators of deception: How iconic gestures reveal thoughts that cannot be suppressed. *Semiotica*, 2010(182), 133–174.
- Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty & deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, 45(6), 1321–1324.

- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2001). Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychological Science, 15*(10), 687–693.
- Connell, L., & Keane, M. T. (2006). A Model of Plausibility. *Cognitive Science, 30*, 95–120.
- Council of Europe. (1988). Protocol to the Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) as amended by Protocol No. 11. *Council of Europe Treaty Series 155*. Strasbourg: Council of Europe.
- Costa, B. M., Hall, L., & Stewart, J. (2007). Qualitative exploration of the nature of grief-related beliefs and expectations. *Omega: Journal of Death and Dying, 55*, 27–56. doi:10.2190=CL20-02G6-607R-8561.
- Costa, B. M., & Stewart, J. (2007). Expectations about grief and evaluation of the griever. *Omega (Westport), 55*(1), 27–56.
- Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika, 16*(3), 297–333.
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests, *Psychological Bulletin, 52*(4), 281–302.
- Culhane, S. E., Kehn, A., Hatz, J., & Hildebrand, M. (2015). Are two heads better than one? Assessing the influence of collaborative judgements and presentation mode on deception detection for real and mock transgressions. *Journal of Investigative Psychology and Offender Profiling, 12*(2), 158–170.
- Dahl, J., Enemo, I., Drevland, G. C. B., Wessel, E., Eilertsen, D. E., & Magnussen, S. (2007). Displayed emotions and witness credibility: a comparison of judgements by individuals and mock juries. *Applied Cognitive Psychology, 21*(9), 1145–1155. doi:10.1002/acp.1320.
- Dando, C., & Bull, R. (2011). Maximising opportunities to detect verbal deception: Training police officers to interview tactically. *Journal of Investigative Psychology and Offender Profiling, 8*(2), 189–202.
- Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35*(10), 867–881.

- Dart, A. (2017). Reporting multiple regression in APA style: Part 1. *AndrewDart.co.uk*. Retrieved from <http://www.adart.myzen.co.uk/reporting-multiple-regressions-in-apa-format-part-one/>.
- Davis, M. S. (1971). That's Interesting: Towards a Phenomenology of Sociology and a Sociology of Phenomenology, *Philosophy of the Social Sciences*, 1(4), 309–344.
- Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior*, 29(6), 683–704.
- DePaulo, B. M., Epstein, J. A., & Wyer, M. M. (1993). Sex Differences in lying: How women and men deal with the dilemma of deceit. In M. Lewis & C. Saarni (Eds.), *Lying and Deception in Everyday Life* (pp. 126–174). New York, NY: Guilford Press.
- DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.), *Credibility assessment* (pp. 51–70). Dordrecht, Netherlands: Kluwer Academic.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129(1), 74–118.
- DePaulo, B. M., & Morris, W. L. (2004). Discerning lies from truths: Behavioural cues to deception and the indirect pathway of intuition. In P. A. Granhag, & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 15–40). Cambridge, UK: Cambridge University Press.
- DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16(3), 249–267.
- DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323–370). New York, NY: McGraw-Hill.
- DePaulo, B. M., Tang, J., & Stone, J. I. (1987). Physical attractiveness and skill at detecting deception. *Personality and Social Psychology Bulletin*, 13(2), 177–187.

- DeTurck, M. A., & Miller, G. R. (1985). Deception and arousal: Isolating the behavioural correlates of deception. *Human Communication Research*, *12*(2), 181–201.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290. doi:10.1037/h0033731.
- Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. Hoboken, NJ: John Wiley & Sons, Ltd.
- Doss, E. (2002). Death, art and memory in the public sphere: The visual and material culture of grief in contemporary America. *Mortality*, *7*(1), 63–82.
- Driskell, J. E., Salas, E., & Driskell, T. (2012). Social Indicators of Deception. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*, 577–588. doi: 10.1177/0018720812446338.
- Dulaney, E. F. (1982). Changes in language behaviour as a function of veracity. *Human Communication Research*, *9*, 75–82. doi:10.1111/j.1468-2958.1982.tb00684.x.
- Dunn, O. (1961). Multiple Comparisons Among Means Author(s): Olive Jean Dunn Source. *Journal of the American Statistical Association*, *56*(293), 52–64.
- Duran, N. D., Dale, R., Kello, C. T., Street, C. N. H., & Richardson, D. C. (2013). Exploring the movement dynamics of deception. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00140.
- Eagly, A. H., & Chaiken, S. (1995). *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.
- Efran, M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, *8*, 45–54.
- Efron, B. (1979). Bootstrap methods: another look at the jack-knife. *Annals of Statistics*, *7*(1), 1–26.
- Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- Ekman, P. (1977). Biological and cultural contributions to body and facial

- movement. In J. Blacking (Ed.). *Anthropology of the Body* (pp. 34–84). London: Academic Press.
- Ekman, P. (1981). Mistakes When Deceiving. *Annals of the New York Academy of Sciences*, *364*, 269–278. doi:10.1111/j.1749-6632.1981.tb34479.x.
- Ekman, P. (1989). Why lies fail and what behaviors betray a lie. In J. Yuille (Ed.), *Credibility assessment* (pp. 71–81). Netherlands: Springer.
- Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, *6*, 169–200.
- Ekman, P. (1996). Why don't We Catch Lairs? *Social Research*, *63*(3) 801–817.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *20*, 288–298.
- Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*, 414–420.
- Ekman, P., Friesen, W. V., & Scherer, K. R. (1976). Body movement and violence pitch in deceptive interaction. *Semiotica*, *16*(1), 23–27.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, *10*(3), 263–266.
- Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Face, voice and body in detecting deceit. *Journal of Nonverbal Behaviour*, *15*(2), 125–135. doi: 10.1007/BF00998267.
- Engel, G. L. (1975). The death of a twin: mourning and anniversary reactions. *International Journal of Psychoanalysis*, *56*(1), 23–40.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724.
- Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and game!* (pp. 1–50). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Evans, J. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*(1), 5–31. doi:10.1080/13546783.2011.637674.
- Evans, J. (2007). On the resolution of conflict in dual process theories of

- reasoning. *Thinking & Reasoning*, 13, 321–339.
- Farah, M. J. (1994). Visual perception and visual awareness after brain damage: A tutorial overview. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 37–76). Cambridge, MA: MIT Press.
- Fay, P. J., & Middleton, W. C. (1941). The ability to judge truth-telling, or lying from the voice transmitted over a public address system. *The Journal of General Psychology*, 24, 211–215.
- Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64(3), 657–674.
- Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, 20(2), 199–223.
- Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage Publications.
- Field, A. (2000a). Mann-Whitney test. *Research Methods 1: SPSS for Windows part 3: Nonparametric tests*. Retrieved from <http://www.statisticshell.com/docs/nonparametric.pdf>.
- Field, N. P., & Filanosky, C. (2010). Continuing bonds, risk factors for complicated grief, and adjustment to bereavement. *Death Studies*, 34(1), 1–29.
- Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, 2006, 291–397.
- Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science*, 11(5), 154–158.
- Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13, 1–17.
- Fiske, S. T., & Taylor, S. E. (1984). *Social cognition* (1st ed.). Reading, MA: Addison-Wesley.

- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*(1), 39–66.
- Frank, M. G., Paolantonio, N., Feeley, T. H., & Servoss, T. J. (2004). Individual and Small Group Accuracy in Judging Truthful and Deceptive Communication. *Group Decision and Negotiation*, *13*, 45–59.
- Franz, V. H., & von Luxburg, U. (2015). No evidence for unconscious lie detection: A significant difference does not imply accurate classification. *Psychological Science*, *26*(10), 1646–1648. doi: 10.1177/0956797615597333.
- Freeman, W. J. (1991). The Physiology of Perception. *Scientific American*, *264*(2), 78–85.
- Freud, S. (1917). Mourning and melancholia. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 14). London: Hogarth Press.
- Frost, J. (2013, May 30). Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit? Retrieved from <http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>.
- Frost, J. (2015, September 3). The Danger of Overfitting Regression Models [The Minitab Blog]. Retrieved from <http://blog.minitab.com/blog/adventures-in-statistics-2/the-danger-of-overfitting-regression-models>.
- Garrett, B. L. (2008). Judging Innocence. *Columbia Law Review*, *108*, 55–141.
- Gervais, W. M., & Henrich, J. (2010). The Zeus Problem: Why Representational Content Biases Cannot Explain Faith in Gods. *Journal of Cognition and Culture*, *10*, 383–389.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.
- Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). Simple

