Title

Evidentiary Instructions Improve Mock Juror Assessment of Feature-Comparison Evidence

Abstract

Feature-comparison evidence has been introduced in court without sufficient scientific validation and has been at the heart of numerous miscarriages of justice. Juror assessment of such evidence and the efficacy of evidentiary instructions were examined through a mock jury experiment with case reports featuring either central or peripheral feature-comparison evidence. In a case-control design (N = 174), the test group was exposed to an evidentiary instruction about the ear print evidence presented in the first case report (adapted from *R* v *Dallagher* [2002] EWCA Crim 1903) whereas the control group did not receive such an instruction. The provision of this instruction resulted in a significant decrease in verdict severity with a large effect size. For the second case report (based on *R* v *George (Barry)* [2007] EWCA Crim 2722), all subjects were asked to return verdicts based on circumstantial evidence, gunpowder residue evidence, and an evidentiary instruction about that gunpowder residue evidence. Verdict severity increased significantly after the provision of gunpowder residue evidence, followed by a subsequent reduction in verdict severity after the introduction of an evidentiary instruction. Furthermore, there was a significant difference in verdict severity between the test and control group, suggesting that the test group exhibited a scepticism effect brought about by the initial evidentiary instruction about ear print evidence. This study demonstrates that although mock jurors consider feature-comparison evidence a convincing indicator of guilt, the provision of an evidentiary instruction has the potential to educate jurors about the limitations of such evidence.

Keywords

Juries, Decision Making, Forensic Science, Scientific Evidence, Evidentiary Instruction, Feature Comparison Evidence

Introduction:

England and Wales adhere to a procedural tradition of adversarial fact-finding and narrative building in which the prosecution and defence are expected to collect evidence in the pre-trial phase to subsequently present these findings in court (Brants & Field, 2016). Within this framework of adversarial litigation jurors are considered impartial decision-makers who decide on innocence or guilt after an adequate evaluation and discussion of the evidence (Brants & Field, 2016; Dunne, 2015). Criminal case judgment in England and Wales is thus strongly wedded to the principle of lay fact-finding by jurors, who are expected to employ common sense reasoning in their assessment of the facts of the case (Jackson et al., 2015).

With the increased popularity and complexity of forensic evidence in court comes the question whether jurors are equipped with sufficient tools to carry out this responsibility. Legal reviews of past convictions have given rise to an increased scepticism about the scientific basis of forensic evidence and have put a spotlight on the occurrence of miscarriages of justice due to misconceptions regarding its reliability (Adam, 2016). The discovery of a single piece of evidence during a criminal investigation can become of crucial importance in court, and the collection and analysis of evidence have therefore become one of the most important and challenging features of the practice of law (Anderson et al., 2005). This article investigates mock jurors' assessment of forensic evidence through a literature review and a mock jury experiment in which the impact of evidentiary instructions on decision-making in cases with central and peripheral feature-comparison evidence is examined.

Scientific Evidence

Forensic science refers to all scientific tests or techniques relevant to legal proceedings. The term masks the complexities that come with the use of forensic science in court as a significant amount of such evidence may come from disciplines outside of what has traditionally been understood as the forensic sciences and can be produced by institutions and individuals whose primary discipline is unrelated to the criminal justice system (e.g. expert medical testimony in Sudden Infant Death Syndrome cases) (Sense about Science & Euroforgen, 2017; Wilson et al., 2014). Due to this disciplinary fragmentation little attention has been paid to underlying issues such as the potential of replication, confirmation through peer review or the risks associated with bias, let alone to the difficulties that come with communicating scientific knowledge to a lay audience (Dror et al., 2017; Jackson et al., 2015).

The increased concentration of expertise within the police effectively limits the availability of equivalent experts to the defence, whose independent examination of the evidence may at times be the only effective way of resolving bias (Wilson et al., 2014). In some fields of expertise it is in fact impossible for a defence expert to obtain domestically recognised qualifications (Wilson et al., 2014). For example, in *R v Smith* [2011] EWCA Crim 1296, the court expressed concern about the fact that there is no opportunity to become qualified as a fingerprint expert in England and Wales except through participation in the police force training. It then seems that the equality of arms on which adversarial fact-finding is thought to depend may not exist in a system in which the prosecution has superior access to witnesses, surveillance materials, forensic science support and databases (Brants & Field, 2016; Wilson et al., 2014). The result of such inequality is that a judge may be more likely to accept evidence

provided by a prosecution expert, even though the context is one in which the expert pool available to the defence is deficient (Wilson et al., 2014).

