

How Do Decision Makers Decide When Witnesses Are Telling The Truth And What Can Be Done To Improve Their Accuracy In Making Assessments Of Witness Credibility?



**Report to The Criminal Lawyers Association of
Australia and New Zealand**

Professor Ian R. Coyle

3 April 2013

**HOW DO DECISION MAKERS DECIDE WHEN
WITNESSES ARE TELLING THE TRUTH AND WHAT CAN BE DONE TO IMPROVE
THEIR ACCURACY IN MAKING ASSESSMENTS OF WITNESS CREDIBILITY?**

Report to The Criminal Lawyers Association of Australia and New Zealand

Professor Ian R. Coyle¹

3 April 2013

ACKNOWLEDGMENTS

Studies involving practical aspects of the administration of justice and issues of forensic science are fraught with difficulties. Without the help, encouragement and assistance from those well versed in the law, they are likely to be relegated to academic obscurity and have no relevance to the most pressing matters involving the day-to-day administration of justice. Equally, without apposite scientific input, such studies are useless. Fortunately, from the inception of this study and to its completion, I had most excellent assistance from both sides of the legal and forensic scientific fence. This assistance is acknowledged in what follows.

This study was made possible by a gift to Bond University Centre for Law Governance and Public Policy by the Criminal Lawyers Association of Australia and New Zealand. This is most gratefully acknowledged, as is the personal assistance of Mr Kevin Borick Q.C., who was instrumental in securing funding to allow this study to proceed and for his most helpful comments upon the draft report. The author would also like to thank Professor Patrick Keyzer and Ms Veronica Jones, Centre for Law, Governance and Public Policy, Faculty of Law, Bond University for their personal encouragement and the highly professional administrative support afforded by the Centre. Professor Don Thomson, apart from his inestimable help in taking part in this study by virtue of being videotaped for one of the DVDs employed therein, also provided most incisive comments in the planning of this study and on the draft report. Associate Professor David Field deserves the most fulsome consideration for his involvement in the preparation of both the DVDs used in this study and for his comments on the draft report. Thanks are also due to Emeritus Professor Ronald Francis for his sage advice both during the planning for this study and for his comments upon the draft.

Professor Ian R. Coyle, 3 April 2013

SUMMARY

Correctly determining witness credibility is integral to a fair trial. Assessments of credibility by the triers of fact are made, *inter alia*, by reference to behavioural stereotypes that are commonly thought to be associated with lying and truth telling. These stereotypes are worthless but pervasive. In this study, venire jurors were given information such as would be given by way of judicial direction and/or expert testimony on those behavioural indicia that are useful in detecting deception. Major changes in venire juror perceptions of what does and does not work were found. This has significant implications for the conduct of criminal trials. Recommendations are presented which, it is argued, can be of real, practical assistance in enabling decision makers to assess the credibility of witnesses.

INTRODUCTION

Objective and Perceived Truth

A trial is an attempt to recreate the past. It involves the presentation and assessment of physical evidence and what may be referred to as memory-trace evidence from eyewitnesses that the triers of fact assess to determine the truth of competing claims. Truth, however, is a somewhat elusive concept and it is necessary to discriminate between perceived and objective truth, as a yawning gulf can exist between them. For example, a common and particularly virulent problem in forensic contexts leading to false perceptions of the truth is post-event contamination of memory, which results in false memories. Thus, a witness can be truthful but nonetheless mistaken. Although the courts have long recognised the potential for prejudice and bias that can occur in eyewitness evidence, attempts to remedy this situation have been limited to giving judicial directions and exhorting jurors to apply their common sense, ill-defined though this is, and their life experiences (Coyle, Field & Miller, 2008).

Prejudice and bias in eyewitness evidence are not the only problems to infect the curial processes. To be sure, there are many other elements of psychology which affect the perception of truth that are much more subtle and require considerable scientific expertise to discern. In many cases, the gap between objective and perceived truth is not so much a yawning gulf as a pygmy strait, but the effect is the same. For example, cognitive biases and heuristics affect our capacity to comprehend statistical information routinely presented in trials involving DNA, latent fingerprint evidence and actuarial information on risk of recidivism. It has been conclusively demonstrated that perceptions of the truth and objective, mathematical, determination of the truth are often very far apart in such cases (Tversky &

Khaneman, 1974; Tversky & Kahneman, 1981). This quintessential element of being human is inimical to the course of justice, yet it has received virtually no attention from the superior courts.

Determining when and how perception, memory and cognition can be affected so as to lead to confusion between perceptions of truth and objective truth involves the study of human behaviour, which is the domain of psychologists. Yet there has been great reluctance to adduce the evidence of psychologists when it is perceived as trespassing upon the finality rule in relation to collateral issues. At the risk of simplification, this rule deals with matters solely going to the credit of a witness. The rule arose so as to impose some sort of reasonable time limits on court proceedings.²

Justice McClellan (2006) made one of the bluntest judicial observations on this point when, in commenting about psychology, common sense and the law, he noted [at paragraph 19]:

Law and psychology can be uneasy partners. The law has traditionally devised its own rules of human behaviour and created its own norms for interpreting that behaviour. Informed by little more than the appellate court's understanding and often classified under the rubric of "common sense" judges are required to direct jurors in a particular manner on a whole range of subjects, with varying degrees of impact on the outcome of the trial – some more easily identifiable than others.

Juror's Assessment of Human Behaviour

For over two centuries the observations of Lord Mansfield in *Folkes v Chadd* (1782) have stood sentinel to the capacity of juries to discharge their fact-finding functions without the assistance of expert opinion in matters of human behaviour. Therein it was noted:

The fact that an expert witness has impressive scientific qualifications does not by that fact alone make his opinion on matters of human behaviour within the limits of normality any more helpful than that of jurors themselves; but there is a danger that they may think it does.

Despite the undoubted advances in medicine and the behavioural sciences that have occurred since *Folkes v Chadd*, the thrust of this opinion still exerts influence on judicial thinking. Thus, Lawton LJ in *R v Turner* (1975), reading the unanimous judgment of the English Court of Appeal, noted [at 843] that: “psychiatry has not yet become a satisfactory substitute for the common sense of juries...within their experience of life.”

This, and similar judicial utterances over the years, suggest that juries are empanelled in the belief that they bring with them into the criminal justice process an awareness of normal human weakness – whether physical or mental – and the usual consequences of the vicissitudes of everyday life. Equally, it has been acknowledged that human behaviour *outside* the “limits of normality” is an area in which jurors without the benefit of academic qualifications in the behavioural/mental health sciences most certainly *will* require assistance from someone who has. In *Murphy v The Queen* (1989), in which the issue was whether or not certain words and phrases used in a contested confession were consistent with those of a person who, like M, had an intellectual age of ten, it was held per Deane J [at 126], that the expert evidence was admissible on that issue, since: “no-one but an expert could properly give evidence grading [the accused’s] level of intellectual functioning...Certainly, the ordinary juror would lack the specialised knowledge and experience.”

The problem areas are to be found within these two extremes, and most of them concern areas of human behaviour that are well within the specialised fields of study of behavioural scientists, but are also arguably contained within normal life experience, and therefore require no scientific elaboration to assist the jury in their fact-finding function. One such area is what might be termed behavioural assessment of witness credibility.

Witness credibility

One “life experience” topic that has straddled the threshold of expert evidence admissibility for many years is that of witness credibility, a matter traditionally regarded as the sole province of the jury in a criminal trial. The historical, and still orthodox, position was best expressed in the majority opinion of the Canadian Supreme Court in *R v Marquard* (1993) held [at 228] that:

It is a fundamental axiom of our trial process that the ultimate conclusion as to the credibility or truthfulness of a particular witness is for the trier of fact, and is not the proper subject of expert opinion.

