

Are police superior to non-police at recognising reliable evidence? are they more likely to use alternative evidence?

Abstract

Sixty police and sixty non-police mock jurors read a fictitious transcript describing how an attempt at blowing up the Channel Tunnel was undertaken and how a bystander was injured by the fleeing culprit. The expert forensic and eyewitness (i.e., psychological) evidence presented in the transcript was either reliable or unreliable (such that there were four possible combinations). Two control transcripts contained expert forensic evidence only (reliable or unreliable). Police officers from Bridgend and Cardiff Central Police Forces and non-police solicited from the universities of Cardiff, South Wales and Sussex acted as 'mock' jurors and rated defendants for guilt and extent of guilt whilst considering evidence most influential in their decision-making. The aims of the current study are therefore to test whether police have a more informed understanding of the reliability and validity of both forensic and eyewitness evidence than non-police mock jurors and, that they are more sceptical of expert eyewitness testimony. The statistical analyses, MANOVA and Jonckheere, showed that police and non-police mock jurors have difficulty differentiating between reliable and unreliable evidence for the count of hitting the bystander. For the count of 'blowing up the tunnel', non-police mock jurors appear to be utilising evidence consistent with its reliability; however, MANOVA and Jonckheere analyses failed to demonstrate this statistically. Findings are accounted for by Hastie et al.¹ 'Story Model' of how jurors interpret information and simplify meaning to help inform their verdicts.

Keywords: evidence reliability, police, decision-making, story model

Volume 5 Issue 4 - 2017

Sandie Taylor

School of Psychology and Therapeutic Studies, University of South Wales, UK

Correspondence: Sandie Taylor, School of Psychology and Therapeutic Studies, University of South Wales, Pontypridd, Wales, CF37 1DL, UK, Email Sandra.taylor@southwales.ac.uk

Received: October 02, 2017 | **Published:** November 01, 2017

Review of literature

There are three main aims underlying a criminal investigation.² Monckton-Smith et al.² claim this involves, "to ascertain if an offence has been committed....to identify who committed the crime....to gather admissible evidence" (p.2). Obtaining evidence that is admissible in court is the task of the police and with this comes a degree of discretion exercised throughout their criminal investigation: reporting and recording crime; decreeing caution; investigating crime; collating evidence; and deciding whether or not to arrest a suspect. Once police have recorded an event as a crime and have identified a suspect, they can then decide on an appropriate deployment option. According to Williams³ and Williams⁴ these options involve decisions of arrest or summons, charge, bail and legal support including advice and caution. Police power operates within the boundaries of written law and the Police and Criminal Evidence Act (PACE) of 1984. The PACE document ensures that police are governed by rules and procedures predicated their investigation process.⁵ Stipulations of legal procedure encapsulated in PACE are there to protect potentially innocent suspects such that police are accountable for their actions during investigation. If their actions fall outside of police procedure then an inquiry is held to establish why this occurred and the consequences for all parties concerned are evaluated and dealt with accordingly. Even with rules and procedures in place, police exercise discretion in determining who is judged to be an offender and put on trial in a courtroom.⁶

In this way police act as an interface between the criminal activity and the implementation of justice. They are the first law enforcing body to be at the scene of a crime and to be in contact with offenders

and those they consider being the suspect of a crime. As police come into contact with offenders and those they consider the suspect of a crime, it is for this reason that it is important at this early stage of their investigation to follow the procedural stipulations described in PACE 1984. The UK and US follow the adversarial legal system differing from the inquisitorial legal system in a number of ways. One difference is that the judge in a criminal trial leaves the jury to deliberate a defendant's guilt after evidence has been presented by both the defence and prosecution lawyers based on the evidential facts gathered by the police. Prior to court proceedings, police in the UK present evidence to the Crown Prosecution Service (CPS) to state their case. The CPS decide whether there is at least a 50 percent chance of winning a case in court based on the police evidence presented. A less than 50 percent chance results in the case being thrown out and legal proceedings dropped until substantive incriminating evidence is obtained. Under the inquisitorial legal system adopted in many European countries, police have more power to investigate and review their findings and decide whether the suspect committed the crime he/she is accused of. Once police make criminal charges, it is the court who then audits and determines the reliability of evidence ultimately influencing the sentence passed - if required.^{7,8}

In both legal systems, police have great power in the determination of case direction. Gathering evidence, however, is less straightforward than it first appears to be. First, police have to be convinced that a criminal event had occurred and to record and recode it accordingly. Second, police require information in support of the suspect being involved in the determination of event outcome. This helps to determine the suspect's level of responsibility at each stage of the criminal event. Hence, police have to code an event as a crime and

obtain evidence of an individual's involvement in the event under question – the acquisition of constitutive facts. To be in a position to establish this, part of the police investigation work must surely involve the separation of relevant and irrelevant facts and how relevant facts inter-relate. Police application to following specific clues is important to the gathering of evidence. The way these clues are investigated is also subject to police decision making.

Questions such as who would be appropriate witnesses and which expert witnesses would best help highlight the significance of clues presented in court. Moreover, police are responsible for presenting uncontaminated evidence (physical and psychological) given they are first at the scene and investigate relevant clues. This supports the assumption that police have a superior understanding of evidence reliability and validity when compared with members of the jury. According to Pennington et al.,⁹ jurors should be able to apply seven conceptually discrete elements of information processing which they captured in their generic model of jury decision making. These elements have been incorporated into their Story Model of 2000 and include:

- i. Assimilation of information presented at trial,
- Continuation of the representation of the facts in a causal-sequence,
- ii. Evaluation of the reliability of evidence presented,
 - iii. Ability to separate admissible from inadmissible evidence,
 - iv. Understanding legal categories,
 - v. Ability to put evidence into legal categories,
 - vi. Application of the legal standard of proof (i.e. beyond reasonable doubt).

Some support for the application of these elements has been found by Ellsworth¹⁰ who showed that jurors concentrate on following the trial, time and sequence framing of the facts and assessing reliability and allocation to legal categories. Ellsworth, however, points out that there are stages jurors omit to do. For example, verdict decision making is based on personal perceptions of crime seriousness adopting a 'dumbed' down appraisal of event occurrence. This leads to the adoption of a story that best depicts the occurrence of events and a verdict that offers an optimal match to the criminal event as presented in court. In this study the stages of story composition and evidence evaluation are important to how jurors (and potential jurors/mock jurors) use evidential information. Story construction by jurors can be considered as a similar process to that observed during the investigative process of constitutive fact gathering by police. The jurors' role, however, is a passive one involving the assimilation of information to construct a story unlike the role of actively collating constitutive facts used in the initial development of a 'story' as seen in the police investigative process. The difference in role of jurors and police suggests that there might be differences in how information is evaluated. In addition to this, police present their 'story' to the CPS before any court trial proceedings. Moreover, the presentation of case material to the CPS requires there to be reliable and valid constitutive and evidential facts to guarantee at least a 50 percent chance of conviction.