- Heuristics That Make Us Smart. New York, NY: Oxford University Press.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York, NY: Guilford.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*(4), 601–613.
- Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read our emotional states. *Journal of Personality and Social Psychology*, *75*, 332–346.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.
- Gottschalk, L. A., & Gleser, G. C. (1969). *The Measurement of Psychological States through the Content Analysis of Verbal Behavior*. Berkeley, CA: University of California Press.
- Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminal's beliefs about deception. *Legal and Criminological Psychology*, *9*, 103–119.
- Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, *14*(3), 189–200. doi: 10.1080/10683160701645181.
- Granhag, P. A., Rangmar, J., & Strömwall, L. A. (2015). Small Cells of Suspects: Eliciting Cues to Deception by Strategic Interviewing. *Journal of Investigative Psychology and Offender Profiling*, *12*, pp. 127–141. doi: 10.1002/jip.1413.
- Granhag, P. A., & Strömwall, L. A. (1999). Repeated Interrogations – Stretching the Deception Detection Paradigm. *Expert Evidence*, *7*(3), 163–174. doi: 10.1023/A:1008993326434.
- Granhag, P. A., & Strömwall, L. A. (2000). Deception detection: Examining the consistency heuristic. In C. M. Breur, M. M. Kommer, J. F. Nijboer & J. M. Reintjes (Eds.), *New trends in criminal investigation and evidence* (Vol. 2, pp. 309–321). Antwerpen, Belgium: Intresentia.

- Granhag, P. A., & Strömwall, L. A. (2004). *The Detection of Deception in Forensic Contexts*. Cambridge University Press, Cambridge, UK.
- Granhag, P. A., Strömwall, L. A., & Jonsson, A-C. (2003). Partners in Crime: How Liars in Collusion Betray Themselves. *Journal of Applied Social Psychology, 33*(4), 848–868.
- Granhag, P. A., Strömwall, L. A., & Reiman, A. C. (2002). Partners in crime: How liars in collusion betray themselves. *Manuscript submitted for publication*.
- Granhag, P. A., Vrij, A., & Verschuere, B. (2015). *Detecting Deception: Current Challenges and Cognitive Approaches*. Chichester, UK: John Wiley & Sons, Ltd.
- Griffel, A. (1999). Countermeasure Analysis: A Strategy for Evaluating and Ranking Safety Recommendations.  
<http://dx.doi.org/10.3389/fpsyg.2013.00068>.
- Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, (Eds.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1253–1262) New York, NY: ACM Press.
- Gudjonsson, G. H. (2003). *The Psychology of Interrogations and Confessions: A Handbook*. Hoboken, NJ: John Wiley & Sons.
- Gumpert, C. H., Lindblad, F., & Johansson, A. (1999). Child Sexual Abuse: Expert Testimony in Swedish District Courts. *Child Maltreatment, 4*(4), 343–352. doi: 10.1177/1077559599004004007.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for configuration of points. *Psychometrika, 33*(4), 465–506.
- Harpster, T., Adams, S. H., & Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence: An exploratory analysis. *Homicide Studies, 13*(1), 69–93.
- Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-

- analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659. doi: 10.1037/a0023589.
- Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661–676. doi: 10.1002/acp.3052.
- Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: on the possibility of suspending belief. *Psychological Science*, *16*(7), 566–571. doi: 10.1111/j.0956-7976.2005.01576.x.
- Helgeson, V. S., & Shaver, K. G. (1990). Presumption of Innocence: Congruence Bias Induced and Overcome. *Journal of Applied Social Psychology*, *20*(4), 276–302. doi: 10.1111/j.1559-1816.1990.tb00412.x.
- Hendriks, M. C. P., & Vingerhoets, A. J. J. M. (2006). Social messages of crying faces: Their influence on anticipated person perception, emotions and behavioural responses. *Cognition and Emotion*, *20*, 878–886. doi: 10.1080/02699930500450218.
- Hinton, P. R. (1993). *The psychology of interpersonal perception*. London: Routledge.
- Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's Blindsight: How Cognitive Load Can Improve Judgments. *Psychological Science*, *24*, 869–879.
- Hogan, N. S., & DeSantis, L. (1992). Adolescent sibling bereavement: an ongoing attachment. *Qualitative Health Research*, *2*(2), 159–177.
- Hogarth, R. M. (2001). *Educating intuition*. Chicago, IL: University of Chicago Press.
- Hughes, S. M., Dispenza, F., & Gallup, G. G., Jr. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*(5), 295–304.
- Hui, C. H., & Yam, Y. M. (1987). Effects of language proficiency and physical attractiveness on person perception. *British Journal of Social Psychology*, *26*, 257–261.
- Huitt, W., Hummel, J., & Kaeck, D. (1999). *Internal and external*

- validity: General issues*. Retrieved from <http://www.edpsycinteractive.org/topics/intro/valdgn.html>.
- Hurd, K., & Noller, P. J. (1988). Decoding deception: A look at the process. *Journal of Nonverbal Behavior*, *12*(3), 217-233. doi: 10.1007/BF00987489.
- Hussain, S., Calvo, R. A., & Chen, F. (2013). Automatic Cognitive Load Detection from Face, Physiology, Task Performance and Fusion during Affective Interference. *Interacting with Computers*, first published online June 6, 2013. doi: 10.1093/iwci/iwt032.
- Inbau, F. E., Reid, J. E., & Buckley, J. P. (1986). *Criminal interrogation and confessions* (3rd ed.). Baltimore, MD: Williams & Wilkins.
- Incoherent. (2016). In Oxford English dictionary online (2nd ed.). Retrieved from <http://www.oup.com>.
- Isaac, S., & Michael, W. B. (1971). *Handbook in research and evaluation*. San Diego: EdITS.
- Jackson, H. J. (1932). *Selected Writings Vol. 2*. London: Hodder & Stoughton.
- Jacobs, L., Keown, C., Worthley, R., & Ghymn, K. (1991). Cross-cultural Colour Comparisons: Global Marketers Beware! *International Marketing Review*, *8*(3), 21–30.
- Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical Attractiveness Biases in Ratings of Employment Suitability: Tracking Down the “Beauty is Beastly” Effect. *The Journal of Social Psychology*, *150*(3), 301–318.
- Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: a novel class of erroneous deductions. *Cognition*, *71*, 191–229.
- Johnson-Laird, P. N. (2012). Mental models and consistency. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 225-243). New York, NY: Guilford Press.
- Jones, C. (2009, September 10). Block 14; Objective 3: Tests of association for nominal data. Retrieved from <http://www.angelo.edu/faculty/ljones/homepage/home.htm>.
- Jundi, S., Vrij, A., Mann, S., Hope, L., Hillman, J., Warmelink, L., & Gahr, E.

- (2013). Who should I look at? Eye contact during collective interviewing as a cue to deceit. *Psychology, Crime & Law*, *19*(8), 661–671. doi: 10.1080/1068316X.2013.793332.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.
- Kassin, S. M. (2012). Why confessions trump innocence. *American Psychologist*, *67*(6), 431–445.
- Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioural Confirmation in the Interrogation Room: On the Dangers of Presuming Guilt. *Law and Human Behavior*, *27*(2), 187–203.
- Kassin, S. M., & Fong, C. T. (1999). “I’m innocent!”: Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, *23*(5), 499–516.
- Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). I’d Know a False Confession if I Saw One: A Comparative Study of College Students and Police Investigators. *Law and Human Behavior*, *29*(2). doi: 10.1007/s10979-005-2416-9.
- Kastor, E. (1994, November 5). The worst fears, the worst reality: For parents, murder case strikes at the heart of darkness. *The Washington Post*, p. A15.
- Kitson, G. C., & Zyzanski, S. J. (1987). Grief in Widowhood and Divorce. *Psychiatric Clinics of North America*, *10*(3), 369–386.
- Klass, D. (2006). Death, grief, religion, and spirituality (Special issue), *Omega, Journal of Death and Dying*, *53*, 1–2.
- Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as communication construct. *Human Communication Research*, *1*, 15–29.
- Köhnken, G., Granhag, P. A., & Strömwall, L. A. (2004). *Statement Validity Analysis and the ‘detection of the truth’*. *The Detection of Deception in Forensic Contexts*. Cambridge: Cambridge University Press.
- Kraut, R. E. (1980). Humans as lie detectors: some second thoughts. *Journal of Communication*, *30*, 209–216.