It is conventional legal wisdom in Australia (see, for example, *R v Martens* (2007)) that this arrangement presents no difficulty, since “witness demeanour” is something that can be safely left to a jury where there is no reason to suspect that the witness is in any way different from his or her fellow citizens. It is, after all, the very “normality” of a jury that is being called in aid to assess the believability of one of their peers, and this is one of the main arguments in favour of jury trial (Connor, 1987).

However, if – unknown to the jury – there are factors present which ought to be taken into account before assessing whether or not a witness should be believed, then should that jury not be advised of those factors? After all, evidence regarding provable witness bias, and a witness’s previous convictions, has long been regarded as admissible on the issue of their credibility.³

In *R v BDX* (2009) the Victorian Court of Appeal held, in a trial for incest in which the credibility of the victim was the central issue in the case, that the defence should have been allowed to adduce the evidence of a former employer of hers, to the effect that she “wouldn’t believe a word she said”, even on oath. This revived a common law exception to the “collateral issues” rule that many observers had come to believe had fallen into disuse. The witness who was allowed to testify in that case was not an “expert” in the professional sense, but someone who knew the witness very well. While this made her “an expert by experience” in the matter to which she was testifying, this fell a long way short of authorising such evidence being given by an expert with relevant qualifications.

In *Aytugrul v The Queen* (2012) Heydon J neatly encapsulated the seminal issues involved in having the superior courts attempt to deal with expert psychological evidence at appeal without such evidence being adduced at trial. The appeal involved issues of expert statistical knowledge and the effect on juror’s perceptions of the evidence based on the manner in which the statistics were presented as to the frequency of a particular type of mitochondrial DNA in various populations. The manner in which statistical evidence is presented has been the subject of significant psychological research and judicial observations in the various countries

that have embraced the common law (Coyle & Halon, 2013). In *Aytugrul*, His Honour commented as follows [at paragraph 71-citations have been inserted]:

Works on psychology have been considered in formulating rules about identification evidence (*Winmar v Western Australia* [2007] WASCA 244; (2007) 35 WAR at 167 and 171-172) both directly (*Smith v The Queen* [2011] HCA 50; (2001) 206 CLR 650 at 667-668) and indirectly (*The People (Attorney-General) v Casey* (No 2) [1963] IR 33 at 39; *R v Gaunt* [1964] NSW 864 at 866). This is not surprising, since the court's recognition of the "inherent frailties of identification evidence" has been said to arise "from the psychological fact of the unreliability of human observation and recollection (*R v Sutton* [1970] 2 OR 358 at 368 per Jessup JA (Gale CJO and Kelly JA concurring), approved in *Alexander v The Queen* [1981] HCA 17; (1981) 145 CLR 395 at 435). *If frailty rests on a psychological fact, and on psychological research (Pollitt v The Queen (2001) HCA35; (1992) 174 CLR 558 at 615), expert material bearing on the psychological fact must have potential significance* (emphasis added). Works on psychiatry have also been considered in explaining why children delay in complaining of sexual assault in relation to the unsafe and unsatisfactory ground of criminal appeal (*Jones v The Queen* (1997) 191 CLR 439 at 463; [1997] HCA 56)...Psychiatric studies on the harm suffered by child victims of sexual offences have been taken into account in developing sentencing principles (*Ryan v The Queen* (2001) 206 CLR 267 at 281; [2001] HCA 21) ...

His Honour then went on [at paragraph 74] to note that if expert evidence:

...were to be taken into account, it was highly preferable that it be presented through expert witnesses, preferably during a pre-trial hearing to determine admissibility. The admissibility and weight of the expert material could then be considered publicly and critically.

Without explicitly commenting upon cognitive heuristics such as framing bias (Tversky & Kahneman, 1974), which affects judgment under conditions of uncertainty (an issue that was exquisitely relevant to the matters raised in *Aytugrul v The Queen*), it is impossible to

properly ascertain the weight, if any, that should be given to expert evidence such as those that were the subject of His Honour's reflections. A proper assessment of the weight that should be afforded to expert evidence on this topic involves complex issues of psychology and epistemology as they apply to the finality rule in relation to collateral issues. Acknowledging this, the decision in *Aytugrul v The Queen* clearly leaves the door open for expert psychological evidence dealing with psychological facts of *all* kinds to be adduced in evidence. In what follows, psychological facts are considered as they apply to the vexed issue of witness credibility and detection of deception.

Witness Demeanour & Witness Credibility

The demeanour of witnesses is a fundamental issue in the assessment of credibility. Although the law is far from settled on this in Australia,⁴ it is considered axiomatic that triers of fact can, and should, consider witness demeanour. In practice, this means attending to verbal and non-verbal indicia of truthfulness, in addition to considering collateral information that may refute or confirm a witnesses' recollection. In many cases, where competing versions of events are given, the assessment of such indicia are of great import: sometimes it may be the only basis on which a decision can be arrived at. Legal training and experience confers no special benefit in detecting deception in forensic contexts. There is no compelling evidence which proves that lawyers generally, and judges in particular, operate at anything better than chance level when detecting truthfulness, although there is a danger that they may think otherwise. And there is an abundance of evidence that jurors attend to unreliable indicia of deception when making their decisions.

In commenting on the problems recognised by the common law of the injustices that have arisen from the large number of "exceptions" to the rules of evidence, Bagaric and McConvill (2005) have made the pungent observation [at 12]:

We also need rules of evidence to be based not on the intuition of lawyers, but rather on the knowledge of experts in the area. We should be listening to behavioural scientists, not more lawyers - they are the ones that got us into this hole in the first place.

This observation is supported, albeit obliquely, in the Equal Treatment Benchbook of the Supreme Court of Queensland wherein, in relation to the putative problems associated with

respect of cultural misunderstanding of witnesses' responses [at 76] it is opined that:

It may be necessary to receive expert evidence in this regard. The areas of potential misunderstanding are likely to include, politeness, body language, power dynamics, metalinguistic factors such as pitch, volume and silence.

This holds out the prospect of relief from the reluctance of the courts to allow such evidence to be adduced: this prospect is a chimera. The thrust of the Equal Treatment Benchbook serves to exclude the reception of expert evidence in such matters for Anglo-Australians.⁵ Although it may be an unintended consequence, Anglo-Australians, as will be apparent in what follows, seem not to be afforded the same consideration given to those of different ethnicity.

What is to be done to overcome such a state of affairs? One way forward has been suggested by Bagaric and McConvill (2005) who propose [at 25] that:

Consideration should be given to having an expert in behavioural science observe the trial process and give expert evidence on the veracity of witnesses. This can only be evaluated by undertaking controlled experiments during mock trials.

There are numerous problems associated with this proposal. For a start, in those jurisdictions that have enacted the Uniform Evidence Laws expert evidence as to the credibility of a witness may only be adduced with the leave of the court: this is by no means guaranteed (see, generally, Coyle, Field, Wilson, Cuthbert & Miller, 2009). In other jurisdictions, the situation is far less certain. It is recognised that section 108(c) of the Uniform Evidence Law allows, in the Commonwealth and those states which have enacted this legislation, expert evidence to be given to as to the credibility of a witness, with the consent of the court. The difficulty is, even where the Uniform Evidence Law is applicable, that such evidence should not, as Kirby J noted in *Farrell v The Queen* (1998) “seek to usurp the decision-makers ultimate assessment of the credibility of the witness” (for discussion see: Coyle, Field, Wilson, Cuthbert & Miller, 2009).

There are also practical problems with having an expert comment on the veracity of

witnesses' evidence given during a trial. Nonetheless, considering the thrust of decision in *Aytugrul v The Queen*, an expert could, *at the very least*, give evidence to assist decision makers as to reliable indicia of truthfulness/deception so as to assist them in their deliberations, without necessarily venturing an opinion as to the credibility of any particular witness.

Witness Demeanour and the Detection of Deception

The detection of deception in forensic contexts has been the subject of voluminous research and many of the characteristics of witnesses' demeanour that decision makers can be expected to attend to have been the subject of extensive research. This research, and the consensus of scientific knowledge that has developed flowing from this research have been almost completely ignored by the law.