The closure of the Forensic Science Service (FSS) in 2012, is believed to have destabilised and even posed a threat to the production of scientific evidence (Wilson et al., 2014). The lack of a common form of governance regulating the forensic sciences has compromised the ability of judges and lawyers to recognise problems with scientific evidence and to improve mechanisms for admitting, representing, and reviewing such evidence (Martire & Edmond, 2017; Roach, 2009). Consequentially, courts may rely on simplistic or even misguided proxies for the evaluation and admission of such evidence while judges are not well positioned to guide jurors on its assessment (Martire & Edmond, 2017). Although validity and reliability studies can produce indicative levels of performance, courts tend to either not require this information or are unnecessarily dismissive of such findings (Martire & Edmond, 2017). The growing importance of forensic evidence in court thus poses a challenge to both experts, who find themselves in a position in which they must communicate complicated scientific findings to a lay audience, and to jurors who must adequately assess the value of such evidence (Diamond, 2007; Hans, 2007). Research has demonstrated that judges often fail to comprehend scientific evidence but nevertheless report high confidence when asked about their scientific literacy (de Keijser & Elffers, 2012; McAuliff & Duckworth, 2010). A survey in the USA alarmingly found that as little as five percent of the participating judges could explain the meaning of falsifiability while an even smaller percentage managed to define error rate (Benforado, 2015). The fact that jurors who may be less knowledgeable and educated than judges are expected to comprehend and assess the validity of scientific evidence admitted in court is concerning.

Admissibility in England and Wales – State of Affairs

Courts in England and Wales have adopted some of the most liberal admissibility practices among advanced common law jurisdictions (Edmond, 2015b). The rules governing the admissibility of expert testimony have developed within the common law system and are referred to as the common-law admissibility test (Henneberg, 2015). This test stipulates that, to be considered sufficiently reliable to be admitted, evidence must be relevant, necessary to assist the jury (helpful), and given by a competent witness (Ward et al., 2017). As jurors may attach great weight to evidence provided by experts, additional rules regulate how the court must be persuaded of the reliability of its scientific foundations (Ormerod & Sturman, 2005). However, there is no preliminary inquiry about the reliability of the science if the court feels acquainted with the type of evidence. The extent to which the courts scrutinise the reliability and validity of scientific evidence, and by which criteria they assess reliability to begin with, thus remains unclear (Ormerod & Sturman, 2005). A plethora of cases exist in which questionable scientific evidence significantly influenced final verdicts to, on appeal, be considered unreliable or at least to require special consideration (Ireland & Beaumont, 2015).

The Turner rule, established in the leading case *R* v *Turner* (1975), limits the reception of expert opinion evidence by stipulating that it is inadmissible unless it provides the court with information outside of the jurors' common experience and knowledge. It can be argued that one should not limit this admissibility requirement to materials the jury does not know about, but extend it to materials the jury mistakenly *think* they know about (Ormerod & Sturman,

2005). This line of reasoning has become increasingly relevant amidst the popularity of television programmes which make the use of forensic techniques in criminal investigations seem nearly infallible. Prosecutors have attributed the jury's unrealistic expectations of crime-solving to such portrayals and have reported concerns that jurors may discredit prosecutions that seem thin on science (Weiss & Xuan, 2015). Academics, on the contrary, fear that jurors may fail to properly evaluate the evidence provided and may blindly accept scientific testimony instead (Benforado, 2015).

According to the Turner rule, expert evidence must furthermore be necessary for a proper resolution of the dispute and must be based on facts which themselves can be proven by admissible evidence (Ebisike, 2008; Freckelton, 2014). These regulations may sound restrictive but do not stipulate that a scientific technique must be generally accepted. It was decided in *R* v *Robb* (1991) that general acceptance is not a necessity for the admissibility of expert testimony as long as the evidence is sufficiently established to be reliable (Ebisike, 2008). However, a new hurdle arose with the increased introduction of novel techniques as it became apparent that courts had at times failed to give new techniques the focused attention they warrant (Ireland & Beaumont, 2015; Ormerod & Sturman, 2005). One of the dangers of introducing novel techniques is that, as there are fewer experts to call upon, the expert in question may be subjected to less effective challenge and the evidence may be afforded a disproportionate weight (Ormerod & Sturman, 2005). On appeal this creates further problems as proclaimed innocence in a conviction based on dubious evidence is not always a convincing reason for the Criminal Cases Review Commission (CCRC) to accept an application or to refer the case to the Court of Appeal (Henneberg, 2017).