We might ask how this informs us of the role played by police in understanding the reliability and quality of evidence presented. It might be taken to suggest that we expect police to have a deeper understanding of evidence than the general public. It is open to debate as to how such expertise is developed. This may develop via contact

with experts such as forensic scientists or criminological psychologists or it may be related to personality dispositions. It is well established that the general public consider the police are able to provide an accurate account of factual information than they are themselves.^{11,12} In fact research findings support this supposition as members of the police force have been shown to have superior recall when compared with the general public. Clifford et al.,¹³ for example, found that police have superior facial recognition when compared with performance of the general public. Clifford and Richards suggested this superiority can be explained by the processing skills that police officers are trained to develop. Yuille,¹⁴ however, considered that this difference may be explained by the fact that police tend to look for criminal activity whereas civilians have quite different goals. There is evidence that professional profilers produce more accurate profile constructions of sex offenders when compared with detectives, psychologists and the general public. This finding, however, does not generalise to cases of homicide.¹⁵

A meta-analysis demonstrated that profilers are more accurate per se than non-profilers.¹⁶ The same data was analysed in 2010 by Snook B, et al.¹⁷ who found the opposite was true. According to¹⁸ it depends on how expertise is defined. "If we exclusively define an expert profiler to include those who have clear training and expertise in the FBI style of profiling then the evidence that profiling experts are better than non-profilers at producing accurate profiles is stronger"¹⁸ He further added that there is little difference in profile performance accuracy between police officers who are perceived as experts but have not received training and non-profilers. This is important as evidential information is taken into consideration when constructing a profile which ultimately is derived from decisions concerning the reliability and validity of criteria used for fact inclusion. Alison et al.¹⁹ considered police attitudes towards the accuracy and reliability of the facts provided in an FBI profile case. The police viewed the facts encompassed in the profile positively. Police positivity, however, extended to a fictitious profile riddled with inaccurate information. When it comes to detecting deception, however, it has been found that police officers are no better than college students even though they are more confident about such judgements.²⁰ Surprisingly, prisoners have been found to be more accurate in their beliefs about deception than members of the police force or students.²¹ This might be taken to suggest that police training does not improve deception detection.²² Research by Bond,²³ however, suggests that there are police officers who perform exceptionally well at detecting when suspects are lying and, called them detection wizards. These police detection wizards were given a series of lie detection tests but only those who performed at an accuracy of 80 percent or higher were subjected to further tests. They wore eye-tracking sensors which highlighted the non-verbal behaviours they focused on to help them decide when a suspect was lying. Bond argues that these findings can enlighten police training in lie detection.

Scholars such as Skolnick²⁴ consider that police training helps maintain a subculture of specific cognitive and behavioural skills: one which looks for confirmatory rather than disconfirmatory evidence of their suspicions. If this is the case then their consideration of the evidence no doubt influences the nature of investigation, especially within an adversarial context. Some consider that the role of police in investigations is by no means impartial since it reflects their prosecutorial role.²⁵ Skolnick²⁴ and Balch²⁶ both consider that police officers tend to have similar personality and cognitive traits which enable them to form close working units. This also suggests they are likely to consider their suspect is guilty and tend to search for

evidence which confirms this belief. This was found to be particularly so during the gathering of evidence by police interviewers who had poor interviewing skills.²⁷ Walsh et al.²⁷ found that interviewers with poor interviewing skills assumed their suspect to be guilty prior to any questioning. This has a biasing effect in the way the interview is composed and the type of questions asked – known as confirmation bias.²⁸ This encourages “questioning styles primed to confirm hypotheses of a suspect’s guilt...at the expense of any other relevant and contradictory information.”²⁹ Inciardi³⁰ found that police subculture includes, “protective, supportive, and shared attitudes, values, understandings and views of the world” (p.227). Nelson et al.³¹ claim the police represent organisational assumptions – in other words they have beliefs which direct their behaviours and how to perceive the actions of others and to cognitise them. Police are trained to be suspicious of others and treat citizens as symbolic assailants.^{32,33} Berg³⁴ claimed that the police personality develops from characteristics and behaviours which typify a police stereotype. Twersky-Glasner³⁵ posed the question of how this police personality develops: a predisposition or organisational socialisation? Although this is beyond the scope of this paper, it does allude to there being a police subculture perpetuated through new police cadets joining the police force and their training. And, this is important because it influences how police investigate.

If Skolnick’s views on police subculture are correct, then they may be taken as a serious criticism of the way that police operationalise investigations. Interestingly, Moston et al.,³⁶ when analysing 1067 cases, found that police interviewers based their beliefs and attitudes of a suspect’s guilt or innocence on the strength of evidence and criminal history. In cases where the evidence was weak, police interviewers were sure of a suspect’s guilt in 31 percent of the time. In cases where the evidence was moderate, this rose to 74 percent and, when evidence was strong, to 99 percent. Where there were no previous convictions, police interviewers were sure of a suspect’s guilt in 69 percent of cases. When, however, suspects had arrests for previous convictions this rose to 76 percent. This might be taken to suggest that a police officer’s belief of a suspect’s guilt is largely because of the evidence uncovered. Moreover, successful outcomes rest largely on how police understand and interpret the validity and reliability of evidence collated and its *res gestae* (defined as “the acts, facts, circumstances, statements, or occurrences that form the environment of a main act or event and especially of a crime”³⁷ and are admitted to the court as evidence). It is interesting that under both inquisitorial and adversarial criminal law, police have great power in influencing the direction taken by the court and Crown Prosecution Service (CPS) respectively.^{1,2,8} One way that expert testimony and its impact on how jurors process evidence and arrive at a verdict, is through the use of mock jury studies.³⁸ Interestingly, courts make use of expert testimony where forensic psychologists are utilised to provide the court and jury with an understanding of the fallibility of human memory when witnessing criminal events. The sophistication and complexity of evidence presented means that jurors often have to process and interpret difficult information; often presented in highly technical language. Richardson³⁹ suggested that jurors interpret what is being presented to them through their own life experiences. Jury members regularly take into account non-evidential evidence (i.e. extra-legal defendant characteristics) of the defendant. Despite this, Visser⁴⁰; Otto et al.⁴¹ found that evidence is three times more important in determining jury deliberation than any extra-legal factor.