Thus, in the Equal Treatment Benchbook of the Queensland Supreme Court (2006) [at p.75] it is stated that:

An impressive witness according to Anglo-Australian culture will look his or her questioner in the eye and answer questions confidently and clearly. In other cultures, however, direct eye contact maybe considered rude and challenging... such responses may be misunderstood as demonstrating evasiveness or shiftiness on the part of the witness."

This unqualified assertion is spoiled by an inconvenient fact: *gaze aversion is the most unreliable indicator of deception in all countries that have been studied*. This misconception it is not unique to English speakers. In a comprehensive, worldwide study (The Global Deception Research Team, 2006), of 2,520 lifelong residents of 63 countries, 71.5% believed that gaze aversion was a reliable indicator of lying. It is, therefore, not surprising that mock jurors are more likely to perceive that expert witnesses, particularly if they are male, have greater credibility if they maintain high levels of eye contact (Neal & Brodsky, 2008).

Techniques for Detecting Deception

Given the importance of evaluating the credibility of witnesses' statements, specific techniques have been developed to assess the validity of claims (Fritzon, 2005; Köhnken, 2004; Raskin & Esplin, 1992). These may be grouped under three main types (Vrij, 2000;

2008). First, observing how people behave: the movements they make gaze aversion, stuttering, their pitch of voice and so on. Second, measuring physiological responses such as the galvanic skin response, which is one of the physiological indicia measured by polygraphs. Third, analysing the speech content.

Measuring putative physiological indicia of lying is obviously not an option for jurors, or experts for that matter, during the course of a trial, neglecting the controversy that surrounds this approach generally and more specifically in the Australian forensic context. The exception to this general statement is the possible utility of measuring the pitch of a witnesses' voice relative to baseline performance when they are giving evidence (Vilar, Arciuli & Paterson, 2013). However, this too has significant practical difficulties. The formal analysis of speech content is not an option for jurors, albeit that various methods of speech analysis are commonly used in other jurisdictions and, at the very least, hold great promise (Memom, Fraser, Colwell, Odinet & Mastroberardino, 2010; Sporer, 2004; Vrij, 2005).⁶ Accordingly, for all practical purposes, the only validated technique that can be employed by decision makers during the course of a trial in assessing witness credibility is to consider those behavioural indicia that are indicative of lying/truth telling and consider the plausibility of the answers given, particularly in the context of collateral information from other witnesses.

Behavioural Indicia of Lying and Truth Telling

Apart from gaze aversion, other global, pan-cultural, stereotypes of lying included the entrenched beliefs that liars shift posture (65.2%), touch and scratch themselves (64.8%) and tell longer stories than usual (62.2%). None of these stereotypes are accurate (The Global Deception Research Team, 2006). More generally, a very significant proportion of laypersons' and professionals' beliefs about non-verbal and verbal cues to deception have been repeatedly demonstrated to be incorrect (De Paulo, Lindsay, Malone, Muhlenbruck, Charlton & Cooper, 2003; Mann, Vrij & Bull 2004; Mann & Vrij, 2006). Apart from the ubiquity of gaze aversion, frequent blinking, fidgeting with objects or self, self-grooming, shifting posture, rate of speaking and brow lowering, none of these indicia are reliable indicators of lying.

Spontaneous corrections and admitting lack of memory are negatively correlated with lying, which is contrary to commonly held beliefs. That is, in ordinary conversation when people

recall events they tell them in a way that is not perfect. They might start to tell a story then realise they have forgotten something and go back and correct themselves. These are signs of credibility. In essence, a significant proportion of cues to lying are counter-intuitive for most laypeople, and for many professionals for that matter. This is especially so for non-verbal behaviour (Bond Jr et al., 2006; De Paulo et al., 2003).

It is not only judges who typically operate at chance level in detecting lying. There is no evidence whatsoever that suggests that police officers, other criminal investigators, social workers or child care workers are better at detecting untruthful/incorrect answers in forensic contexts than anyone else (Porter & ten Brinke, 2012; Vrij, Mann, Robbins & Robinson, 2006). Typically, these groups also perform at about chance level when trying to determine whether an individual is truthfully recalling something the individual claims to have observed or experienced if they rely on unstructured judgment. The more experienced they are the more confident they feel in their assessment but this confidence is entirely unsupported by evidence. For example, there is evidence that police officers tend to over-emphasise non-verbal cues at the expense of more reliable indicators of lying such as changes in pitch, hesitation and speech error and micro-momentary facial expressions of emotion (Bull, 2004; O'Sullivan & Ekman, 2004).

As far as observing behaviour is concerned, there are some reliable indicators of truthful recall, such as change in pitch of voice, time taken to respond to questions and frequency of hand movements designed to illustrate or supplement verbal utterances (De Paulo et al., 2003; Mann et al., 2004; 2006; Vrij, 2008). To reiterate, gaze aversion, which is often thought to be a reliable indicator of lying is not, in fact, a good indicator of lying; nor are excessive finger and hand movements. Conversely, lack of body movements is a good indicator of lying. This deserves some explanation since it appears counter-intuitive. The reason is that because telling lies is more cognitively intensive than truth telling, liars minimise extraneous movements so as to concentrate on the lies they are presenting.

The problem with attempting to determine some the behavioural indicia that are diagnostic of lying/truth telling in the forensic context is simple: courtroom design, or the lack thereof. The design of many courtrooms does not permit the triers of fact to see the witness with sufficient clarity to be able to observe minor movements of the facial features, which can be a good indicator of deception according to some authors. Further, other indicators of deception (such

as lack of movement of the feet) simply cannot be observed at all due to the design of the witness box. In fact, in many courts, the jurors and judge can only see a side profile of the witness.

That said, the most reliable general non-verbal and verbal indicators of lying, as distinct from nervousness, which is often conflated with lying, are as follows (De Paulo et al., 2003; Strömwall, Granhag & Hartwig, 2004; Vrij, 2008; Vrij, Granhag & Porter, 2010; Vilar, et al., 2013):

- Liars tend to speak in a higher pitch relative to baseline performance.
- Liars make fewer movements with hands/arms/fingers.
- Liars use fewer illustrators (hand and arm movements to demonstrate and illustrate points they are making).
- Liars take longer pauses.
- Liars make fewer movements with their hands and feet.
- Liars' answers are less plausible.
- Liars' stories contain fewer details.
- Liars give more indirect answers.
- Liars contain less temporal, perceptual and spatial information.
- Liars make fewer spontaneous corrections.

It needs to be emphasised that, in the absence of compelling collateral information that disproves a version of events, there is no one indicator of lying that is universally accurate. That is, there is no such thing as Pinocchio's nose.

Although there is general consensus in the literature that the behavioural indicia iterated above are reliable, there remains the problem of determining the baseline, or usual behaviour, of individuals when they are telling the truth. After all, it is departure from individuals' usual behaviour when they are telling the truth that enables observers to detect when someone is lying insofar as they rely upon behavioural indicia. Thus, it is easier to detect if someone well known to the observer is lying as opposed to a stranger because departure from their usual mannerisms/ behavioural indicia can be compared to those present when lying was not an issue. For example, it is within common experience that some individuals will naturally speak in a higher pitched voice than others. If an individual's typical speech pattern is well

known to an observer it is far easier to detect departures from their normal pitch of speaking than if this is not known. *Inter alia*, this is why it is far easier for parent's to determine if their children are lying than complete strangers.

Training in Detecting Deception and Truth-Telling

Summarising studies of training in typical interrogation deception detection, Reidlich and Meissner (2009) observed [at 126] that:

Even professionals who have to make daily decisions of whether people are lying do not demonstrate high rates of accuracy when detecting deception (Meissner & Kassin, 2002; O'Sullivan & Ekman, 2004). Indeed, training on typical interrogation deception detection techniques has been shown to have a deleterious effect on accuracy (Kassin & Fong, 1999; Meissner & Kassin, 2002; see also, Bond & DePaulo, 2006). That is, studies with college students and police officers found that trained participants were less accurate than naïve participants, but were nevertheless significantly more confident in their abilities to detect deception.