- Kübler-Ross, E. (2005) *On Grief and Grieving: Finding the Meaning of Grief through the Five Stages of Loss*. London: Simon & Schuster Ltd.
- Kujipers, M. M. (2014). *Absorbing stories. The effects of textual devices on absorption and evaluative responses* (Unpublished doctoral dissertation). Utrecht University, Netherlands.
- Kujipers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *SSOL Scientific Study of Literature*, 4(1), 89–122. doi: 10.1075/ssol.4.1.05kui.
- Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology. University of Illinois Press*, 113(3), 387–404. doi: 10.2307/1423365.
- Labov, W. (1972). *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
- Lam, R. (2001). Stories and assessments of plausibility. In J. van Benthem (Ed.), *TARK 2001 Proceedings of the 8<sup>th</sup> conference on Theoretical aspects of rationality and knowledge* (pp. 39–49). San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.
- Landström, S., Ask, K., Sommar, C., & Willén, R. M. (2013). Children's testimony and the emotional victim effect. *Legal and Criminological Psychology*, 20(2), 365–383. doi: 10.1111/lcrp.12036.
- Landström, S., Granhag P. A., & Hartwig, M. (2007). Children's live and videotaped testimonies: How presentation mode affects observers' perception, assessment and memory. *Legal and Criminological Psychology*, 12, 333–347. doi: 10.1348/135532\_506X133607.
- Lane, D. (2013, December 1). *Type I and II errors (1 of 2)*. Retrieved from <http://davidmlane.com/hyperstat/A18652.html>.
- Lane, J. D., & Harris, P. L., (2014). Confronting, representing, and believing supernatural. *Perspectives on Psychological Science*, 9, 144–160. doi:

10.1177/1745691613518078.

- Lange, B. P., Zaretsky, E., Schwarz, S., & Euler, H. A. (2014). Words won't fail: Experimental evidence on the role of verbal proficiency in mate choice. *Journal of Language and Social Psychology, 33*(5), 482–499. doi: 10.1177/0261927x13515886.
- Lapata, M., McDonald, S., & Keller, F. (1999). Determinants of Adjective-Noun Plausibility. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 30-36). Bergen, Norway: Association for Computational Linguistics.
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *Cognitive Science, 9*(2), 75–82.
- Leal, S., & Vrij, A. (2012). The occurrence of eye blinks during a guilty knowledge test. *Psychology, Crime and Law, 16*(4), 349–357.
- Leal, S., Vrij, A., Mann, S., Fisher, R. P. (2010). Detecting true and false opinions: The Devil's Advocate approach as a lie detection aid. *Acta Psychologica, 134*(3), 323–329. doi: 10.1016/j.actpsy.2010.03.005.
- Lensing, V. (2001). Grief support: The role of funeral service. *Journal of Loss and Trauma, 6*(1), 45–63.
- Leventhal, G., & Krate, R. (1977). Physical Attractiveness and Severity of Sentencing. *Psychological Reports, 40*, 315–318.
- Levine, T. R., & Bond, C. F., Jr. (2014). Direct and indirect measures of lie detection tell the same story: A reply to ten Brinke, Stimson, and Carney (2014). *Psychological Science, 25*, 1960–1961. doi: 10.1177/0956797614536740.
- Levine, T. R., & McCornack, S. A. (1991). The dark side of trust: Conceptualizing and measuring types of communicative suspicion. *Communication Quarterly, 39*, 325–340.
- Levine, T. R., & McCornack, S. A. (2001). Behavioral adaption, confidence, and heuristic-based explanations of the probing effect. *Human Communication Research, 27*(4), 471–502.

- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect”. *Communication Monographs*, *66*(2), 125–144.
- Lindemann, E. (1944). Symptomatology and management of acute grief. *American Journal of Psychiatry*, *101*(2), 141–148.
- Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition*, *22*, 395–410.
- Longstaffe, K. A., Hood, B. M., & Gilchrist, I. D. (2014). The influence of cognitive load on spatial search performance. *Attention, Perception, & Psychophysics*, *76*, 49–63.
- Machin, L. (2001). *Exploring a framework for understanding the range of response to loss: A study of clients receiving bereavement counselling* (Unpublished doctoral dissertation). Keele University, United Kingdom.
- Maciejewski, P. K., Zhang, B., Block, S. D., & Prigerson, H. G. (2007). An Empirical Examination of the Stage Theory of Grief. *Journal of the American Medical Association*, *297*(7), 716–723. doi: 10.1001/jama.297.7.716.
- Mann, S., Vrij, A., & Bull, R. (2002). Suspects, Lies, and Videotape: An Analysis of Authentic High-Stake Liars. *Law and Human Behavior*, *26*(3), 365–376.
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers’ ability to detect suspects’ lies. *Journal of Applied Psychology*, *89*(1), 137–149. doi: 10.1037/0021-9010.89.1.137.
- Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic Versus Systematic Processing of information in Detecting Deception: Questioning the Truth Bias. *Psychological Reports*, *105*, 11–36.
- Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78.
- McAdams, D. P. (2006). *The redemptive self: Stories Americans live by*. New York, NY: Oxford University Press.
- McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. In M. L. McLaughlin (Ed.), *Communication yearbook 9* (pp. 377–389). Beverly Hills, CA: Sage.

- McGaugh, J. L. (1989). Involvement of hormonal and neuromodulatory systems in the regulation of memory storage. *Annual Review of Neuroscience, 12*, 2552–2587.
- McGuire, W. J. (1968). Personality and susceptibility to social influence. In E. F. Borgatta & W. W Lambert (Eds.), *Handbook of personality theory and research* (pp. 1130–1187). Chicago, IL: Rand McNally.
- Meijer, E. H., Bente, G., Ben-Shakhar, G., & Schumacher, A. (2013). Detecting concealed information from groups using a dynamic questioning approach: Simultaneous skin conductance measurement and immediate feedback. *Frontiers in Psychology, 4*, 1–6.
- Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29*(1), 176–186.
- Meissner, C. A., & Kassin, S. M. (2002). “He’s guilty!”: investigator bias in judgements of truth and deception. *Law and Human Behavior, 26*(5), 469–480.
- Mesquita, B., & Walker, R. (2003). Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour Research and Therapy, 41*(7), 777–793.
- Metric and non-metric scaling. (n.d). Retrieved from [http://www.simfit.org.uk/metric\\_and\\_non-metric\\_scaling.pdf](http://www.simfit.org.uk/metric_and_non-metric_scaling.pdf).
- Miles, J. N. V., Shevlin, M. E., & McGhee, P. (1999). Examining gender differences in reliability of the EPQ: A bootstrapping approach. *British Journal of Psychology, 90*(1), 145–154.
- Miller, G. R., & Stiff, J. B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*(2), 81–97.
- Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? *Personality & Social Psychology Bulletin, 10*, 243–252. doi: 10.1177/0146167284102010.
- Moshman, D. (2000). Diversity in reasoning and rationality: Metacognitive

- and developmental considerations. *Behavioral & Brain Sciences*, 23, 689–690.
- Moyle Wright, P., & Hogan, N. S. (2008). Grief Theories and Models: Applications to Hospice Nursing Practice. *Journal of Hospice and Palliative Nursing*, 10(6), 350–356.
- Mueller, C. B., & Laird, C. (2009). *Evidence* (4th ed.). New York, NY: Aspen Publishers.
- Mumby, P. J. (2002). Statistical power of non-parametric tests: a quick guide for designing sampling strategies, *Marine Pollution Bulletin*, 44(1), 85–87.
- Murray, D. J., Rowan, A. J., Smith, K. H. (1988). The effect of articulatory suppression on short-term recognition. *Canadian Journal of Psychology*, 4(42), 424–436. doi: 10.1037/h0084204.
- Myers, D. G. (2002). *Intuition. Its powers and perils*. New Haven, CT: Yale University Press.
- Nelson, T. O. (2002). *Metacognition: Core readings*. Boston: Allyn and Bacon.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Personality & Social Psychology Bulletin*, 29(5), 665–675. doi: 10.1177/0146167203251529.
- Ng, M., & Youngs, D. (2016). Veracity Assessment: aspects of the Account, the Source and the Judge that influence judgements of plausibility. *Crime Psychology Review*, 1(1), 135–154. doi: 10.1080/23744006.2016.1177946.
- Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220.
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, 109(3), 504–511.
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3(5), 400–424. doi: 10.1111/j.1745-6924.2008.00088.x.
- Norton, E. C., Bieler, G. S., & Ennett, S. T. (1996). Analysis of prevention program effectiveness with clustered data using generalized estimating equations. *Journal of Consulting and Clinical Psychology*, 64(5), 919–926.