An important question addressed by Kovera and McAuliffe⁴² is to what extent are potential jurors able to detect methodological flaws in expert evidence? In particular, they considered whether judges

and jurors were sensitive to variations in the methodological quality of expert evidence. Their findings showed how potential jurors had difficulty in identifying reliable expert evidence and differentiating it from that which was unreliable. Although Kovera and McAuliffe also found that some judges had problems with separating reliable from unreliable evidence, they were on the whole more sensitive to internal validity issues such as the absence of a control group. Judges also believed that inaccurate science could be eliminated by the cross-examination process. Kovera and McAuliffe, however, pointed out that many judges, despite managing expert evidence, were, “ill equipped to identify flawed research designs in social science evidence”.³⁹ Furthermore, they found that judges had a low opinion of psychology *per se* and often excluded psychological research as constituting reliable expert evidence. Potential jurors find it difficult to understand the implication of control groups and why cross-examination will iron-out any internal reliability and validity issues. These findings favour the supposition that police as opposed to the layperson (i.e., potential/mock juror) will have more of an understanding and knowledge of evidence reliability. This is probably a result of their role in obtaining evidence and courtroom experience.

Leippe et al.,⁴⁴ considered the impact of expert eyewitness testimony on jurors’ beliefs and judgements about reliability of an eyewitness’ statement. Interestingly, their findings suggest that there is little gain in trying to educate jurors about the pitfalls of eyewitness reliability and, improve sensitivity to the quality of the evidence provided. The provision of expert eyewitness evidence tends to reduce conviction rates and increase acquittals. Loftus⁴⁵ examined juror scepticism and sensitivity to expert eyewitness evidence. A transcript containing numerous factors impacting on eyewitness identification was presented to mock jurors. Exposure to expert testimony increased mock juror scepticism which was found to be the case some 29 years later by Leippe et al.,⁴⁴ Mock jurors who had been exposed to expert eyewitness testimony convicted 39 percent of the time compared to 58 percent for those who had not received such evidence. Little research has been conducted to examine the extent to which police are open to expert eyewitness testimony about the reliability of eyewitness evidence.

Problem of the study

The following hypotheses were tested:

- a. H¹ states that police mock jurors will be better at differentiating between reliable and unreliable expert evidence than non-police mock jurors.
- b. H² states that police mock jurors will use reliable expert evidence to inform their verdict more than non-police mock jurors.
- c. H³ states that non-police mock jurors are more open to expert eyewitness evidence than police mock jurors.

Method

Participants

Sixty participants solicited from the Bridgend and Cardiff Central Police Forces comprised the police group. Sixty participants ranging from students, lecturers and other staff attending Cardiff University, University of South Wales and the University of Sussex comprised the non-police group. The age range for police and non-police participants was 21-50. All participants were considered as mock jurors.

Design

An unrelated measures design was used: two control and four experimental conditions. A captive police audience was allocated to conditions blindly. A contact senior police officer provided permission and actively distributed transcripts to the police population in a classroom set up. Transcripts were not labelled in any way therefore assuming a similar appearance. Transcript distributed to the non-police sample was undertaken using a snowball effect; whereby experimenter contacts within the universities allocated mock jurors to conditions on an *ad hoc* basis. In each condition there were ten police and ten non-police mock jurors. Given that in the UK courts it is the judge who reminds jurors of the fallibility of eyewitness memory and not an expert psychologist (as is often practised in the US), two control conditions were devised. Only expert forensic evidence and the testimonies from eyewitnesses were presented in the control conditions – one control condition having reliable forensic expert evidence (Cg) while the other having unreliable forensic expert evidence (Cb). In addition to reliable and unreliable forensic expert evidence, experimental conditions included either reliable or unreliable psychological expert evidence. Hence, there were four possible experimental combinations: reliable forensic and reliable psychological expert evidence (Exp_{gg}); reliable forensic and unreliable psychological expert evidence (Exp_{gb}); unreliable forensic and reliable psychological expert evidence (Exp_{bg}); unreliable forensic and unreliable psychological expert evidence (Exp_{bb}). The reason for regarding conditions including psychological expert evidence as experimental relates to the fact that evidence of this nature atypically appears in British courts.

Materials

Mock jurors were presented with a written transcript of a fictitious court case outlining how terrorist activity damaged part of the British Channel Tunnel. Although the plot was fictitious, the script format adhered to a realistic trial procedure where direct and cross-examination of witnesses, courtroom dialogue, expert forensic and medical evidence, police evidence and the judge’s instruction to jurors followed as near possible that used in court. Reliability of forensic evidence was controlled by using forensic tests well known to experts working in the field. Tests known to be less reliable were presented in the unreliable forensic evidence transcript, whereas tests renowned for their reliability were presented in the reliable forensic evidence transcript. In the case of psychological evidence, psychological research on eyewitness reliability and face recognition were combined to form a body of either reliable or unreliable evidence.

The summary of the case is as follows:

- i. The bystander receives a bump to the head, which aggravates a past medical problem. Damage is done to part of the British Channel Tunnel using explosives. Police find the defendant hiding in the vicinity of the crime scene who is later identified by witnesses. Forensic scientists conclude that the explosive material used to blow the tunnel is the same as that found on the defendant’s hands.

Procedure

Mock jurors were provided with instructions to read the transcript and behave as criminal trial jurors. The questions, attached at the end of the transcript, were to be answered independent of any other person. Using a rating scale, mock jurors had to decide on the extent of guilt they believed the defendant was for the two counts charged.

In this case the scale was from ‘1’ not guilty at all, ‘3’ uncertain to ‘5’ highly guilty. For the two counts which the defendant was charged, a guilty or not guilty verdict was also decided where the answer they believed to be true was circled. A further decision was made regarding the evidence contributing most to their final verdict: forensic; psychological; both or neither.

Ethical issues

This study was passed by the Ethics Committee of the School of Social Sciences at the University of Cardiff.

Results

There were two criminal counts for which the defendant was charged and presented to the court: ‘hitting the bystander’ and ‘blowing up the tunnel’. These will be considered in turn.

Count of ‘hitting the bystander’

Descriptive analysis: The mode scores were calculated for all conditions. Refer to Figure 1 for the police mode scores and Figure 2 for the non-police mode scores on count of ‘hitting the bystander’ across conditions. Refer to Figure 3 for the police and non-police mode scores collapsed across conditions for count of ‘hitting the bystander.’ Figures 1 & 2 regarding the count of ‘hitting the bystander’ are very informative; providing a distribution pattern of mode scores across conditions of evidence reliability. In the case of police mock jurors, it appears that although uncertain as to the extent of the defendant’s guilt in the Cg condition they opted for a guilty verdict and claimed to have relied on forensic and eyewitness testimonies (Figure 1). Given that the Cg condition is a control in that only reliable forensic expert evidence is presented in a comprehensive way their decision to be reliant on both appears unfounded (there is no expert psychologist). A guilty verdict, however, is logical given the reliable forensic evidence presented. The only other condition where the majority of police gave a guilty verdict is the Exp_{bg} condition. Here, both forensic (albeit unreliable) and psychological evidence is presented which police claimed both informed their decision. Although they had problems deciding the extent of guilt they were able to deliberate a guilty verdict. The Cb, Exp_{gg}, Exp_{gb} and Exp_{bb} conditions all received ‘not guilty’ verdicts. The ‘extent of guilt’ ratings go in the direction of the verdict with the exception of Exp_{bb} where despite a ‘probably guilty’ rating they deliberated a not guilty verdict.