While a significant proportion of studies where training has been given in detecting behavioural indicia indicative of lying and truth telling have failed to demonstrate robust results, this does not prove that such training and/or education is doomed to fail. The most parsimonious explanation for this failure to improve detection of deception is training in the wrong techniques (Frank & Feeley, 2003). Even the most cursory review of many police training manuals, for example, demonstrates glaring errors in so-called deception training and interview techniques. The Reid Technique, which has been employed for decades in the USA in various iterations, is but one of a number of egregious examples (Inbau, Reid & Buckley, 1986) that fail to take into account fundamental psychological principles involved in interrogation and/or detection of deception, as has been commented on by Gudjonsson (1992) and others.

Then there is the problem that the detection of deception is intrinsically difficult. Humans have evolved to lie and lies have social benefit in many situations: imagine a world without *any* lies, however well intentioned, for a moment. While humans have also evolved to detect lies, the contest between liars and those seeking to catch out the liars is fraught with difficulties. Hartwig and Bond Jr (2011) argue that the failure to reliably detect lies is not

because people rely on the wrong cues, but it is attributable to the weakness of behavioural cues to deception. These authors may well be correct in saying that the behavioural differences between liars and truth-tellers are small but it does not proceed from this that inability to discriminate between liars and truth-tellers is due solely to inability to detect small differences in relevant behavioural indicia. For example, it has been demonstrated that observers can reliably detect behavioural differences between liars and truth-tellers and that they can make accurate veracity judgements on such observed differences (Vrij, Akehurst, Soukara & Bull, 2006). In addition, it has been repeatedly demonstrated that poor accuracy in discriminating liars from truth-tellers is largely due to individuals paying attention to behavioural indicia that are *not* diagnostic of deception, such as gaze aversion (Hartwig et al., 2004; Kassin & Fong, 1999; Mann et al., 2004; Vrij & Mann, 2001).

It proceeds from this, that one way of improving the detection of deception is to improve the abilities of triers of fact to discriminate between liars and truth tellers. There is a number of ways in which this can be achieved, such as increasing the cognitive workload of the suspected liar, which would be expected to increase the frequency of behavioural indicia which are associated with deception (Vrij, Mann et al., 2010; Warmelink, Vrij, Mann, Leal & Poletiek, 2013). Most of these are unlikely to be of any great value in a courtroom for the triers of fact, unless they were very well trained indeed. Having said this, after participants were given instruction vis-à-vis behavioural cues that are diagnostic and non-diagnostic of deception and/or undertook some practise/feedback tasks (Porter, McCabe, Woodworth & Pearce, 2007; Porter, Juodis, ten Brinke, Klein & Pearce, 2010) their discrimination between liars and truth tellers improved significantly. This, then, is worthy of investigation with venire jurors, insofar as it has practical and theoretical implications both from a forensic psychological and jurisprudence perspective.

Judicial Directions and Expert Evidence on Detecting Deception and Truth Telling

It is an article of faith amongst the judiciary that properly framed judicial directions are capable of effectively educating jurors on many topics. The evidence does not support this entrenched belief; jurors' comprehension of judicial instructions is generally poor (Blankenship, Luginbuhl, Cullen, & Redick, 1997; English & Sales, 1997; Luginbuhl, 1992; Ogloff & Rose, 2003; Ogloff, Nadjovski-Terziovski, Spivak, & Clough, 2011; Rose & Ogloff, 2001; Severance & Loftus, 1982; Spivak, Ogloff, & Clough, 2011). In the Australian forensic context, this problem extends to judicial directions such as *Longman*, *Domican* and

Alexander directions (*Alexander v The Queen* (1981); *Longman v The Queen* (1989); *Domican v the Queen* (1992)) routinely employed in all areas of the criminal law where witnesses' credibility is at issue. As Coyle and Field (2013) note [at 101] in connection with the efficacy of judicial direction in complex areas of psychology:

It is independent of whether or not these directions are framed as obligatory directions or as opinions that the jury is free to accept or reject. It requires a suspension of disbelief to accept long cherished legal maxims as to the efficacy of judicial directions in such circumstances.

Although some studies that have found that using judicial instructions and/or expert testimonies caused a response shift in the perceptions that venire jurors commonly have of eyewitness identification and rape (Cutler, Dexter and Penrod, 1989) these results are not unequivocal. To amplify this point, Ellison & Munro (2009) found that such education did not dispel the stereotype that genuine rape victims always fight back. Considering the manifest problems identified in jurors comprehending judicial directions it is stretching a long bow to argue that such direction can be expected to cause a significant response shift in jurors' perceptions. However there are, to the author's knowledge, no studies on the efficacy of judicial instructions and expert evidence vis-à-vis sensitising jurors to the behavioural cues that are diagnostic of deception and truth telling.

One way of considering this problem is to ask a very simple question: does education of venire jurors, such as may be practically obtained via judicial direction and/or expert testimony during the conduct of case, change their perceptions as to what behavioural indicia of deception/truth telling discriminate between liars and truth tellers? After all, if this does not change perceptions of what does and does not work, which is seriously flawed in virtually all venire jurors, then there is simply no point in bothering about the ultimate effect of mooted judicial directions and/or expert testimony on their assessments of credibility of witnesses. This question was the focus of this study.

METHOD

Participants, who remained anonymous, were attracted by advertising and a snowballing technique and were paid \$50 to participate. They were a mixture of the general public and university undergraduates/graduates. The participants' pre-conceptions about which indicia of

truth telling or deception are, or are not indicative of deception, was determined by a questionnaire (see Appendix 1), which measured their baseline knowledge vis-à-vis diagnostic behavioural indicia of deception. The questionnaire contained items that extensive research has demonstrated to be diagnostic and not diagnostic of deception. The order of presentation of the questionnaire items was randomised, to prevent participants from deducing which behaviours are diagnostic of deception and which behaviours are not. To prevent response set bias, the verbal anchors of the Likert scales were randomly presented in increasing or decreasing numerical order. Verbal anchors were paired with each number in the six-point Likert scale employed. A six-point Likert scale was employed so as to enable comparisons with the mean.

The participants were randomly assigned to one of three experimental conditions: JR1- JR3. In JR1, after completing the baseline questionnaire (Q1), the participants read what they were informed were a judge's directions to a jury (see Appendix 2). They then completed the questionnaire again (Q2). In JR2, the participants, after completing Q1, saw a short DVD (10 minutes duration) of an expert giving video-link evidence on behavioural indicia that are diagnostic of deception before completing Q2. The "evidence" given therein mirrored the information contained in the "judge's direction to the jury" utilised in JR1. JR3 was the same as JR2, with the exception that, after seeing the DVD, participants also saw an extract from a DVD where an expert witness was being cross-examined. The participants in JR3 were not asked to make a credibility assessment of the witness (who was lying): the second DVD was simply there to provide a context in which they could consider the initial DVD. Both the DVD's were taped in courtrooms and both experts were full professors in forensic psychology with more than three decades of experience as expert witnesses.

RESULTS

Statistical Concepts

Determining whether two or more groups, of whatever type, truly are different or if the difference is merely due to some random event is a common problem. Suppose, for example, that one wished to test the hypothesis that all humans born with hands had five fingers. The only way this could be completely proven would be to observe every human on the planet. However one could disprove the opposite, or null, hypothesis, to a specific degree of certainty or probability, by observing a sufficiently large sample of humans, which would then infer that the original hypothesis was correct. The degree of certainty in such a process is never

absolute; it may approach a probability of $p=1.0$ (i.e. 100%) but it can never obtain this level of certainty unless *all* members of a class (in these case humans) are observed. This example is not far-fetched or fanciful since polydactyly (having more than five digits on either hand or foot) clearly exists, albeit it is rare.