Although the admissibility rules may seem to establish safeguards against the admittance of junk science in court, the House of Commons' Science and Technology Committee has voiced concerns that scientific expert witness testimony is admitted in court without sufficient scrutiny. The Law Commission (2011) of England and Wales has proposed that judges should provide a cautionary instruction to the jury if a case hinges on disputed scientific evidence (Child et al., 2015). Although the Ministry of Justice shared the concern that courts are ill-equipped to resolve arguments about the scientific authority of experts, the Law Commission's draft bill was rejected in 2013 with expenses cited as the main reason for this rejection (Henneberg, 2015; Wilson et al., 2014). Instead of accepting the proposed bill, the government asked the Criminal Procedure Rule Committee to consider amendments to the Criminal Procedure Rules to introduce the spirit of the Law Commission's recommendations (Stockdale & Jackson, 2016). A series of enquiries and reports on expert evidence and the law have taken place since, but this has ultimately led to a retention of the status quo (Adam, 2016). Even if the Law Commission produced and promoted an improved standard, the issue at stake is that it proposes to impose it on a system that is insufficiently prepared to apply it effectively (Edmond & Roberts, 2011).

A promising recent development aimed to stimulate a better understanding of complex evidence is the creation of scientific guides to aid judges in the assessment of scientific validity and admissibility (Ghosh, 2017). However, if academics have not yet reached consensus on the limitations of a certain science, or a new forensic technique emerges, there will be no such guide available to aid judges in its assessment. If unreliable forensic evidence is admitted in court as a result of this, safeguarding the jury's ability to understand and evaluate forensic

evidence remains imperative to a proper functioning of the judicial decision-making process (McAuliff & Duckworth, 2010; Thomas, 2010).

Judicial Safeguards

Although judicial safeguards such as evidentiary instructions have been hypothesized to provide the education necessary to enable jurors to make informed decisions about scientific evidence, the provision of such instructions is still uncommon. The court has, however, established guidelines for judges in trials that involve disputed identification evidence. The Turnbull directions (*R* v *Turnbull* [1997] QB 224) allow judges to provide jurors with precautions before the introduction of eyewitness testimony evidence (Bromby et al., 2007). These precautions are intended to raise awareness of issues that could impact the reliability of the eyewitness testimony evidence. One could argue that similar directions should be provided before the introduction of other types of evidence. Indeed, the provision of evidential instructions in an adversarial system comes with distinct advantages for such instructions are focused and concise, authoritative when coming from the trial judge, less costly than competing expert testimony, and can avoid confusion created by duelling experts (Jones, 2017). To be effective, safeguards should maximise juror sensitivity to factors that influence the reliability of forensic evidence, and the timing of an instruction may impact this effectiveness (Cutler et al., 1990; Jones et al., 2017). Research suggests that instructions on the burden of proof and reasonable doubt are most effective when delivered prior to the introduction of evidence (Jones et al., 2017). In line with these findings it can be hypothesised that, if provided before the introduction of forensic evidence, evidentiary instructions can assist jurors in evaluating evidence in line with academic and legal standards (Jones et al., 2017). However, for evidentiary instructions to be effective jurors must be capable and willing to follow them. Unfortunately, research has revealed that jurors experience difficulties in understanding and applying many types of judicial and evidentiary instructions (Baguley et al., 2017; Jones et al., 2017; Valentine & Fitzgerald, 2016).

It is important to bear in mind that the introduction of an evidentiary instruction is not without its problems as it could come with a risk of undue caution (Bromby et al., 2007). The desired result of a jury direction is to induce juror sensitivity, not scepticism. The latter can be defined as a general distrust of all evidence, even when this caution is not merited (Leverick, 2014). It is possible that the introduction of an evidentiary instruction gives rise to a scepticism effect (McAuliff & Duckworth, 2010). This is in line with findings of opposing expert testimony research by Levett and Kovera (2009), who demonstrated that the mere presence of an opposing expert causes jurors to become more sceptical of the initial expert's testimony. However, scepticism effects found within expert testimony research may be due to the unique way in which experts educate the jury relative to other safeguards (Jones et al., 2017). Such findings do not necessarily demonstrate that an evidentiary instruction would cause a similar scepticism effect.