- O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioral & Brain Sciences*, *22*(1), 127–196.
- Okada, T., & Simon, H. A. (1997). Collaborative Discovery in a Scientific Domain. *Cognitive Science*, *21*(2), 109–146.  
doi:10.1207/s15516709cog2102\_1.
- Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 211–238). New York: Guilford Press.
- Onyebadi, U., & Park, J. (2012). 'I'm Sister Maria. Please help me': A lexical study of 4-1-9 international advance fee fraud email communications. *The International Communication Gazette*, *74*(2), 181–199.
- Operario, D., & Fiske, S. T. (1999). Social Cognition Permeates Social Psychology: Motivated Mental Processes Guide the Study of Human Social Behavior. *Asian Journal of Social Psychology*, *2*(1), 63–78.
- Organised. (2017). In *OxfordDictionaries.com*. Retrieved from <https://en.oxforddictionaries.com/definition/organised>.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. doi: 10.1037/h0043424.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*, 988–1010.
- O'Sullivan, M. (2008). Home runs and humbugs: Comment on Bond and DePaulo (2008). *Psychological Bulletin*, *134*, 493–497.
- O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behaviour*, *12*(3), 203–216.
- O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*, *33*, 542–543.
- Otten, L. J., Alain, A., & Picton, T. W. (2000). Effects of visual attentional load on auditory processing. *Neuroreport*, *11*(4), 875–880. doi: 10.1097/00001756-200003200-00043.

- Palmer, C. L., & Peterson, R. D. (2012). Beauty and the Pollster: The Impact of Halo Effects on Perceptions of Political Knowledge and Sophistication. Midwest Political Science Association. Retrieved from [http://wayback.archive.org/web/20140309023825/http://cas.illinoisstate.edu/clpalme/research/documents/Beauty\\_and\\_the\\_Pollster\\_revision.pdf](http://wayback.archive.org/web/20140309023825/http://cas.illinoisstate.edu/clpalme/research/documents/Beauty_and_the_Pollster_revision.pdf).
- Park, C. L. (2013). The Meaning Making Model: A framework for understanding meaning, spirituality, and stress-related growth in health psychology. *The European Health Psychologist, 15*(2), 40–47.
- Park, J., & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology, 78*(6), 1005–1023.
- Park, H. S., & Levine, T. R. (2001). A probability model of accuracy in deception detection experiments. *Communication Monographs, 68*, 201–210.
- Parker, R. L., & Szymanski, E. M. (Eds.). (1992). *Rehabilitation counselling: Basics and beyond* (2nd ed.). Austin, TX: Pro-Ed.
- Parkes, C. (1972). *Bereavement: Studies of grief in adult life*. London, UK: Tavistock.
- Patry, M. W. (2008). Attractive but Guilty: Deliberation and The Physical Attractiveness Bias. *Psychological Reports, 102*(3), 727–733.
- Paunonen, S. (2006). You are honest, therefore I like you and find you attractive. *Journal of Research in Personality, 40*(3), 237–249.
- Peace, K. A., & Sinclair, S. M. (2012). Cold-blooded lie catchers? An investigation of psychopathy, emotional processing, and deception detection. *Legal and Criminological Psychology, 17*(1), 177–191. doi: 10.1348/135532510x524789.
- Penman, E. L., Breen, L. J., Hewitt, L. Y., & Prigerson, H. G. (2014). Public Attitudes About Normal and Pathological Grief. *Death Studies, 38*(8), 510–516.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC): LIWC 2001*. Mahwah, NJ: Erlbaum.

- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. (2003). Psychological aspects of natural Language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.
- Pennington, N., & Hastie, R. (1981). Juror decision making models: The generalization gap. *Psychological Bulletin, 89*, 246–287.
- Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 521–525.
- Pennington, N., & Hastie, R. (1990). Practical implications of psychological research on juror and jury decision making. *Personality and Social Psychology Bulletin, 16*, 90–100.
- Petty, R. E., & Cacioppo, J. T. (1981). Attitudes and persuasion: Classic and contemporary approaches. Dubuque, IA: William C. Brown.
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology, 19*, 123–181. doi:10.1016/s0065-2601(08)60214-2.
- Porter, S., & ten Brinke, L. (2010). The Truth About Lies: What Works in Detecting High-Stakes Deception? *Legal and Criminological Psychology, 15*, 57–75.
- Practice. (2017). In *OxfordDictionaries.com*. Retrieved from <https://en.oxforddictionaries.com/definition/practice>.
- Price, I. (2000). Chapter 2: Research Design. *Threats to Internal Validity*. Retrieved from [http://webstat.une.edu.au/unit\\_materials/c2\\_research\\_design/validity\\_threat\\_internal.htm](http://webstat.une.edu.au/unit_materials/c2_research_design/validity_threat_internal.htm).
- Raveh, A., & Landau, S. (1993). Partial Order Scalogram Analysis with Base Coordinates (POSAC): Its Application to Crime Patterns in All the States in the United States. *Journal of Quantitative Criminology, 9*(1), 83–99.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York, NY: Oxford University Press.
- Rebhun, U. (2012). Facet approach in the sociology of religion: Theories, techniques and analysis of religious identification in the United States.

- Annual Review of Sociology of Religion*, 3, 249–281.
- Reinhard, M-A., & Schwarz, M. N. (2012). The Influence of Affective States on the Process of Lie Detection. *Journal of Experimental Psychology: Applied*, 18(4), 377–389.
- Reinhard, M-A., & Sporer, S. L. (2008). Verbal and nonverbal behaviour as a basis for credibility attribution: The impact of task involvement and cognitive capacity. *Journal of Experimental Social Psychology*, 44, 477–488.
- Reinhard, M-A., & Sporer, S. L. (2010). Content versus Source Cue Information as a Basis for Credibility Judgments: The Impact of Task Involvement. *Social Psychology*, 41, 93–104. doi: 10.1027/1864-9335/a000014.
- Reynolds, E., & Rendle-Short, J. (2010). Cues to deception in context: Response latency/gaps in denials and blame shifting. *British Journal of Social Psychology*, 50(3), 431–449. doi: 10.1348/014466610X520104.
- Richter, T., Schroeder, S., & Wohrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, 96(3), 538–558.
- Risinger, D. M., Saks, M. J., Thompson, W. C., & Rosenthal, R. (2002). The Daubert/Kumho implications of observer effects in forensic science: Hidden problems of expectation and suggestion. *California Law Review*, 90(1), 1–56.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2006). MIRRORS IN THE MIND. *Scientific American*, 295(5), 54–61.
- Rosenblatt, P. C. (1988). Grief: The Social Context of Private Feelings. *Journal of Social Issues*, 44, 67–78. doi: 10.1111/j.1540-4560.1988.tb02077.x.
- Rosenblatt, P. C., & Elde, C. (1990). Shared reminiscence about a deceased parent: Implications for grief education and grief counseling. *Family Relations*, 39(2), 206–210.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1), 16–20. doi: 10.1007/BF02322211.

- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320.
- Rubin, M., & Badaea, C. (2010). The central tendency of a social group can affect ratings of its intragroup variability in the absence of social identity concerns. *Journal of Experimental Social Psychology*, 46, 410–415. doi: 10.1016/j.jesp.2010.01.001.
- Salerno, J. M., & Bottoms, B. L. (2009). Emotional evidence and jurors' judgments: The promise of neuroscience for informing psychology and law. *Behavioural Sciences and the Law: Special Issue: The Neuroscience of Decision Making and Law*, 27, 273–296.
- Sapir, A. (2005). *The LSI course on scientific content analysis (SCAN)*. Phoenix, AZ: Laboratory for Scientific Interrogation.
- Sarter, H. (2012). Language and Emotion. *Applied Psycholinguistics. Positive Effects and Ethical Perspectives*, 1, 61–73.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22(1), 36–71.
- Schwarz, N. (2002). Situated cognition and the wisdom of feelings: Cognitive tuning. In L. Feldman Barrett & P. Salovey (Eds.), *The wisdom in feelings* (pp. 144–166). New York, NY: Guilford.
- Schwarz, N. (2012). Feelings-as-information theory. In P. A. M. Van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 289–308). Thousand Oaks, CA: Sage Publishing.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information. *Advances in Experimental Social Psychology*, 39, 127–161.
- Semmler, C., & Brewer, N. (2002). Effects of mood and emotion on juror processing and judgments. *Behavioral Sciences & the Law*, 20(4), 423–436.
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49, 11–36.
- Shear, M. K., & Mulhare, E. (2008). Complicated grief. *Psychiatric Annals*, 39,