Thus, inferential statistical tests enable the probability that a difference between groups is due to some casual factor rather than mere chance variation to be inferred. This procedure was used in the Matched Pair analysis where the effects of the various education/training were measured on the same individuals. In other words, the probability that the difference in the same group of participants' perceptions of what behavioural indicia of deception discriminate between liars and which do not was due to a real effect and not chance was determined. By convention, a result is statistically significant if the probability of it occurring by chance is a minimum of $p < 0.05$. That is, this result will only occur by chance less than 5% of the time and so there is at least a 95% probability that the results actually reflect a true underlying difference. The lower the statistical significance level, the higher the probability that the difference is real.

Factor Analysis involves a totally different concept. Factor analysis seeks to discover simple patterns in the pattern of relationships among variables, in the case of this study individual questionnaire items. More specifically, it seeks to discover if the observed variables can be explained largely or entirely in terms of a much smaller number of variables called *factors*. Thus rather than trying to consider differences between all 31 questionnaire items in this study, Factor Analysis seeks to identify a much smaller number of underlying factors in the individual questionnaire items that “go together” in a mathematical sense.

A perfect Factor Analysis solution would account for all the variability (i.e. variance) in the sample of things studied (in this case questionnaire items) but this almost never happens. In practice, there are usually a small number of factors that account for the majority of the variance with the rest being almost completely meaningless in real terms. The higher the proportion of variance accounted for by any particular factor the more important it is in what might be called “explanatory power.” This explanatory power is maximized, in a mathematical sense, in the final “rotated factor pattern.”

The importance of each questionnaire item on any particular factor is determined from the final factor pattern with correlations ranging from 0.3 (by convention, any correlation below this was considered to be too small to be important and was discarded from the final analysis) up to a theoretical maximum of 1. It is important to note that the higher the correlation coefficient the bigger the contribution of any specific item to any specific factor. To the extent that different patterns of factors are identified, particularly with respect to the number of items indicative of deception contributing to (loading on in mathematical terms) factors as a result of the education/training the participants received, this would provide support for the notion that such training was effective. Such a result would also confirm the result of the Matched Pair analysis.

Demographics

A total of 98 valid questionnaires were completed: one was rejected, as some questionnaire items were not coded. Demographic statistics are shown in Table 1.

Table 1 – Demographic Information

	Age		Sex		Education	
	Mean	s.d. ^a	Male	Female	High school qualifications	Completed or partially completed university studies
Total of all groups	33.76	15.04	42.7%	57.3%	41.8%	58.2%
JR1 (n=31)	26.74	9.75	48.4%	51.6%	25.8%	74.2%
JR2 (n=30)	31.87	14.68	43.3%	56.7%	46.7%	53.3%
JR3 (n=37)	30.24	11.65	40%	60%	51.4%	48.6%

a. Standard Deviation.

Matched Pair Analysis

For some questionnaire items, low scores were indicative of deception (items 14, 28 and 30). Scores on these items were transformed prior to analysis so that higher scores were always indicative of deception (i.e. 1 was transformed to 6, 2 to 5 etc.). The scores on items indicative of deception (item numbers: 2, 3, 6, 11, 14, 17, 19, 21, 25, 28, 29 & 30) were summed for each participant and Wilcoxon Signed Tests were conducted to determine whether the mean scores on items indicative of deception diverged between Q1 and Q2 for all three experimental groups. One-tailed tests were employed as differences were only to be expected in increased scores on items indicative of deception. For JR1, the mean score (rounded) of summed items indicative of deception on Q1 and Q2 was 3.034 and 3.634

respectively (mean difference=0.598, standard error=0.175); $S=141.50$, $p<0.001$ (one tailed). For JR2, the mean score (rounded) of summed items indicative of deception on Q1 and Q2 was 3.345 and 3.62 respectively (mean difference=0.275, standard error=0.130); $S=76.0$, $p<0.04$ (one tailed). For JR3, the mean score (rounded) of summed items indicative of deception on Q1 and Q2 was 3.316 and 3.753 respectively (mean difference=0.437, standard error=0.091); $S=235.50$, $p<0.001$ (one tailed).

Factor Analysis

The pre-intervention questionnaire (Q1) and post-intervention questionnaire (Q2) were analysed using Kaiser-Meyer-Olkin (KMO) Maximum Likelihood Factor Analysis with Varimax Rotation, employing Kaiser normalisation, using the SAS statistical programme. In this analysis the three experimental groups (JR1-JR3) were collapsed. Prior to final rotation individual items loading less than 0.3 on any factor were discarded. The individual items loading on the final rotated factor solutions for Q1 and Q2 are set out in Tables 1 and 2: items indicative of deception (either by their presence or absence) are indicated by shading.

KMO Measures of Sampling Adequacy for Q1 and Q2 were 0.728 (Approximate Chi-square=1,308.475) and 0.859 (Approximate Chi-square=1,703.22) respectively. Bartlett's test for Sphericity was significant for Q1 (df, 465, significance <0.000) and Q2 (df, significance < 0.000). These results strongly support the validity of factor analysis. Tables 2 and 3 set out the rotated factor matrices for Q1 and Q2. The order of presentation of the items loading on factors has been standardised to enable ease of comparison.

For Q1, 9 factors, accounting for 53.71%, of the variance were identified. For Q2, 6 factors, accounting for 54.12%, of the variance were identified. No statistical test was employed to determine whether the factor patterns diverged, as none was considered appropriate. Either the factor patterns are identical or they are not: they are not. As can be seen from inspection of Tables 2 and 3, the number of items indicative of deception loading on Factor 1 in Q1 and Q2 was 4/12 and 9/12 respectively. Further, in Factor 1, in Q2 accounted for approximately double the variance of Factor 1 in Q1.

Table 2 – Q1 Rotated Factor Matrix ^a

	Factor								
	1	2	3	4	5	6	7	8	9
Ambivalent responses			.395		.351				
Body movements				.392		.415			
Delayed responses	.307				.434				
High pitched voice							.915		
Illustrators			-.690						
Implausible responses	.378	.407	.458						.626
Lacks logical structure	.413		.537			.406			
Pauses							.368		
Responses - few details	.626		.471						
Responses- many details								.534	
Short responses						.353			
Spontaneous corrections					.441				
Anxious Behaviour	.663	.422							
Blinking				.672					
Confidence			-.447						
Covering the mouth		.906							
Crossed Arms				.421	.459				
Fidgeting	.391		.446	.382					
Gaze Aversion			.532		.554				
Interrupting the questioner	.335							.472	
Mumbling		.665							
Nervous behaviour	.675								
Pressed lips				.513				.348	
Repetition words/phrases						.669			
Restless behaviour	.336						.310		
Self-grooming				.511					
Self-manipulators/scratch		.353		.445		.319			
Speech disturbances	.432	.300						-.466	
Stuttering	.347								
Tenseness	.596								

a. Rotation converged in 24 iterations. Shaded items are indicative of deception. Variance accounted for (rounded): Factor 1=10.24%, Factor 2=7.7%, Factor 3=7.6%, Factor 4=7.4%, Factor 5=5.1%, Factor 6=5.1%, Factor 7= 4.6%, Factor 8=4.0%, Factor 9= 2.1%. Total Variance accounted for by Factors 1-9: 53.71%.