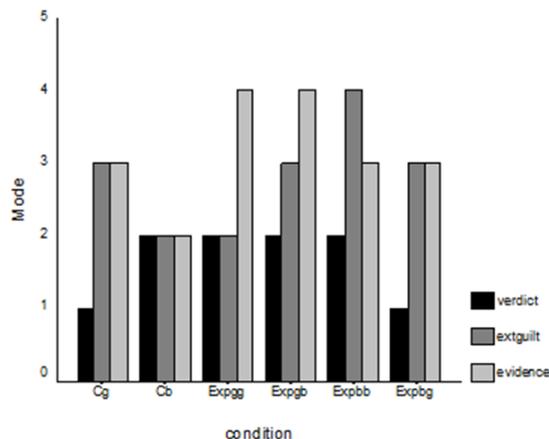


Figure 1 Police mock jurors’ mode scores for verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)) on count of ‘hitting the bystander’ across the conditions of evidence reliability.

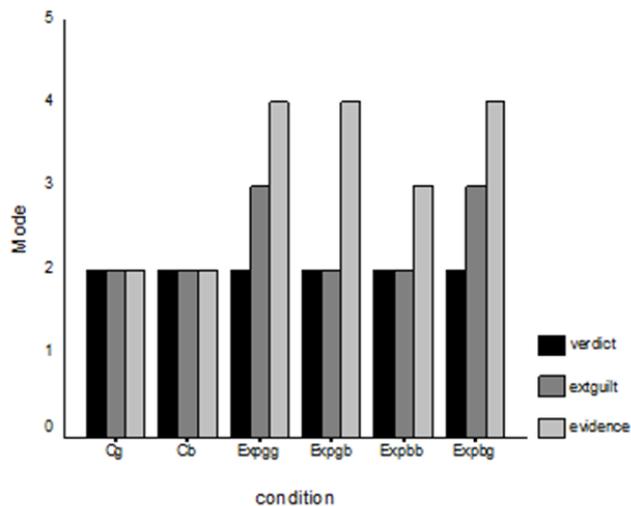


Figure 2 Non-police mock jurors' mode scores for verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)) on count of 'hitting the bystander' across the conditions of evidence reliability.

The non-police data are similar barring the fact that in all conditions a 'not guilty' verdict was deliberated (Figure 2). It appears that for the count of 'hitting the bystander' both police and non-police are unable to clearly differentiate between reliable and unreliable evidence. In fact, the evidence that both police and non-police claimed to have helped inform their verdict does not map onto the experimental conditions of varying combinations of reliable and unreliable forensic and psychological evidence. In Figure 3 extent of guilt, verdict and evidence data for police and non-police were collapsed across conditions. Both police and non-police have a mode score of 2 meaning 'not guilty.' Police rate 'extent of guilt' as uncertain whereas non-police 'probably not guilty' – both ratings are fitting for a 'not guilty' verdict. Whilst non-police seem to be relying on expert psychological evidence, police are relying on neither psychological nor forensic expert evidence. Extrapolating beyond this one can speculate that perhaps police rely on their 'gut instinct' or on their own 'story telling' skills of defining constitutive facts based on their years of police experience.

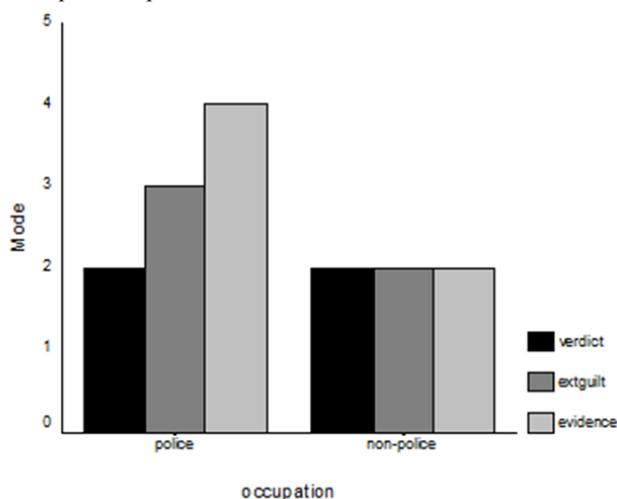


Figure 3 Police and non-police mock jurors' mode scores collapsed across conditions for the count of 'hitting the bystander'. Verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)).

Inferential analysis: A MANOVA analysis was conducted on these data. Occupation yielded a significant difference between police and non-police for extent of guilt ($F(1,108)=8.09, p<0.01$) and evidence ($F(1,108)=11.19, p<0.001$). Verdict, however, failed to reach significance ($F(1,108)=3.24, p>0.05$). The difference in mean scores verifies this: both police and non-police indicate a guilty verdict. For extent of guilt, police tend to be 'uncertain' and non-police tend to choose 'probably not guilty.' In the case of evidence, police tend to use both types of expert evidence and non-police use expert psychological evidence. Across conditions verdict was only just significant ($F(5,108)=2.31, p<0.05$). This is primarily because of the only significant Scheffepost hoc comparison: Cb and Expbg ($p<0.05$). Both extent of guilt and evidence were not significant ($F(5,108)=2.05, p>0.05$ and $F(5,108)=2.11, p>0.05$ respectively). The only significant interaction between occupation and condition was for verdict ($F(5,108)=2.75, p<0.05$). As one of the aims of this study was to ascertain whether there are any differences in verdict between conditions of varying expert forensic and psychological evidence reliability, it was decided to look for a trend using 'extent of guilt' scores. Under conditions of unreliable forensic and psychological expert evidence, a preponderance of ratings by police and non-police should lie in the lower end of the scale (below 3). For the more reliable evidence conditions a preponderance of ratings should lie in the higher end of the scale (above 3). Thus an increasing progression can be postulated as below:

- The data met the requirements of the Jonckheere Trend Test. For the count of 'hitting the bystander' a significant trend for both police and non-police was found ($S=-390, cv=256, p<0.01$ and $S=-334, cv=256, p<0.05$) respectively. Considering the raw data by looking at the rating most frequent under each condition, the trend fails to fit the predicted order above. In fact for the police the ratings for each condition are 2,4,3+4,3,3 and 2. The fours are in the incorrect position. A similar finding for the non-police data: 2,2,3,2,2 and 3. The most frequent extent of guilt ratings do not fit with the conditions of increasing reliability nor the direction of the predicted trend (despite being significant). Taking all analysis into consideration it appears that for the count of 'hitting the bystander,' mock jurors have difficulty identifying reliable from unreliable forensic and psychological expert evidence.