- 662–670.
- Shoemaker, D. J., South, D. R., & Lowe, J. (1973). Facial stereotypes of deviants and judgments of guilt or innocence. *Social Forces*, *51*(4), 427–433. doi: 10.2307/2576687.
- Shuchter, S. R., & Zisook, S. (1988). Widowhood. The continuing relationship With the dead spouse. *Bulletin of Menninger Clinic*, *52*(3), 269–279.
- Shuy, R. W. (1998). *The Language of Confession, Interrogation and Deception*. Thousand Oaks, CA: Sage Publications Inc.
- Shye, S. (1978). Partial order scalogram analysis. In S. Shye (Ed.), *Theory construction and data analysis in the behavioural sciences*. San Francisco: Jossey-Bass.
- Shye, S. (1985). *Multiple Scaling: The Theory and Application of Partial Order Scalogram Analysis*. New York, NY: Elsevier Inc.
- Sigall, H., & Ostrove, N. (1975). Beautiful but Dangerous: Effects of Offender Attractiveness and Nature of the Crime on Juridic Judgment. *Journal of Personality and Social Psychology*, *31*(3), 410–414.
- Simon, H. A. (1990). Invariants of human behaviour. *Annual Review of Psychology*, *41*, 1–20.
- Slepian, M. L., Bogart, K. R., & Ambady, N. (2014). Thin-slice judgments in the clinical context. *Annual Review of Clinical Psychology*, *10*(1), 131–153.
- Sloman, S. A. (1996). "The empirical case for two systems of reasoning". *Psychological Bulletin*, *119*, 3–22. doi: 10.1037/0033-2909.119.1.3.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*, 1333–1352.
- Slutz, S., & Hess, K. L. (2017). Increasing the Ability of an Experiment to Measure an Effect. *Science Buddies*. Retrieved from [http://www.sciencebuddies.org/science-fair-projects/top\\_research-project\\_signal-to-noise-ratio.shtml](http://www.sciencebuddies.org/science-fair-projects/top_research-project_signal-to-noise-ratio.shtml).
- Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*, 279–281. doi:10.1214/aoms/1177730256.

- Soufan, A. H. (2011). *The Black Banners: The Inside Story of 9/11 and the War Against Al-Qaeda*. London: Penguin Books.
- Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P. A. Granhag & L. Strömwall (Eds.), *Deception detection in forensic contexts* (pp. 64–102). Cambridge, UK: Cambridge University Press.
- Sporer, S. L. (1997). The Less Travelled Road to Truth: Verbal Cues in Deception Detection in Accounts of Fabricated and Self-Experienced Events. *Applied Cognitive Psychology, 11*, 373–397.
- Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic Synthesis. *Applied Cognitive Psychology, 20*, 421–446.
- Sprang, V. M., McNeil, J. S., & Wright, R. (1989). Psychological changes after the murder of a significant other. *The Journal of Contemporary Social Work, 70*, 159–164.
- Srull, T. K. (1983). Affect and memory: The impact of affective reactions in advertising on the representation of product information in memory. In R. P. Bagozzi & A. M. Tybout (Eds.), *Advances in consumer research* (Vol. 10, pp. 520–525). Ann Arbor, MI: Association for Consumer Research.
- Stage, F. K. Carter, H. C., & Nora, A. (2004). Path analysis: An introduction and analysis of a decade of research. *Journal of Educational Research, 98*(1), 5–12.
- Stanovich, K. E., & West, R. F. (2000). Individual difference in reasoning: implications for the rationality debate? *Behavioural and Brain Sciences, 23*, 645–726. doi:10.1017/s0140525x00003435.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives*. Norwood, NJ: Ablex.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217–245). New York, NY: Springer Verlag.
- Stenning, K., & van Lambalgen, M. (1999). Is psychology hard or impossible? Reflections on the conditional. In M. M. J. Gerbrandy, M. Marx, M. de Rijke, & Y. Venema (Eds.), *Liber amicorum for*

- Johan van Benthem's 50th birthday* (pp. 1–29). Amsterdam: Amsterdam University Press.
- Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research, 19*, 326–345.
- Strömwall, L. A., Granhag, P. A., & Jonsson, A-C. (2003). Deception among pairs: “Let’s say we had lunch and hope they will swallow it!”. *Psychology, Crime & Law, 9*(2), 109–124. doi: 10.1080/1068316031000116238.
- ten Brinke, L., Stimson, D., & Carney, D. (2014). Some evidence for unconscious lie detection. *Psychological Science, 25*(5), 1098–1105. doi: 10.1177/0956797614524421.
- Street, C. N. H. (2013). *Lie detection: Cognitive processes* (Doctoral dissertation, University College London, London, UK). Retrieved from <http://discovery.ucl.ac.uk/1414942/>.
- Street, C. N. H., & Kingstone, A. (2015). Aligning Spinoza with Descartes: An informed Cartesian account of the truth bias. *British Journal of Psychology*. doi: 10.1111/bjop.12210.
- Street, C. N. H., & Masip, J. (2015). The source of the truth bias: Heuristic processing? *Scandinavian Journal of Psychology, 56*(3), 254–263. doi: 10.1111/sjop.12204.
- Street, C. N. H., & Richardson, D. C. (2014). Lies, damn lies, and expectations: How beliefs about the base rates inform lie-truth judgments. *Applied Cognitive Psychology, 29*(1), 149–155.
- Street, C. N. H., & Richardson, D. C. (2015). The focal account: Indirect lie detection need not access unconscious, implicit knowledge. *Journal of Experimental Psychology: Applied, 21*(4), 342–355. doi: 10.1037/xap0000058.
- Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners’ beliefs about deception. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229–250). Cambridge: Cambridge University Press.
- Suhr, D. (2008). Step your way through Path Analysis. In *Western users of SAS Software Conference Proceedings*, W. Universal City, California (pp. 1–

10).

- Sukumar, D., & Wade, K. A. (2016). Strategic Disclosure of Evidence: Perspectives from Psychology and Law. *Psychology, Public Policy and Law*, 22(3), 306–313.
- Surawski, M. K., & Ossoff, E. P. (2006). The effects of physical and vocal attractiveness on impression formation of politicians. *Current Psychology*, 25(1), 15–27.
- Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, 119(534), 47–60.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138.
- Sweeney, C. D., & Ceci, S. J. (2014). Deception detection, transmission, and modality in age and sex. *Frontiers in Psychology*, 13(5), 590. doi: 10.3389/fpsyg.2014.00590.
- Tamborini, R., Huang, R-H., Mastro, D., & Nabashi-Nakahara, R. (2007). The Influence of Race, Heuristics, and Information Load on Judgments of Guilt and Innocence. *Communication Studies*, 58(4), 341–358.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Taylor, J. R. (1999). *An introduction to error analysis: The Study of Uncertainties in Physical Measurements*. Herndon, VA: University Science Books.
- Taylor, P. J. (2002). A partial order scalogram analysis of communication behaviour in crisis negotiation with the prediction of outcome. *The International Journal of Conflict Management*, 13(1), 4–37.
- ten Brinke, L. (2015, August 26). Personal communication.
- ten Brinke, L., & Porter, S. (2011). Cry Me a River: Identifying the Behavioral Consequences of Extremely High-Stakes Interpersonal Deception. *Law and Human Behavior*. Advance online publication. doi: 10.1037/h009392.
- Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

- The Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74.
- The SPSS Logistic Regression Output. (2011, 22 July). Retrieved from <http://www.restore.ac.uk/srme/www/fac/soc/wie/research-news/srme/modules/mod4/12/>.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29. doi:10.1037/h0071663.
- Toma, C., & Hancock, J. T. (2010). Looks and Lies: The Role of Physical Attractiveness in Online Dating Self-Presentation and Deception. *Communication Research*, *37*(3), 335–351.
- Toris, C., & DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology*, *47*, 1063–1073.
- Trabasso, T, Secco, T., & van den Broek., P. W. (1984). *Causal cohesion and story coherence*. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. *Science: New Series*, *185*(4157), 1124–1131.
- Validity in scientific investigation. (2016). Retrieved from <http://www.statsdirect.com/help/default.htm#basics/validity.htm>.
- Van Der Zee, S., Poppe, R., Taylor, P. J., & Anderson, R. (2015). To freeze or not to freeze: A motion-capture approach to detecting deception. *Conference Proceedings of the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, 48th HICSS (5-8 January, 2015).
- Vance, J. C., Najman, J. M., Thearle, M. J., Embelton, G., Foster, W. J., & Boyle, F. M. (1995). Psychological Changes in Parents Eight Months After the Loss of an Infant from Stillbirth, Neonatal Death, or Sudden Infant Death Syndrome—A Longitudinal Study. *Pediatrics*, *96*(5), 933–938.
- Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and its Detection Effects of Monetary Incentives and Personal Relationship History. *Communication Research*, *39*(2), 217–238.
- Vernham, Z., & Vrij, A. (2015). A review of the collective interviewing approach