Table 3 – Q2 Rotated Factor Matrix ^a

	Factor					
	1	2	3	4	5	6
Ambivalent responses	.730					
Body movements				.562		
Delayed responses	.566					
High pitched voice	.619			.381		
Illustrators		.483				
Implausible responses	.796					
Lacks logical structure	.813					
Pauses	.650					
Responses - few details	.809					
Responses - many details				.522		
Spontaneous corrections	.471		.347			
Short responses	.474					
Anxious behaviour	.478	.485				
Blinking		.558				
Confidence			.484			
Covering the mouth	.485	.353				
Crossed arms		.678				
Fidgeting	.332	.307	.446		.644	
Gaze aversion	.418	.389	.426			
Interrupting the questioner	.348		.435			
Mumbling	.612		.446			
Nervous behaviour			.347			
Pressed lips		.609			.376	
Repetition of words/phrases	.338		.438			
Restless behaviour	.552	.394	.327			.468
Self-grooming			.569			
Self-manipulators/scratch		.506	.444			
Speech disturbances	.369		.350	.568		
Stuttering	.485	.353				
Tenseness	.455	.541	.378			

a. Rotation converged in 7 iterations. Shaded are indicative of deception. Variance accounted for (rounded): Factor 1=21.60%, Factor 2=11.98%, Factor 3=8.91%, Factor 4=6.1%, Factor 5=3.6%, Factor 6=1.98%. Total Variance accounted for by Factors 1-6: 54.12%.

DISCUSSION

This study provides support for the proposition that information provided to jurors *either* by way of judicial directions or by way of expert evidence can correct venire juror's well-entrenched misconceptions of which behavioural indicia are indicative of deception. This was unexpected since it was hypothesised that judicial directions would not have the same effect as the presentation of oral expert evidence. Conversely, the finding that oral expert evidence in conjunction with a context in which to consider such evidence resulted in a more pronounced main effect than all other groups was expected. What might explain these findings?

First, the judicial directions were provided in written format and took less time to read than it did for participants in the other two experimental groups to listen to oral expert evidence covering the same main points. Conceivably, this could have resulted in less decay of memory for the group provided with written directions. Of course, the group with the greatest time between exposure to oral expert evidence and completing the post-intervention questionnaire was the group who also saw the additional DVD (JR3). However, apart from providing context for them in which to consider the expert oral evidence they had viewed, this also provided them with the opportunity of rehearsing this information. Thus, the effects of decay would have been countervailed by rehearsal, albeit that such rehearsal may not have been deliberate. Second, the written directions were incorporated in a document comprising the baseline and follow up questionnaires. Although instructions were given, both in the questionnaire and orally by the supervising research assistants, to turn the page after all participants had completed the baseline questionnaire it is possible that some participants used these instructions to check their answers (i.e. they cribbed from them) when completing the second questionnaire. Third, there is some evidence from related research into the efficacy of judicial directions, that written judicial directions twice as effective as oral directions (Thomas, 2010). Assuming that this research is directly applicable to the current study this is of some note.

The factors derived in the rotated factor matrix are not amendable to categorical reduction insofar as verbal descriptors of these factors are not obvious. Further, to the author's knowledge no one has employed this technique in studies on deception and therefore the final rotated factor matrix should be regarded as exploratory. That said, the results of the factor

analysis show that the information participants in this study received resulted in a very significant increase in those items indicative of deception loading of Factor 1 in the post-intervention questionnaire (Q2), compared to the baseline questionnaire (Q1). Thus, to reiterate, considering Factor 1, for Q1 and Q2, the number of items indicative of deception loading on this factor included 4/12 items indicative of deception in Q1 and 9/12 items indicative of deception in Q2. Further, in Q2, Factor 1 accounted for *double* the variance of Factor I in Q2. The robustness of these results was unexpected. The factor analysis results further support the proposition that education, either by way of judicial directions and/or expert evidence as to which factors are indicative of deception can cause a major shift in perceptions of what works and what doesn't work vis-à-vis behavioural indicia of deception. By way of elaboration, the greatly increased proportion of questionnaire items that are indicative of deception loading on, or contributing to, the most important factor in Q2 (Factor 1) can only be realistically explained as being due to the exposure of participants to information as to what works and what doesn't work in detecting deception on the basis of behavioural indicia. The fact that other questionnaire items not indicative of deception also loaded on Factor 1 serves to confirm that stereotypes are hard to overcome. Nonetheless, this result demonstrates that considerable progress can be made in overcoming incorrect stereotypes on deception in a manner that could be practically employed in court. Indeed, it may be that not overloading venire jurors with too much information was partly responsible for the significant and positive shift, in the sense that they were more likely to be accurate, in their perceptions of the efficacy of various behavioural indicia in detecting deception.

Lack of knowledge of the behavioural indicia that discriminates between liars and truth tellers is not the only obstacle that interferes with the ability of decision makers when assessing the credibility of witnesses. Considering the pervasive effects of cognitive heuristics on decision-making, it has been proposed that triers of fact may disregard education/training/judicial directions due to their tendency to adhere to their initial hypotheses (Porter & ten Brinke, 2009). Even worse, there is some evidence that triers of fact, as a group, often make judgments very early in a trial based on inaccurate perceptions of what constitutes indicia of truthfulness/deception (Porter & ten Brinke, 2009). This has been referred to as Dangerous Decisions Theory. Partly as a result of the well-established concept of cognitive dissonance (Festinger, 1957), triers of fact are resistant to changing their initial judgment – even in the face of clear evidence to the contrary.

These arguments raise a number of issues. First, there is the practical consideration as to whether or not the results of this study would apply to the final credibility judgments of decision makers. If Dangerous Decisions Theory is indeed correct, and the jury is still out on this, then the answer is clear: no. Yet, if this is the case, there is simply no point in trying to adduce expert evidence on detection of deception expert evidence since it won't make any difference because of the way trials are conducted in an adversarial system. By way of explanation, by the time the defence could adduce expert evidence on detection of deception, the triers of fact would have already made up their minds according to Dangerous Decisions Theory and thus there would be no point bothering. This is not only a counsel of despair but it also runs counter to the entire basis on which jury trials are conducted.

Second, the problem that changes in behavioural indicia indicative of deception are best measured by reference to the baseline performance for any particular person is well nigh intractable in the context of an actual trial. This study shows that venire juror's perceptions of behavioural indicia discriminating between liars and truth tellers can be very significantly improved, but whether this will result in an increased accuracy in detecting liars in the absence of knowledge of the usual behavioural presentation of witnesses is another thing. That said, a number of studies have shown that increased deception accuracy is certainly possible in the absence of significant knowledge of individuals' usual behavioural mannerisms (Porter, McCabe, Woodworth & Pearce, 2007; Porter, Juodis, ten Brinke, Klein & Peace, 2010). Further, some clues can be gleaned from this study that support the proposition that increased credibility assessments are likely to flow from increasing the knowledge of venire jurors, irrespective of lacking knowledge of a witnesses' usual presentation of behavioural indicia indicative of truth telling and deception. The fact that ambivalent responses, spontaneous corrections and short responses loaded on Factor 1 in Q2 but not in Q1 is of some note in this regard, since assessment of these items does not require significant, or even any knowledge, of individuals' baseline behaviour. Similarly, the fact there was a significant increase in the loading of questionnaire items dealing with responses with few details and responses that lack logical structure on Factor 1 in Q2 relative to Q1 may be interpreted in the same way. That is, observation of these behavioural responses does not require significant knowledge of individuals' baseline behaviour, if at all. Whichever way one looks at the results of this study, they clearly support the contention that adducing expert evidence *and* ensuring that judicial directions based on scientific fact as to what works and

what doesn't work in detecting deception are provided to jurors can only be a great advance on the current state of affairs.