Count of 'blowing up the tunnel'

Descriptive analysis: The mode scores were calculated for all conditions. Refer to Figure 4 for the police mock jurors' mode scores and Figure 5 for the non-police mock jurors' mode scores on count of 'blowing up the tunnel' across conditions. Refer to Figure 6 for the police and non-police mock jurors' mode scores collapsed across conditions for the count of 'blowing up the tunnel.' Figures 4 & 5 regarding the count of 'blowing up the tunnel,' show a similar distribution pattern of mode scores across conditions of evidence reliability. In the case of police mock jurors the only 'not guilty' verdict given is for the Cb condition which interestingly scored '4' on the extent of guilt scale – 'probably guilty' (Figure 4). And conversely for the Cg condition despite a verdict of guilty, they were uncertain of the extent of guilt. For both Cb and Cg conditions, police mock jurors relied on expert forensic and psychological evidence (despite there being no expert psychological evidence presented in the control conditions). For the remaining experimental conditions, verdicts were 'guilty' with 'probably guilty' to 'clearly guilty' ratings. In the Expgg condition, police mock jurors claimed not to have used expert forensic or psychological evidence to inform their verdict when in fact both

sources of evidence were reliable. In the Expgb condition, they relied on expert psychological evidence (interestingly this is unreliable in this condition). For Expbb and Expbg conditions, both forensic and psychological evidence informed their verdict (under these conditions forensic evidence was unreliable). Police mock jurors appear to be unsuccessfully discriminating reliable from unreliable forensic and psychological evidence – as their deliberations appear to be independent of the appropriate evidence. For example, in the Expgg condition they rely on neither sources of evidence when in fact both forensic and psychological evidence are reliable. Likewise in the Expgb condition, they rely on psychological evidence to inform their verdict when it is an unreliable source: using forensic evidence would have been more appropriate as it is the reliable source of information. A similar interpretation for Expbb and Expbg can be used.

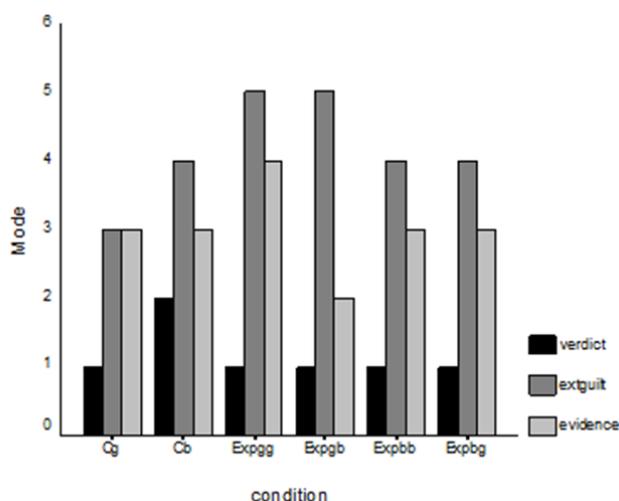


Figure 4 Police mock jurors' mode scores for verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)) on count of 'blowing up the tunnel' across the conditions of evidence reliability.

Non-police mock jurors, however, appear to be utilising evidence consistent with its reliability (Figure 5). For example, in the Cg condition they use forensic evidence to inform their verdict. In the Expgg condition, both sources of evidence inform their verdict which is consistent with both expert forensic and psychological evidence being reliable. The reliance on forensic evidence to inform their verdict in the Expgb condition is again consistent with the reliable source of evidence presented: forensic. For Cb, Expbb and Expbg conditions they rely on both forensic and psychological evidence to inform their verdict. In the case of the Cb condition the verdict is 'not guilty' and the rating is 'probably not guilty' and yet, both sources of unreliable evidence are relied upon. In this way, they are using the evidence to support the fact that the defendant must be 'not guilty' because the evidence is 'doggy.' In this case, the evidence is used appropriately and it can be argued that they realise the evidence should not be relied upon to find the defendant guilty. It appears that non-police mock jurors might have difficulty discriminating the reliable from the non-reliable evidence under the Expbb and Expbg conditions, because here, they deliberate a 'guilty' verdict and rely on both sources of evidence to inform their decision.

It could be argued that for some conditions both non-police and police mock jurors have difficulty discriminating reliable from unreliable evidence (more so the police). Or as suggested earlier, the

independent variable of evidence reliability across the conditions might not be sensitive enough to enable detection of reliability difference. This explanation falls short because as already noted, the non-police mock jurors are generally quite good at deciding the source of information they consider to be most credible. Variability of evidence reliability across the conditions clearly shows a good degree of manipulation. The problem seems to lie with knowing when evidence is unreliable and this is definitely more applicable to police mock jurors. The recognition of reliable evidence appears to be an easier task than the recognition of unreliable evidence. Interestingly, when the conditions are collapsed, the pattern of modal score distribution for extent of guilt, verdict and evidence are the same for police and non-police mock jurors (Figure 6). This can be explained by the overall modal scores for extent of guilt, verdict and evidence being the same for police and non-police mock jurors, however, the modal scores are simply distributed differently across the conditions.

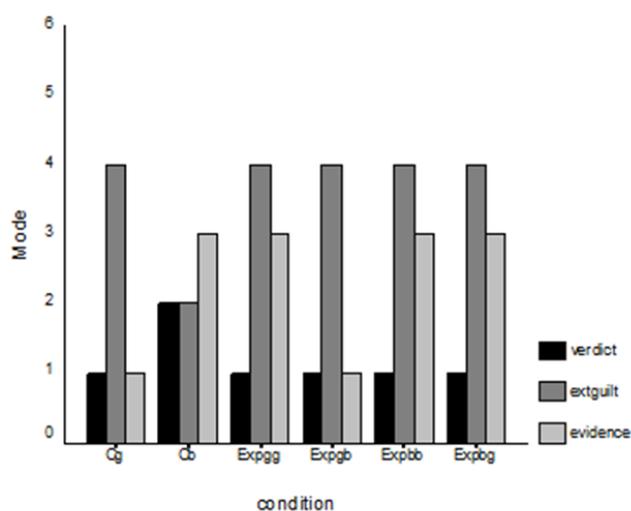


Figure 5 Non-police mock jurors' mode scores for verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)) on count of 'blowing up the tunnel' across the conditions of evidence reliability.