- to detecting deception in pairs. *Crime Psychology Review*, 1(1), 43–58.
- von Hippel, W., Baker, E., Wilson, R., Brin, L., & Page, L. (2015). Detecting deceptive behaviour after the fact. *British Journal of Social Psychology*, 55, 195–205. doi: 10.1111/bjso.12129.
- Vredeveldt, A., Hitch, G. J., & Baddeley, A. D. (2011). Eye-closure helps memory by reducing cognitive load and enhancing visualisation. *Memory & Cognition*, 39, 1253–1263. doi: 10.3758/s13421-011-0098-8.
- Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.
- Vrij, A. (2005). Cooperation of liars and truth tellers. *Applied Cognitive Psychology*, 19(1), 38–50. doi: 10.1002/acp.1050.
- Vrij, A. (2005a). Criteria-based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology, Public Policy and Law*, 11(1), 3–41. doi: 10.1037/1076-8971.11.1.3.
- Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection: a plea to change police practice. *Criminal Justice and Behaviour*, 35(10), 1323–1336. doi: 10.1177/0093854808321530.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8–41.
- Vrij, A., Edward, K., & Bull, R. (2001). Police officers' ability to detect deceit: The benefit of indirect deception detection measures. *Legal and Criminological Psychology*, 6(2), 185–196.
- Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24(4), 239–263. doi: 10.1023/A:1006610329284.
- Vrij, A., & Fischer, A. (1995). The expression of emotions in simulated rape interviews. *Journal of Police and Criminal Psychology*, 10(3), 64–67.
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Science*, 10(4), 141–142.
- Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What

- matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, 1(2), 110–117. doi: 10.1016/j.jarmac.2012.02.004.
- Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, 11(3), 89–121.
- Vrij, A., Jundi, S., Hope, L., Hillman, J., Gahr, E., Leal, S., Warmelink, L., Mann, S., Vernham, Z., & Granhag, P. A. (2012). Collective Interviewing of Suspects. *Journal of Applied Research in Memory and Cognition*, 1, 41–44. doi: 10.1016/j.jarmac.2011.12.002.
- Vrij, A., Kneller, W., & Mann, S. (2000). Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, 5(1), 57–70.
- Vrij, A., Leal, S., Granhag, P.A., Mann, S., Fisher, R., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, 33(2), 159–166.
- Vrij, A., & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high stake lies. *Journal of Psychology, Crime and Law*, 7, 119–132.
- Vrij, A., & Mann, S. (2001a). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology*, 15(2), 187–203.
- Vrij, A., & Mann, S. (2004). Detecting deception: The Benefit of Looking at a Combination of Behavioral, Auditory and Speech Content Related Cues in a Systematic Manner. *Group Decision and Negotiation*, 13, 61–79.
- Vrij, A., Mann, S., & Leal, S. (2013). Deception traits in psychological interviewing. *Journal of Police and Criminal Psychology*, 28(2), 115–126. doi: 10.1007/s11896-013-9125-y.
- Vrij, A., Mann, S., Leal, S., & Granhag, P. (2010). Getting into the minds of pairs of liars and truth tellers: An examination of their strategies. *Open Criminology Journal*, 3(1), 17–22.

- Vrij, A., Mann, S., Robbins, E., & Robinson, M. (2006). Police officers' ability to detect deception in high stakes situations and in repeated lie detection tests. *Applied Cognitive Psychology, 20*, 741–755. doi: 10.1002/acp.1200.
- Vrij, A., Semin, G. R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research, 22*, 544–562.
- Wade, T. J., & DiMaria, C. (2003). Weight Halo Effects: Individual Differences in Perceived Life Success as a Function of Women's Race and Weight. *Sex Roles, 48*(9), 461–465. doi: 10.1023/A:1023582629538.
- Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology, 92*, 821–833.
- Weaver, E. A., & Stewart, T. R. (2012). Dimensions of Judgment: Factor Analysis of Individual Differences. *Journal of Behavioural Decision Making, 25*, 402–413. doi: 10.1002/bdm.748.
- Weerman, F. M. (2001). *Samenplegen, over criminele samenwerking en groepsvorming*. Nijmegen: Ars Aequi Libri.
- Well thought out. (2017). In *OxfordDictionaries.com*. Retrieved from [https://en.oxforddictionaries.com/definition/well\\_thought\\_out](https://en.oxforddictionaries.com/definition/well_thought_out).
- Wells, G. L., Steblay, N. K., & Dysart, J. E. (2011). A Test of the Simultaneous vs. Sequential Lineup Methods: An Initial Report of the AJS National Eyewitness Identification Field Studies. Des Moines, Iowa: American Judicature Society. Retrieved from: <http://www.popcenter.org/library/reading/PDFs/lineupmethods.pdf>.
- Wessel, E. M., Bollingmo, G. C., Sønsteby, C., Nielsen, L. M., Eliersten, D. E., & Magnussen, S. (2012). The emotional witness effect: story content, emotional valence and credibility of a male suspect. *Psychology, Crime and Law, 18*(5), 417–430.
- Wessel, E., Magnussen, S. and Melinder, A. M. D. (2013). Expressed Emotions and Perceived Credibility of Child Mock Victims Disclosing Physical Abuse. *Applied Cognitive Psychology, 27*, 611–616. doi: 10.1002/acp.2935.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80–83.

- Wilder, D. A., Simon, A. F., & Faith, M. (1996). Enhancing the Impact of Counterstereotypic Information: Dispositional Attributions for Deviance. *Journal of Personality and Social Psychology*, *71*(2), 276–287.
- Winokur, M. (2005, August 30). *Errors and Uncertainties*. Retrieved from <http://badger.physics.wisc.edu/lab/manual/node4.html>.
- Woolley, J. D., & Ghossainy, M. E. (2013). Revisiting the fantasy-reality distinction: Children as naïve skeptics. *Child Development*, *84*(5), 1496–1510.
- Worden, J. W. (1983). *Grief counselling and grief therapy* (1st ed.). London, UK: Tavistock.
- Wright, S. (1921). Correlated and causation. *Journal of Agricultural Research*, *20*, 557–585.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*(3), 161–215. doi: 10.1214/aoms/1177732676.
- Whelan, C., Wagstaff, G. F., & Wheatcroft, J. M. (2014). High-Stakes Lies: Verbal and Nonverbal Cues to Deception in Public Appeals for Help with Missing or Murdered Relatives. *Psychiatry, Psychology and Law*, *21*(4), 523–537. doi: 10.1080/13218719.2013.839931.
- Yogev-Seligmann, G., Rotem-Galili, Y., Dickstein, R., Giladi, N., & Hausdorff, J. M. (2012). Effects of explicit prioritization on dual task walking in patients with Parkinson's disease. *Gait Posture*, *35*(4), 641–646. doi: 10.1016/j.gaitpost.2011.12.016.
- Yuille, J. C. (1989). *Credibility Assessments*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Zajonc, R. B. (1965). Social Facilitation. *Science*, *149*(3681), 269–274.
- Zajonc, R. B. (2001). Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228. doi: 10.1111/1467-8721.00154.
- Zaparniuk, J., Yuille, J. C. & Taylor, S. (1995). Assessing the Credibility of True and False Statements. *International Journal of Law and Psychiatry*, *18*(3), 343–352.
- Zhang, Z. (2014). Too much covariates in a multivariable model may cause the

- problem of overfitting. *Journal of Thoracic Disease*, 6(9), 196–197. doi: 10.3978/j.issn.2072-1439.2014.08.33.
- Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication. *Group Decision and Negotiation*, 13, 81–106.
- Zhou, L., & Zhang, D. (2006). A Comparison of Deception Behavior in Dyad and Triadic Group Decision Making in Synchronous Computer-Mediated Communication. *Small Group Research*, 37, 140–164. doi: 10.1177/1046496405285125.
- Zimmerman, L., Carlucci, M. E., & Compo, N. S. (2012). Lie Detection during High-Stakes Truths and Lies. *Legal and Criminological Psychology*, 2, 314–323.
- Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, 43(2), 347–357.
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication in deception. In L. Berkowitz, (Ed.), *Advances in experimental social psychology* (Vol. 14, pp.1–59). New York, NY: Academic Press.
- Zuckerman, M., Fischer, S.A., Osmun, R. W., Winkler, B. A., & Wolfson, L. R. (1987). Anchoring in Lie Detection Revisited. *Journal of Nonverbal Behavior*, 11(1), 4–12.