RECOMMENDATIONS

Having regard to uncontested findings in the literature on detection of deception some of the problems confronting decision makers in assessing witness credibility are beyond doubt. The following are recommended: -

- Considering the vagaries of courtroom design and the utter impossibility of decision makers being able to accurately assess some behavioural indicia in the overwhelming majority of courtrooms, it is apparent that this needs to be remedied. Having CCTV footage of witnesses that can be displayed from both a wide angle and close-up perspective on large format television screens can do this. This should be mandatory in all courts so that the triers of fact have the capacity to closely observe a range of behavioural indicia of witnesses.
- Since misleading, if not demonstrably wrong guidance to judges, as may be found in the Queensland Supreme Court Equal Treatment Benchbook for example, is based on a misunderstanding of the science on which it is ostensibly founded such guidance should be rewritten in accordance with uncontested psychological fact.
- The issue of witness credibility is not confined to behavioural indicia that can be observed in the courtroom. It may be founded in numerous aspects of the investigatory phase of the legal process (Coyle & Field, 2013) and thus continuing dialogue between lawyers and psychologists, as McClellan J (2006), among others, has called for is essential. Continuing legal education on the interaction between psychology and law is clearly essential.

As far as the experimental results reported herein are concerned the following are recommended: -

- This study needs to be replicated, preferably with venire jurors who can view actual witnesses in mock trials, or, preferably, actual trials and make final credibility assessments after listening to expert evidence and receiving appropriate judicial direction. The latter option will require revisiting the long-standing reluctance to allow videorecording of trials in Australia and allow such recordings to be used for research purposes. There is no reason why appropriate safeguards

could not be put in place to ensure that such videorecording did not interfere with the administration of justice. Removing the barriers to post-polling of jurors for the purpose of research would enable questions about witness credibility that can only be inferred at present to be precisely explicated. Although some may regard this as a bridge by far, previously unthinkable steps (such as pre-polling of jurors to exclude those who are irretrievably biased – see *R v Patel* QSC (2013, unreported)) have found their way into the courts in Australia.

Notes

1. Professorial Associate, Centre for Law Governance and Public Policy, Law Faculty, Bond University; Adjunct Professor, Department of Psychology, Faculty of Sciences, University of Southern Queensland; Director, Safetysearch Forensic Consultants, Gold Coast, Queensland safetysearch2001@yahoo.com.au
2. See, generally: Field, D. *Queensland Evidence Law*, 2nd Edition, LexisNexis Butterworths Chatswood. At p. 159 it is noted that: - “It is sometimes difficult to determine what is and what is not a matter going solely to the witness’s credit.” Field cites *R v Hitchcock* (1847) 154 ER 38 where it was affirmed that an issue “is not collateral if, as part of the case for the cross-examining party, it could be proved.”
3. These operate as exceptions to what is generally termed the “collateral evidence rule” both at common law and under statute; see *Cross on Evidence* (note 6) at 17595. See also *Evidence Acts 1995* (Cth & NSW), s 106.
4. See, generally: *CSR v Della Maddalena* (2006) 224 ALR 1; *Fox v Percy* (2003) 214 CLR 118. In other jurisdictions such evidence is clearly admissible. See, generally: Ogloff, J.R.P. (1996). The admissibility of expert testimony regarding malingering and deception. *Behavioral Sciences and Law*, 8, 27-43.
5. S.21A of the *Evidence Act 1977(Qld)* states that a witness’s cultural background may be a “relevant matter” which would be likely to cause a witness to be disadvantaged-but Anglo-Australian’s are not afforded such consideration.
6. Vrij (2005) takes issue with the use of some of these techniques, SVA and CBCA, as accuracy rates may be as low as 75% – but even this is much higher than is typically observed with other techniques.

References

- Alexander v The Queen* (1981) 145 CLR 395.
- Aytugrul v The Queen* [2012] HCA 15, (18 April 2012).
- Bagaric, M. & McConvill, J. (2005). Evidence law: when you are in a hole stop digging. *The Original Law Review*, 1 (1), 10-25.
- Blankenship, M. B., Luginbuhl, J., Cullen, F. T., & Redick, W. (1997). Jurors comprehension of sentencing instructions: A test of the death penalty process in Tennessee. *Justice Quarterly*, 14 (2), 325-351.
- Bond Jr., C. F., & DePaulo, B. M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review*, 10, 214-234.
- Bull, R. (2004). Training to detect deception from behavioural cues: attempts and problems. In: Granhag, P.A. and Strömwall, L.A. (Eds.). *The Detection of Deception in Forensic Contexts* (pp. 251-268). Cambridge, Cambridge University Press.

- Connor, X. Trial by jury –can it survive? *Law Institute Journal*, 61, 818-819.
- Coyle, I.R., Field, D. & Miller, G. (2008). The Blindness of the eye-witness. *The Australian Law Journal*, 82, 471-498.
- Coyle, I.R., Field, D. Wilson, P. Cuthbert, C. & Miller, G. (2009). Out of the mouths of babes: the case for an increased use of expert evidence in rebuttal of sexual abuse allegations by child witnesses. *Criminal Law Journal*, 33, 139-164.
- Coyle, I. R. & Field, D. (2013). Psychology from the bench. *Psychiatry, Psychology and Law*, 20 (1), 91-104.
- Coyle, I. R. & Halon, R.L. (2013). Humpty Dumpty and Risk Assessment: A Reply to Slobogin. In: Keyzer, P. (Ed.), *Preventive Detention: Asking the Fundamental Questions*. Intersentia, Cambridge, In Press.
- Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences & the Law*, 7 (2), 215-225.
- 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.
- Domican v R* (1992) 66 ALJR 285; 173 CLR 555.
- English, P. W., & Sales, B. D. (1997). A ceiling or consistency effect for the comprehension of jury instructions. *Psychology, Public Policy, and Law*, 3 (2/3), 381-401.
- Ellison, L., & Munro, V. E. (2009). Turning mirrors into windows? Assessing the impact of (mock) juror education in rape trials. *British Journal of Criminology*, 49, 363-383.
- Equal Treatment Benchbook Supreme Court of Queensland (2006). Brisbane, Supreme Court of Queensland Library.
- Farrell v The Queen* (1998). CLR 286.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford C.A.
- Folkes v Chadd* (1782). 3 Doug KB 157; 99 ER 589.
- Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detecting training. *Journal of Applied Communication Research*, 31 (3), 58-75.
- Fritzon, K. (2005). Assessing the reliability of interviews with vulnerable witnesses. In: Alison, L.J. (Ed.) *Offender Profiling and Crime Analysis: A Forensic Psychology Case Book* (pp. 278-295). Devon, William Publishing.
- Gudjonsson, G. (1992). *The Psychology of Interrogations, Confessions and Testimony*. New York, John Wiley & Sons.
- Hartwig, M., & Bond Jr., C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgements. *Psychological Bulletin*, 137 (4), 643-659.
- Inbau, F. E., Reid, J. E., & Buckley, J. P. (1986). *Criminal Interrogation and Confessions* (3rd ed. ed.). Baltimore, MD: Williams & Wilkins.
- Kassin, S. M., & Fong, C. T. (1999). "I'm Innocent!": Effects of training on judgements of truth and deception in the interrogation room. *Law and Human Behavior*, 23 (5), 499-516.
- Köhnken, G. (2004). Statement Validity Analysis and the 'detection of the truth'. In: Granhag, P.A. and Strömwall, L.A. (Eds.) (2004). *The Detection of Deception in Forensic Contexts*, pp. 41-63. Cambridge, Cambridge University Press.
- Longman v The Queen* (1989) CLR 79; 43 A Crim R 463.
- Luginbuhl, J. (1992). Comprehension of judges' instructions in the penalty phase of a capital trial. *Law and Human Behavior*, 16 (2), 203-218.
- McClellan, P. (2006). Who is telling the truth? Psychology, commonsense and the law. Local Courts of New South Wales Annual Conference.
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89 (1), 147-149.
- Mann, S., & Vrij, A. (2006). Police officers' judgements of veracity, tensions, cognitive load

and attempted behavioural control in real-life police interviews. *Psychology, Crime & Law*, 12 (3), 307-319.

Memon, A., Fraser, J., Colwell, K., Odnot, G. & Mastroberardino, S. (2010). Distinguishing truthful from invented accounts using reality monitoring. *Legal and Criminological Psychology*, 15(2), 177-194.

Murphy v The Queen (1989). 167 CLR 194.