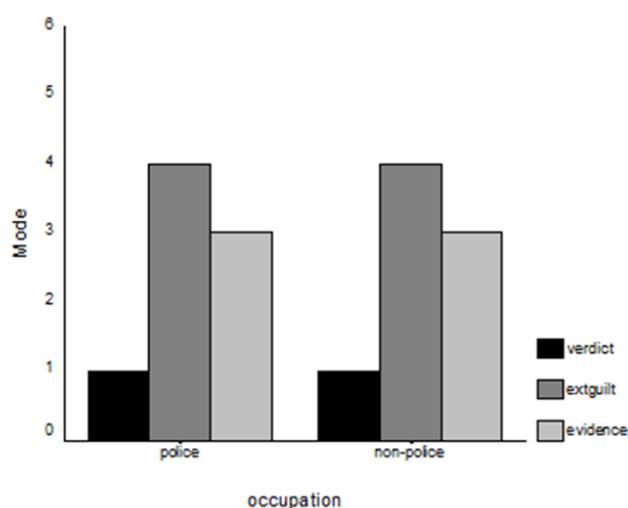


Figure 6 Police and non-police mock jurors' mode scores collapsed across conditions for the count of 'blowing up the tunnel'. Verdict (guilty (1) or not guilty (2)) extent of guilt (clearly not guilty (1) to clearly guilty (5)) and evidence (forensic (1), psychological (2), both (3) or neither (4)).

Inferential analysis: A MANOVA analysis was conducted on these data. Occupation yielded a significant difference between police and non-police mock jurors for evidence ($F(1,108)=20.79$, $p<0.001$). Verdict and extent of guilt, however, were not significant ($F(1,108)=0.05$, $p>0.05$ and $F(1,108)=0.04$, $p>0.05$ respectively). The difference in mean scores verifies this: both police and non-police mock jurors indicate a guilty verdict. For 'extent of guilt' both police and non-police mock jurors tend to be 'uncertain.' In the case of evidence, police mock jurors tend to use both types of evidence whereas non-police mock jurors use psychological. Across conditions significant differences were found for verdict ($F(5,108)=7.47$, $p<0.001$), extent of guilt ($F(5,108)=4.44$, $p<0.001$) and evidence ($F(5,108)=4.35$, $p<0.001$). In the case of verdict, significant differences using Scheffepost hoc comparisons were found for: Cb and Cg ($p<0.01$); Cb and Expagg ($p<0.01$); Cb and Expbg ($p<0.001$); Cb and Expbg ($p<0.01$). For 'extent of guilt' significant differences using Scheffepost hoc comparisons were found: Cb and Expagg ($p<0.05$); Cb and Expbg ($p<0.01$). And in the case of evidence a significant difference using Scheffepost hoc comparisons was found for: Expagg and Expbg ($p<0.05$). There were no significant interactions between occupation and condition found.

The Jonckheere Trend Test using the same criteria as previously was carried out on the count of 'blowing up the tunnel.' A significant trend for the police mock jurors was found ($S=-306$, $cv=256$, $p<0.05$) and a non-significant trend for the non-police mock jurors ($S=-38$, $cv=256$, $p>0.05$). Considering the raw data by looking at the rating most frequent under each condition, the trend for police mock jurors goes in the direction of the predicted order. In fact for the police mock jurors the ratings for each condition are 4,4,4,5,3 and 5. With the exception of the uncertainty expressed for the Cg condition, the 'extent of guilt' scores are increasing – but they start from a high extent of guilt score even when the evidence is unreliable! In the case of non-police mock jurors the ratings are: 2,4,4,4,4+5 and 4. Although not that dissimilar from the police trend, this was not significant.

Discussion

Findings from the MANOVA analysis suggest that both police and non-police mock jurors are deliberating similar verdicts when the six conditions are collapsed for both counts. For the count of 'hitting the bystander' there was a difference in 'extent of guilt' and 'evidence' ratings between police and non-police mock jurors: where non-police mock jurors were rating 'probably guilty' and using expert psychological evidence and police mock jurors were rating 'uncertain' and using both expert forensic and psychological evidence. In the case of 'blowing up the tunnel' police and non-police mock jurors alike are rating 'probably guilty' but non-police mock jurors are still more likely to rely on expert psychological evidence alone. Considering the six conditions, in the case of the first count of 'hitting the bystander,' verdict only just reached significance and this was the result of one comparison (Cb and Expbg where the Cb condition accrued more not guilty verdicts).

Results are thus consistent across the conditions. For the second count of 'blowing up the tunnel,' differences across the six conditions are found for all three scales. Verdict showed the major differences with there being fewer guilty verdicts for the Cb condition as compared with Cg, Expagg, Expbg and Expbg conditions. For the 'extent of guilt' ratings the lower ratings (probably not guilty and certainly not guilty) were found in the Cb condition as compared with the Expagg

and Expbg conditions (and this is primarily a result of the non-police mock jurors rating the Cb condition as 'probably not guilty'). Finally, the only significant Scheffepost hoc finding for the evidence scale was between Expagg and Expbg conditions (this is the result of police mock jurors not using any expert evidence to inform their verdict in the Expbg condition and non-police mock jurors relying on forensic expert evidence only).

It appears that when the modal scores for extent of guilt are considered across the six conditions, non-police mock jurors might be using evidence more appropriately to inform their verdict, especially for the count of 'blowing up the tunnel'. However, there was no statistical evidence to support this – the Jonckheere Trend Test failed to find a trend in the direction of increasing reliable expert evidence using the 'extent of guilt' scores. A trend was found for police mock jurors but they were almost scoring at ceiling in the conditions of unreliable evidence to begin with. Nevertheless, when police and non-police mock jurors' scores were collapsed, differences in verdict across the conditions could be seen. Clearly the level of evidence reliability is having some influence on the verdict ratings, but for all mock jurors regardless of police and non-police categories. What this seems to suggest, certainly for 'blowing up the tunnel,' is that the reliability of evidence does affect verdict deliberations regardless of occupation. In other words, police mock jurors are no more proficient at discriminating the reliable from the unreliable evidence than are non-police mock jurors. Hence, H_1 and H_2 fail to have statistically significant support. Although police are not involved directly in the deliberation process, they are involved in the initial stages of gathering evidence and formalising a case based on the evidence collected. It is therefore important to build a case on reliable and valid information that is likely to be accepted by the CPS on the grounds of a 50 percent chance of successful prosecution. The fact that non-police and police mock jurors for the first count of 'hitting the bystander' found it hard to differentiate between the reliable and unreliable conditions of evidence as determined from their scores for verdict, extent of guilt and evidence, suggests that they were using some other means of decision making not readily accounted for in this study. What they are using to base their decision making on is difficult to ascertain. It is plausible, however, that they failed to understand the differences in evidence reliability; failed to comprehend or entertained different interpretations of the meaning of 'probability of guilt' or 'beyond reasonable doubt'⁴⁶ or formulated their own story modelling hypotheses⁴⁷ based on their perceptions of crime seriousness.¹⁰ This could have further interacted with their model of perceived justice.