## APPENDIX A: Content dictionary of variables

Number on SSA figure	Key	Definition	Frequency for Genuine Appealers (%)	Frequency for False Appealers (%)
<i>Self-centred</i>				
1	Conditional	Appealer in reference to victim, mentions the conjunctions "if you cannot contact" and "or"	0	17.6
2	Ididntdo	Appealer mentions "I didn't do it", "I'm not a suspect", "I had nothing to do with it"	0	17.6
3	Idontknow	Mention of "I don't know" or "don't know" any time per narrative	4.5	17.6
4	Somebody	Appealer mentions "somebody/some people"	13.6	29.4
<i>Tense</i>				
5	Vicpresenttense	Appealer speaks about the victim in the present tense regardless of whether the victim has been found, is dead, or is still missing	72.7	41.2
<i>Victim-centred</i>				
6	Vicimagined	Appealer fantasises about what victim would be doing if they were with them, what they are going through, or wonder about their condition	9.1	0
7	keepcount	Appealer keeps count of time since victim has been missing or dead	27.3	5.9

***Social***

8	communality	Appealer cooperates with community, builds rapport and a sense of everyone helping together to find victim and/or killer; the community holding the appellant up in support to keep going	54.5	11.8
9	religion	Appellant mentions or refers to "God", "faith", "prayer"	22.7	0

***Emotional***

10	Hope	Appellant mentions "hope", "hopefully", "not give up"	22.7	0
11	Emotions	Mention of basic emotions such as 'sadness', cognitive state of appellant i.e. "peace", or the lack of emotions i.e. "I've got no feelings"	45.5	5.9
12	Miss	Appellant states that he/she misses the victim	18.2	0

***Grammar***

13	Wronggrammar	Grammatical errors, syntax structure errors in appeal	18.2	41.2
----	--------------	---	------	------

***Chronology***

14	Eventref	Appellant mentions or refers to the event i.e. abduction, killing, in terms of time i.e. morning of, day of the week, date	9.1	35.3
----	----------	--	-----	------

***Movement***

15	Activity	Defined as presence of movement, activity, directional, change, with appellant mentioning phrases such as "If anybody knows anything, let us/police know", "forward", "call for search", "moving", "looking"	45.5	11.8
----	----------	--	------	------

**APPENDIX B: Consent form**  
**TELEVISION APPEALS PROJECT**

As part of a study of people making appeals or interviews on television where their loved ones have disappeared or been killed, I will show you 4 short films taken from actual televised appeals. Some of them are guilty and some of them are not. I would then like you to answer some questions about each film. It should all take less than 30 minutes.

There is no right or wrong answer. It is your own personal opinion that is wanted.

We do not need to know your name, so the whole study is anonymous and your responses will be confidential to the research team.

Some people may find these short videos distressing. So if, at any time for whatever reason, you feel like you wish to stop taking part just let me know and we'll stop the experiment and you can leave without any consequences. You can withdraw from this experiment at any time but once the questionnaire is handed in it is subject to analysis and your name will be replaced by a code number so all anonymity is ensured. If you feel you have been affected by any of the issues surrounding this experiment you can approach the Samaritans for emotional support; alternatively you can seek counselling service offered by CareLine and Victim Support.

My name is Magdalene Ng. I am supervised by Professor David Canter and Dr Donna Youngs. If you would like any further information regarding the study please contact Dr. Youngs at [d.youngs@hud.ac.uk](mailto:d.youngs@hud.ac.uk).

If you do wish to continue with the experiment please sign this consent form below.

Thank you for your time and agreement to take part in this study.

..... Date: .....

### APPENDIX C: Debrief form

Thank you very much for taking the time to complete this experiment.

The aim of this experiment was to investigate the phenomena of lying and telling the truth behaviours specifically in the context of false appeals from the receiver's end. A false appeal is where a family member or partner of a person who has gone missing or has been injured or murdered publically appeals for information regarding the crime either in a press conference or on while being interviewed by a journalist. However this person is either involved in the disappearance or is complicit in it but pretends not to know anything about it and deceives both the investigating police officers and attempts to extend this to the general public.

In order to explore false appeals I found cases of false appeals and matched them with true appeals for participants to watch and/or listen to them to see whether they could decide whether the person speaking in each case was lying or telling the truth. I also matched these appeals according to other variables such as age and attractiveness, as well as the narrative contents and roles. The study aims to explore whether people typically perceive certain characteristics in true and false appeals and whether these impact on their judgment of veracity.

If you were not able to correctly decide whether they were lying or telling the truth you must not worry. Detecting lies in other people is a notoriously difficult task and that is why psychology and other disciplines for many decades now has been trying to explore ways to detect lies and identify the verbal and nonverbal behaviour that may signal deceit.

If you feel you would like more information on any of the issues this experiment addresses, or on the experiment in general you can contact my supervisor Dr Donna Youngs at [d.youngs@hud.ac.uk](mailto:d.youngs@hud.ac.uk). Or alternatively if you could contact me at [magdalene.sophie@yahoo.com](mailto:magdalene.sophie@yahoo.com).

If you feel you have been affected by any of the issues surrounding this experiment you can call the **Samaritans** for emotional support on 0845 909090. Alternatively you could call the counselling service **CareLine** on 020 85141177 or Victim Support 0845 30 30 900.

**APPENDIX D: Question sheet**

1. Do you think the appellant in the video is innocent or guilty?

Please rate the following statements on a scale of 1 to 5, with options ranging from 1 for Strongly Disagree, 2 for Disagree, 3 for Undecided, 4 for Agree, and 5 for Strongly Agree.

2. The appellant's story makes sense.
3. The appellant sounds like he or she has had practice in making the appeal
4. The appellant tells a story that seems plausible.
5. I felt absorbed into the story the appellant was telling.
6. The appellant's emotions do not seem to match his or her story.
7. The appellant has a pleasant voice.
8. The appellant sounds like they had thought out well their appeal.
9. The appellant words his or her sentences in a manner that is pleasant to hear.
10. The appellant has an attractive face.
11. When I was listening to the appeal I was focused on what happened during the appeal.
12. I was worried about what was going to happen to the appellant and the missing person.
13. I felt sympathy for the appellant.
14. The appellant displays less emotion than seems appropriate for their situation.
15. The appellant makes an appeal that affected me emotionally.
16. The appellant gives an organised speech.
17. I believe that the appellant sounds credible overall.
  
18. Do you know personally anyone who featured in the appeal you've just heard?
  
19. Do you know of the appeal you have just heard i.e. you've watched or heard it on TV?
  
20. If you ticked yes to knowing the appeal, what did you believe the outcome was? Was the appellant guilty or innocent?

If you have any other comments or notes for this particular appeal, please state them on the space given on the answer sheet.

## APPENDIX E: Question sheet for Audio-only condition in Study 2

1. Do you think the appellant in the video is innocent or guilty?

Please rate the following statements on a scale of 1 to 5, with options ranging from 1 for Strongly Disagree, 2 for Disagree, 3 for Undecided, 4 for Agree, and 5 for Strongly Agree. Please ignore questions labelled 'N/A' - these apply to a different version of the appeals.

2. The appellant's story makes sense.
3. The appellant sounds like he or she has had practice in making the appeal
4. The appellant tells a story that seems plausible.
5. I felt absorbed into the story the appellant was telling.
6. The appellant's emotions do not seem to match his or her story.
7. The appellant has a pleasant voice.
8. The appellant sounds like they had thought out well their appeal.
9. The appellant words his or her sentences in a manner that is pleasant to hear.
10. N/A
11. When I was listening to the appeal I was focused on what happened during the appeal.
12. I was worried about what was going to happen to the appellant and the missing person.
13. I felt sympathy for the appellant.
14. N/A
15. The appellant makes an appeal that affected me emotionally.
16. The appellant gives an organised speech.
17. I believe that the appellant sounds credible overall.
  
18. Do you know personally anyone who featured in the appeal you've just heard?
  
19. Do you know of the appeal you have just heard i.e. you've watched or heard it on TV?
  
20. If you ticked yes to knowing the appeal, what did you believe the outcome was? Was the appellant guilty or innocent?

If you have any other comments or notes for this particular appeal, please state them on the space given on the answer sheet.