Neal, T.M.S. & Brodsky, S.L. (2008). Expert witness credibility as a function of eye contact behavior and gender. *Criminal Justice and Behavior*, 35 (12), 1515-1526.

Ogloff, J. R. P., & Rose, V. G. (2003). The comprehension of judicial instructions. In N. Brewer, & K. D. Williams (Eds.), *Psychology and Law: An empirical perspective* (pp. 407-444). New York: The Guildford Press.

Ogloff, J. R. P., Nadjovski-Terziovski, E., Spivak, B., & Clough, J. (2011, August). *Juror comprehension of legal instructions in a sexual assault trial*. Paper presented at the Australian Psychological Society Forensic Psychology National Conference, Noosa.

O'Sullivan, M. & Ekman, P. (2004). The wizards of deception detection. In: Granhag, P.A. and Strömwall, L.A. (Eds.), *The Detection of Deception in Forensic Contexts*, pp. 269-286. Cambridge, Cambridge University Press.

Porter, S. & ten Brinke, L. (2009) Dangerous decisions: a theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology*, 14, 119-134.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high stakes deception? *Legal and Criminological Psychology*, 15, 57-75.

Porter, S., Juodis, M., ten Brinke, L. M., Klein, R., & Wilson, K. (2010). Evaluation of the effectiveness of a brief deception detection training program. *The Journal of Forensic Psychiatry and Psychology*, 21 (1), 66-76.

Porter, S., McCabe, S., Woodworth, M., & Peace, K. A. (2007). 'Genius is 1% inspiration and 99% perspiration'...or is it? An investigation of the impact of motivation and feedback on deception detection. *Legal and Criminological Psychology*, 12, 297-309.

Raskin, D.C. & Esplin, P.W. (1992). Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. *Behavioral Assessment*, 13, 265-291.

Rose, V. G., & Ogloff, J. R. P. (2001). Evaluating the comprehensibility of jury instructions: A method and example. *Law and Human Behavior*, 25 (4), 409-431.

Severance, L. J., & Loftus, E. F. (1982). Improving the ability of jurors to comprehend and apply criminal jury instructions. *Law & Society Review*, 17 (1), 153-198.

Spivak, B, Ogloff, J. R. P., & Clough, J. (2011, August). *A psycho-legal analysis of jury instructions*. Paper presented at the Australian Psychological Society Forensic Psychology National Conference, Noosa.

R v BDX (2009). VSCA 28

R v Marquard (1993). 4 SCR 223; (1993) 108 DLR (4th) 47; 85 CCC (3d) 193.

R v Martens (2007). QCA 137.

R v Turner (1975). 1 Q.B. 835.

Redlich, A.D. & Meissner, C. (2009). Techniques and Controversies in the Interrogation of Suspects: The Artful Practice versus the Scientific Study In :J. L. Skeem, K. Douglas, & S. Lilienfeld (Eds.), *Psychological science in the courtroom: Controversies and consensus*. New York, The Guildford Press, pp. 124-149.

Sporer, S.L. (2004). Reality Monitoring and the Detection of Deception. In: Granhag and Strömwall (Eds.), *The Detection of Deception in Forensic Contexts*, pp. 64-102. Cambridge, Cambridge University Press.

Strömwall, L.A., Granhag, P.A. & Hartwig, M. (2004). Practitioners' beliefs about deception. In Granhag, P.A. & Strömwall, L.A. (Eds.). *The Detection of Deception in Forensic Contexts*, pp. 229-250. Cambridge, Cambridge University Press.

The Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology*, 37(1), 60-74. Note: this team consists of 90 researchers.

Thomas, C. (2010). Are juries fair? UK Ministry of Justice Research Series, 1/10, 1-65.

Tversky, A & Khaneman, D. (1974). Judgment under uncertainty: heuristics and bias, *Science*, 185, 1124-1131.

Tversky, A & Khaneman, D. (1981). The framing of decisions and the psychology of choice, *Science*, 211, 453-458.

Vilar, G., Arciuli, J. & Paterson, H. (2013). Vocal pitch production during lying: beliefs about deception matter. *Psychiatry, Psychology and Law*, 20 (1), 123-132.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and its implications for professional practice*. Chichester, England: John Wiley & Sons.

Vrij, A. (2005). Criteria based content analysis: a qualitative review of the first 37 studies. *Psychology Public Policy and Law*, 11, 3-41.

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2nd ed.). Chichester, John Wiley & Sons.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2006). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, 30, 8-41.

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). 'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, 16 (4), 327-348.

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.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, 11, 89-121.

Warmelink, L., Vrij, A., Mann, S. Leal, S & Poletiek. (2013). The effects of unexpected questions on detecting familiar and unfamiliar lies. *Psychiatry, Psychology and Law*, 20 (1), 29-35.

Appendix 1: Questionnaire

This questionnaire was based on one developed as part of a B.A. (Hons.) thesis by Melissa Toh at Macquarie University where the author was a co-supervisor.

Instructions:

You are asked to indicate what factors you think would enable you to determine whether someone is truthful if they were giving evidence in a court case. Circle the number that best describes how you feel. **If you make a mistake, put a cross through the circle and circle the appropriate number.** There are no correct answers; we are merely seeking your opinion

1. Confidence

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

2. Pauses

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

3. High Pitched Voice

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

4. Fidgeting

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

5. Restless Behaviour

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

6. Implausible Responses

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

7. Postural Changes

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

8. Speech disturbances/fillers i.e. “ah’s” and “umm’s”

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

9. Anxious behaviour.

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

10. Self-manipulators i.e. hand movements that involve relieving a bodily need like scratching an itch.

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

11. Short responses

Extremely Frequently 6	Very Frequently 5	Somewhat Frequently 4	Somewhat Infrequently 3	Very Infrequently 2	Extremely Infrequently 1
------------------------------	-------------------------	-----------------------------	-------------------------------	---------------------------	--------------------------------

12. Mumbling

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

13. Covering the mouth

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

14. Body/hand/finger/leg/foot movements

Extremely Frequently 6	Very Frequently 5	Somewhat Frequently 4	Somewhat Infrequently 3	Very Infrequently 2	Extremely Infrequently 1
------------------------------	-------------------------	-----------------------------	-------------------------------	---------------------------	--------------------------------

15. Stuttering

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

16. Interrupting the questioner

Extremely Infrequently 1	Very Infrequently 2	Somewhat Infrequently 3	Somewhat Frequently 4	Very Frequently 5	Extremely Frequently 6
--------------------------------	---------------------------	-------------------------------	-----------------------------	-------------------------	------------------------------

17. Responses that lack logical structure

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

18. Pressed lips

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

19. Repetition of certain words and phrases

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

20. Blinking

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

21. Responses that contain very few details

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

22. Tenseness

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

23. Self-grooming

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

24. Nervous behaviour

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

25. Ambivalent responses

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

26. Gaze aversion

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

27. Crossed arms

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

28. Illustrators i.e. hand movements that accompany and illustrate the content of the speaker's responses

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

29. Delayed responses

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

30. Spontaneous corrections

Extremely Frequently	Very Frequently	Somewhat Frequently	Somewhat Infrequently	Very Infrequently	Extremely Infrequently
6	5	4	3	2	1

31. Responses that contain a lot of detail

Extremely Infrequently	Very Infrequently	Somewhat Infrequently	Somewhat Frequently	Very Frequently	Extremely Frequently
1	2	3	4	5	6

Appendix 2: Judicial Directions

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

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

- Liars tend to speak in a higher pitch relative to their normal speaking voice when they are telling the truth.
- Liars make fewer movements with their hands, arms and fingers.
- Liars use fewer illustrators (hand and arm movements to demonstrate and illustrate points they are making).
- Liars take longer pauses.
- Liars' answers are less plausible.
- Liars' stories contain fewer details.
- Liars give more indirect answers.
- Liars' answers contain less temporal, perceptual and spatial information.

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