Lerner⁴⁸ addressed people's notion of perceived justice in his 'just deserts hypothesis.' Lerner's 'just deserts' hypothesis is important here. Police are constantly faced with inequitable situations where offenders do wrong to victims and they are expected to find and apprehend such individuals. The nature of the questions asked by police and the course of interrogative action taken, relates to how police perceive the crime event. Police may not always be aware or have access to all the facts, however, and when evidential facts are limited they may have to rely on assumptions: the constitutive facts which they have already formed. These could potentially dictate the direction of the investigation. Such assumptions by police, when evidential facts are limited can lead to misinterpretations of what evidence is there and its relevancy and reliability and consequently its implications. In effect police can make mistakes which could be due to limited evidential facts of the case or to misinterpretation of the reliability of information.

It is beyond the scope of this study to predict how mock jurors arrived at their decisions. The modal scores when collapsed across conditions (Figure 3) do provide some insight to the extent to which police mock jurors used expert forensic or psychological evidence to inform their verdict. More often than not police mock jurors ignored both and based their decision of guilt on something else. Hence, it appears that there is some support for H3 that they are less open to expert eyewitness evidence. Again by referring to Figure 3, it is obvious that non-police mock jurors are more often relying on psychological expert evidence to inform their decision of guilt. This makes perfect sense considering that for the count of 'hitting the bystander' eyewitness testimony would have been useful - hence providing support to H₃. In the case of count two, 'blowing up the tunnel,' a different picture emerges. It appears that both police and non-police mock jurors considered evidence reliability more than for the first count of 'hitting the bystander.' However, when studying the modal scores (Figures 4&5) it becomes apparent that non-police scores across verdict, extent of guilt and evidence tended to be in keeping with the condition of evidence reliability. The non-police mock jurors are getting it right more often than the police mock jurors. Hence, under the condition of poor expert psychological reliability, police mock jurors are giving a guilty verdict and relying on psychological evidence to inform their decision, when they should be relying on expert forensic evidence. The two problem conditions for police and non-police mock jurors as regards to identifying evidence reliability, appears to be Expbb and Expbg. Both police and non-police mock jurors are inferring guilty verdicts for these two groups and relying on expert forensic and psychological evidence to do so. It might be the case that reliable evidence is easier to recognise than unreliable evidence. Again there appears to be little support for H₁ and H₂. In the case of openness to expert forensic and psychological evidence for the count of 'blowing up the tunnel', it appears that both police and non-police mock jurors are making use of these forms of information - hence, providing some support to H₃.

Accounting for why police are more likely than non-police mock jurors to determine a 'guilty' verdict based on neither forensic nor psychological expert evidence is easier to explain than why both police and non-police mock jurors have difficulty identifying that Expbb and Expbg conditions contain unreliable evidence. As suggested earlier, police role and culture might play a hand in how they perceive the constitutive facts and therefore piece the evidence together to 'story-tell.' It could be that police training encourages them to interpret situations in a particular way and to detect suspicious behaviour. This perceptual set could encourage police to cognitise the meaning of constitutive facts in a negative or suspicious manner, therefore increasing the likelihood of wanting evidence and seeking evidence that could support their suspicions. Support for this comes from a host of different research studies.^{14,21,49} Hence, what Skolnick, for instance, seems to be suggesting is that police look for confirmatory rather than disconfirmatory evidence to support their suspicions. This can only be speculated as there was no way of knowing what suspicions the police mock jurors may have held when reading the transcripts. Equally it could be argued that non-police mock jurors are using a 'story telling' tactic. They too are influenced as pointed out by Pennington et al.,⁹ by constructing a 'causal-sequential representation' of factual information presented in court. They formulate a narrative story that best fits the facts and this can be based on their world knowledge about events containing similar facts. The only difference potentially between police and non-police is that in the real world police develop

the story and jurors develop the story from one given to them. In this study both police and non-police mock jurors are provided with the story, so how their world knowledge influences their understanding, interpretation and decisions should be similar in this respect. Police additionally have 'police knowledge' to add to their story. But as has been found there does not appear to be a significant contribution of this to the findings.

Conclusion

Pennington et al.,⁹ model can be applied to both police and non-police decision making. The seven elements of processing information that jurors should try to achieve can equally apply to police and non-police mock jurors in this study. Both assimilate information whilst reading the transcript. Both will try to construct a 'causal-sequential representation' of the factual information presented in the transcript. In other words, they will formulate a narrative story that best fits the facts.⁴⁷ In this study findings appear to suggest that both police and non-police mock jurors have difficulty discriminating reliable from unreliable evidence (although not statistically supported, non-police mock jurors appear slightly better at the task). It was assumed that police are more likely to develop a sequence of constitutive facts (similar to storytelling) than non-police because of their persecutory role of piecing information together and then finding the evidence.⁵⁰⁻⁵² However, it could be that as information processors we are always looking to make sense of events and making attributions as to cause. It seems a likely explanation that all mock jurors do this regardless of position and occupation.

Acknowledgments

None.

Conflicts of interest

The author declares that there are no conflicts of interest.

References

1. Hastie R, Pennington N. *Explanation-based decision-making*. In: Connolly T, Arkes HR editors. *Judgement and decision-making: An interdisciplinary reader*. Cambridge University Press, USA. 2000; p.212-228.
2. Monckton-Smith J, Adams T, Hart A, et al. *Introducing forensic and criminological investigation*. Sage Publications Limited, UK. 2013. p. 1-232.
3. Williams H. *Criminology and Criminal Justice*. Butterworths, UK. 1982.
4. Williams K. *Textbook on Criminology*. 3rd edn. Blackstone Press Limited, UK. 1991. p. 1-664.
5. Zander M. *The Police and Criminal Evidence Act 1984*. 2nd ed, Sweet and Maxwell, UK. 1990.
6. La Fave W. *Arrest: The decision to take a suspect into custody*. In: Williams J, editors. *Criminology and Criminal Justice*, Butterworths, UK. 1965.
7. Anderson TJ. *The Netherlands criminal justice system: An audit model of decision making*. In: Malsch M, Freckelton I, editors. *Expert bias and partisanship: A comparison between Australia and the Netherlands*. Psychology, Public Policy and Law. 1999;11(1):42-61.
8. Malsch M, Freckelton I. *Expert bias and partisanship: A comparison between Australia and the Netherlands*. Psychology, Public Policy and Law. 2005;11(1):42-61.