**APPENDIX F: Demographic sheet**

**Finally, please answer the following demographic questions:**

- a) What is your gender? Please circle: **Male / Female**
- b) What is your age?
- c) How would you describe your ethnic origin? Please circle below:  
**From the Asian subcontinent / African / White European / Chinese / Other**  
**(Please specify:  ) / Mixed (Please specify:  )**
- d) Please state your nationality:
- e) What is your occupation? E.g. academic / police officer:
- f) How long have you been in your current occupation?

**This is the end of the questionnaire. Thank you for your participation.**

APPENDIX G: Sample answer sheet

Video 1		Video 2		Video 3		Video 4	
Question	Answers	Question	Answers	Question	Answers	Question	Answers
1	Guilty	1	Guilty	1	Innocent 	1	Guilty
	Strongly Disagree Disagree Undecided Agree Strongly Agree						
2	<input checked="" type="checkbox"/>						
3	<input checked="" type="checkbox"/>						
4	<input checked="" type="checkbox"/>						
5	<input checked="" type="checkbox"/>						
6	<input checked="" type="checkbox"/>						
7	<input checked="" type="checkbox"/>						
8	<input checked="" type="checkbox"/>						
9	<input checked="" type="checkbox"/>						
10	<input checked="" type="checkbox"/>						
11	<input checked="" type="checkbox"/>						
12	<input checked="" type="checkbox"/>						
13	<input checked="" type="checkbox"/>						
14	<input checked="" type="checkbox"/>						
15	<input checked="" type="checkbox"/>						
16	<input checked="" type="checkbox"/>						
17	<input checked="" type="checkbox"/>						
18	Yes 	29	Yes 	29	Yes 	29	Yes 
19	Yes 	30	Yes 	30	No	30	Yes 
20	Innocent Guilty	31	Innocent Guilty	31	Innocent 	31	Innocent Guilty

Comments:

Video 1	
Video 2	
Video 3	
Video 4	

**APPENDIX H: Frequency of evidence types used to establish ground truth**

<b>Evidence type</b>	<b>Genuine (<i>N</i> =23)</b>	<b>Deceptive (<i>N</i> =16)</b>
Forensic evidence (e.g. soil, pollen, fibre, blood spatter patterns, arson tracks)	1	2
DNA evidence (e.g. body fluids, hair, skin)	9	7
CCTV/video/speed/security camera footage	1	2
Presence of victim's blood (e.g. in car; home of the accused)	4	2
Knowledge of location of victim's body or body parts	13	12
Confession un-recanted	7	12
Computer search history of location of crime/body	1	0
Post mortem evidence	2	0
Medical evidence (e.g. expert testimony)	2	0
Eyewitness testimony of appellant's whereabouts	2	2
Possession of murder weapon	1	1
Possession of items from crime scene	0	1
Phone records/phone range/phone tap	1	2
Incriminating monetary transactions	2	2
Inadequate account/alibi inconsistent with evidence	1	3
Public tip (e.g. finding burnt victim in car)	4	3

**NOTE: Totals exceed sample size because majority of cases were classified based on several pieces of evidence**

**APPENDIX I: Not-sure percentages reported by judges for the High and  
Low Emotionality conditions**

<b>Variables</b>	<b>High Emotionality (%)</b>	<b>Low Emotionality (%)</b>
Emotions match	53.3	57.1
Voice	35.6	31.7
Face	36.4	43.5
Plausible	11.1	31.7
Practice	31.1	20.6
Credible	24.4	32.3
Wording	20.0	28.6
Well-thought-out	17.8	14.3
Organised	11.1	12.7
Sense	11.1	19.0
Meta-emotion	2.2	1.6

**APPENDIX J: Not-sure percentages reported by judges for the Audio-only and Audio-visual conditions**

<b>Variables</b>	<b>Audio-only (%)</b>	<b>Audio-visual (%)</b>
Sense	24.5	19.2
Voice	20.3	24.0
Plausible	21.8	20.1
Credible	18.7	28.8
Practice	16.4	27.5
Organised	12.6	9.6
Wording	10.5	5.7
Well-thought-out	5.6	6.1
Meta-emotion	1.0	0.9

**APPENDIX K: Not-sure percentages reported by judges for the High and  
Low Sense conditions**

<b>Variables</b>	<b>Low Sense (%)</b>	<b>High Sense (%)</b>
Voice	30.0	23.7
Plausible	28.0	11.4
Practice	24.4	13.8
Wording	20.1	20.5
Organised	12.5	12.4
Face	22.0	33.8
Credible	22.0	14.6
Emotionsmatch	13.2	14.2
Well-thought-out	8.9	10.0
Less emotions	8.9	8.3
Meta-emotion	1.2	0.5

**APPENDIX L: Not-sure percentages reported by judges for the Paired  
and Solo conditions**

<b>Variables</b>	<b>Paired (%)</b>	<b>Solo (%)</b>
Face	36.8	22.4
Voice	27.6	20.8
Sense	25.8	31.6
Practice	22.2	12.0
Wording	22.7	22.3
Credible	24.9	21.2
Plausible	21.8	30.7
Emotions match	20.9	21.1
Less emotions	20.4	19.9
Well-thought-out	18.7	15.4
Organised	16.0	18.1
Meta-emotion	4.0	0.6

**APPENDIX M: Hypothesised cues for genuine and deceptive appeals by aspects**

<i>Hypothesised genuine cues</i>		<i>Brief definition</i>
<b>Tense</b>	Presenttense	Speaking about the victim in the present tense
<b>Victim-centred</b>	Vicimagined	Wondering or fantasizing about the victim's condition
	Keepcount	Keeping count of time since victim has been missing/dead
<b>Social</b>	Religion	Mentioning religion
	Communality	Mentioning and building rapport with community
<b>Emotional</b>	Emotions	Mentioning emotions or the lack of emotions
	Hope	Mentioning hope
	Miss	Mentioning missing the victim
<b>Movement</b>	Activity	Presence of movement words and activity in appeal
<i>Hypothesised deceptive cues</i>		<i>Brief definition</i>
<b>Self-centred</b>	Conditional	Mentioning conjunctions regarding victim coming home
	Ididntdoit	Mentioning "I didn't do it"
	Somebody	Mentioning "somebody/some people"

	Idontknow	Mentioning "I don't know"
<b>Expletives</b>	Swear	Cursing
<b>Grammar</b>	Wronggrammar	Grammatical/syntax errors that deviates from the official English language
<b>Chronology</b>	Eventref	Referencing event i.e. abduction/killing in terms of time
	Repetition	Mentioning words/phrases thrice or more in succession in the same sentence, excluding 'uhms' and 'ahs'

---

**APPENDIX N: Empirical support for Eight Language Areas**

<b>Areas</b>	<b>Variables</b>	<b>Empirical Support</b>
Tense	Presenttense	Dulaney, 1982; Ebesu & Miller, 1994; Harpster et al., 2009
Victim-centred	Vicimagined Keepcount	Klass, 2006; Maciejewski et al., 2007; Field & Filanosky, 2010; Shuchter & Zisook, 1988 Engel, 1975
Social	Religion	Bjork & Thurman, 2007
Emotional	Communality Emotions Hope Miss	Stone & Pennebaker, 2002; Cohn et al., 2001; Vrij & Mann, 2004; Bond et al., 2014; DePaulo et al., 2003; Anderson et al., 1999 Granhag et al., 2015; Sporer, 1997; Sapir, 2005; ten Brinke & Porter, 2012 Whelan et al., 2014 Granhag et al., 2015; Sporer, 1997; Sapir, 2005
Movement	Activity	Vrij & Mann, 2004; Harpster et al., 2009
Self-centred	Conditional Ididntdoit/Somebody/Idontknow	Harpster et al., 2009 Granhag et al., 2015; Davis et al., 2005
Grammar	Wronggrammar	Davis et al., 2005; Whelan et al., 2014; Zuckerman et al., 1981; Goldman-Eisler, 1968; Vrij & Mann, 2001; Whelan et al., 2014
Form/Structure	Eventref	Vrij, 2005a; Vrij et al., 2010

**APPENDIX O: Direct and indirect appeals according to appeal veracity**

<b>Appeals</b>	<b>Direct</b>	<b>Indirect</b>
Innocent	13	8
Guilty	7	11

**APPENDIX P: Monotonicity coefficients between  
variables with POSAC map axes *X* and *Y***

<b>Feature</b>	<b>Base coordinate</b>	
	<b><i>X</i></b>	<b><i>Y</i></b>
Self-centred	0.64	0.91
Vicpresenttense	0.86	0.58
Victim-centred	0.82	0.23
Social	0.80	0.35
Emotional	0.10	1.00
Grammar	0.74	0.59
Chronology	0.60	0.86
Movement	0.99	0.13