9. Pennington N, Hastie R. *Juror decision making models: The generalization gap*. *Psychological Bulletin*. 1981;89:246–287.
10. Ellsworth PC. *Are 12 heads better than one? Law and Contemporary Problems*. 1989;52:205–224.
11. Leippe MR. The appraisal of eyewitness testimony. In: Ainsworth P, editor. *Psychology, Law and Eyewitness Testimony*, Wiley, UK. 1994. p. 385–418.
12. Loftus E. Eyewitnesses: Essential but unrealistic. In: Ainsworth P, editor. *Psychology, Law and Eyewitness Testimony*, Wiley, UK. 1984.
13. Clifford BR, Richards VJ. Comparison of recall by policemen and civilians under conditions of long and short duration of exposure. *Perceptual and Motor Skills*. 1977;45:503–512.
14. Yuille JC. Research and teaching with police: A Canadian example. *International Review of Applied Psychology*. 1984;33:5–23.
15. Ellis N. *Improving guesswork: An evaluation of Canter's multivariate model of offender profiling*. In the Proceedings of the British Psychological Society. 1999;7(2).
16. Kocsis RN, Middeldorp J, Karpin A. Taking stock of accuracy in criminal profiling: The theoretical quandary for investigative psychology. *Journal of Forensic Psychology Practice*. 2008;8(3):244–261.
17. Snook B, Eastwood J, Gendreau P, et al. The importance of knowledge cumulation and the search for hidden agendas: A reply to Kocsis, Middeldorp and Karpin. *Journal of Forensic Psychology Practice*. 2010;10(3):214–223.
18. Howitt D. *Introduction to Forensic and Criminal Psychology*. 4th ed. Pearson, UK. 2011; p. 1–604.
19. Alison LJ, Smith MD, Morgan K. *Interpreting the accuracy of offender profiles*. *Psychology, Crime and Law*. 2003;9(2):185–195.
20. De Paulo PJ, Pfeiffer RL. On the job experience and skill at detecting deception. *Journal of Applied Social Psychology*. 1986;16(3):249–267.
21. Vrij A, Semin GR. Lie experts' beliefs about nonverbal indicators of deception. *Journal of Non Verbal Behaviour*. 1996;20: 65–80.
22. Vrij A. The impact of information and setting on detection of deception by police detectives. *Journal of Non Verbal Behaviour*. 1994;18(2):117–137.
23. Bond CF. A few can catch a liar, sometimes. *Applied Cognitive Psychology*. 2008;22(9):1298–1300.
24. Skolnick J. *Justice Without Trial: Law Enforcement in Democratic Society*. Wiley and Sons Inc, UK, 1966. p. 1–309.
25. Stephenson GM. *The Psychology of Criminal Justice*. Blackwell Publishing Ltd, UK. 1992. pp. 1–300.
26. Balch RW. The police personality: Fact or fiction. *Journal of Criminal Law, Criminology and Police Science*. 1972;63(1):106–119.
27. Walsh D, Milne R. Giving PEACE a chance. *Public Administration*. 2007;85(3):525–540.
28. Hill C, Memon A, McGeorge P. The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*. 2008;13(2):357–371.
29. Taylor S. *Crime and criminality: A multidisciplinary approach*. Taylor and Francis, UK. 2016. p. 1–822.
30. Inciardi JA. *Criminal justice*. 3rd ed, Harcourt Brace Jovanovich, UK. 1990.
31. Nelson DL, Quick JC. *Organizational behaviour: Foundations, realities and challenges*. Thomson South-Weston, USA, 2006. p. 217–636.
32. Barker T, Hunter R, Rush J. *Police systems and practices: An introduction*. Prentice Hall Career and Technology, USA, 1994. p. 1–354.
33. Johnson LB, Todd M, Subramanian G. Violence in police families: Work-family spillover. *Journal of Family Violence*. 2005;20(1):3–12.
34. Berg BL. *Policing in modern society*. Butterworth-Heinemann, USA. 1999.
35. Twersky-Glasner A. Police personality: What is it and why are they like that? *Journal of Police and Criminal Psychology*. 2005;20(1): 56–67.
36. Moston S, Stephenson GM, Williamson TM. Police interrogation styles and suspect behaviour. In: Losel F, Bender D, editors. *Psychology and Law International Perspectives*. Walter de Gruyter, USA. 1990.
37. <https://www.merriam-webster.com/dictionary/res%20gestae>
38. Penrod SD, Cutler BL. Assessing the competence of juries. In: Weiner IB, Hess AK, editors. *A Handbook of Forensic Psychology*. Wiley and Sons, USA. 1987. p. 293–318.
39. Richardson L. Narrative and Sociology. *Journal of Contemporary Ethnography*. 1990;19(1):116–135.
40. Visher CA. Juror decision making: The importance of evidence. *Law and Human Behaviour*. 1987;11(1):1–17.
41. Otto AL, Penrod SD, Dexter HR, et al. The biasing impact of pretrial publicity on juror judgements. *Law and Human Behaviour*. 1994;18(4):453–469.
42. Kovera MB, McAuliff BD. The effects of peer review and evidence quality on judge evaluations of psychological evidence: Are judges effective gatekeepers? *Journal of Applied Psychology*. 2000;85(4):574–586.
43. Wilson RA. *Writing history in international criminal trials*. Cambridge University Press, USA. 2011; p. 1–272.
44. Leippe MR, Eisenstadt D. *The influence of eyewitness expert testimony on jurors' beliefs and judgements*. In: Cutler BL, editors. *Expert testimony on the psychology of eyewitness identification*. Oxford University Press, USA. 2009; p.169–199.
45. Loftus E. Impact of expert psychological testimony on the unreliability of eyewitness identification. *J Appl Psychol*. 1980;65:9–15.
46. Thomas EAC, Hogue A. Apparent weight of evidence, decision criteria and confidence ratings in juror decision making. *Psychological Review*. 1976;83(6):442–465.
47. Pennington N, Hastie R. Explaining the evidence: Tests of story model for juror decision making. *Journal of Personality and Social Psychology*. 1992;62(2):189–206.
48. Lerner M. *The unjust consequences of the need to believe in a just world*. In: Sales BD, editors, *Perspectives in law and psychology*. The jury, judicial and trial process, Plenum, USA. 1966.
49. Winkel W, Koppelaar L, Vrij A. Creating suspects in police-citizen encounters: Two studies on personal space and being suspect. *Social Behaviour*. 1988;3(4):307–319.
50. Cooper J, Neuhaus IM. The 'Hired Gun' effect: Assessing the effect of pay, frequency of testifying, and credentials on the perception of expert testimony. *Law Hum Behav*. 2000;24(2):149–171.
51. Cutler BL, Dexter HR, Penrod SD. Expert testimony and jury decision making: An empirical analysis. *Behavioural Sciences and Law*. 1989;7(2):215–225.
52. Diamond SS, Casper JD. Blindfolding the jury to verdict consequences: Damages, experts and the civil jury. *Law and Society Review*. 1992;26(3):513–563.