Jury Research

Legal practitioners have long held the opinion that jurors are capable of assessing scientific evidence despite a lack of legal or scientific sophistication (Boudreau & McCubbins, 2009). This lack of specialized knowledge and experience was considered advantageous for it can act

as a safeguard against a biased judiciary or overzealous prosecution (Hans, 2012). However, it is important not to take at face value the assumption that jurors, through general education and experience, have become sufficiently equipped to cope with the demands of interpreting complex scientific reasoning (Jackson et al., 2015). The problem at stake is that findings of scientific studies are often counter-intuitive and outside the knowledge of an average juror (Leverick, 2016). Academics have therefore expressed concerns that jurors are not capable of assessing scientific evidence and may ignore crucial evidence or take it at face value by focusing on the expert's credentials (Diamond, 2007; McAuliff & Duckworth, 2010). This heuristic, known as source expertise, could be particularly present in legal settings if jurors are aware that experts must meet certain criteria to provide testimony in court and may therefore assume that the testimony must be relevant and reliable.

Research has demonstrated that, when confronted with complex scientific testimony, jurors considered scientists with a PhD from a prestigious university more persuasive (Cooper & Hall, 2000). Jurors may moreover believe that expertise is linked to excellent performance or skill in a particular task rather than to a broad range of professional competencies (Martire & Edmond, 2017). In reality experts may become respected and considered experienced for critical thinking, competent use of assessment tools, rapport building with clients, or communication skills instead of for making more correct predictions (Martire & Edmond, 2017). Such observations underscore the responsibility of judges and lawyers to ensure that evidence is presented in a comprehensible manner, but courts have expressed discomfort with the task of reviewing methodological foundations of scientific research and tend to avoid methodological explorations of expert evidence during trial (Jackson et al., 2015; Martire & Edmond, 2017; Newirth, 2016).

Jury comprehension research in England and Wales has demonstrated that more than half of jurors do not entirely comprehend what happens in court, citing legal terminology as the main impediment (Bromby et al., 2007). Thomas (2010) conducted a study in which 797 jurors at three courts in England observed the same simulated trial and heard the same directions and found that, although over half of the jurors reportedly perceived the directions easy to understand, only a minority of jurors actually fully comprehended the directions. It has been argued that at least some barriers to comprehension may stem from limitations in working memory. Asking jurors to retain all the information discussed at trial may be unreasonable (Leverick, 2014). Indeed, in the aforementioned study a written summary of the judge's direction, provided at the time of the judge's verbal instructions, improved juror comprehension of the law (Thomas, 2010). In line with this finding, a trial judge in England provided written directions to jurors in all criminal cases for several months and concluded that it almost eliminated request for reminders and further guidance on the law (Leverick, 2014). Research has moreover suggested that juries are more likely to follow directions if it is explicitly explained why these directions are given (Leverick, 2014). In the case of forensic evidence, a jury instruction could stipulate that individuals have been wrongfully convicted on the basis of flawed evidence of the same nature before. Providing written instructions is however not an adequate solution if the directions are inherently unclear. An unawareness of the limitations of certain types of evidence can only be addressed by judicial training (Leverick, 2014).

The provision of jury instructions, whether verbal or in writing, has been subject of ongoing debate. Much research has focused on simplifying complex instructions but, as of yet, not one technique was found to consistently improve mock juror comprehension of complex evidence (Baguley et al, 2017). Cicchini and White (2016) provided 300 mock jurors with one of three instructions on the burden of proof after reading a summary of a hypothetical criminal case. After reading the case participants either received an instruction to search for the truth, a proper instruction on reasonable doubt, or a combined instruction on reasonable doubt followed by a stipulation to search for truth instead of doubt. The authors reported a near double conviction rate amongst jurors who had been instructed to search for the truth (Cicchini & White, 2016). This result seems to indicate that jury instructions can significantly alter jury decision-making. Other studies, however, have uncovered no effect or more moderate results. Ellison and Munro (2015), for example, explored the extent to which participants can understand and apply a judicial direction provided both verbally and in writing by asking participants to deliberate in groups towards a unanimous verdict after observing a rape trial re-enactment. The instruction intended to guide mock jurors' assessment of the evidence and stipulated what legal tests were applicable. The authors found that participants made limited use of the instruction and instead relied on personal recollections and (mis)interpretations of legal tests as well as on personal evaluations of weight and credibility (Ellison & Munro, 2015).

The jury instructions discussed up until this point can be classified as procedural instructions whereas this study is particularly concerned with evidentiary instructions. O'Donnell and Safer (2017) examined whether an enhanced evidentiary instruction, in which empirical findings were added to a standard instruction, could sensitize mock jurors to confession evidence in a criminal trial transcript in which a defendant's recanted confession was either gathered using coercive or appropriate tactics. The results were more promising and indicated that, compared to the standard instruction, the enhanced instruction successfully sensitized mock jurors to the strength of the confession evidence rather than merely inducing a scepticism effect (O'Donnell & Safer, 2017). In 2011, the New Jersey judiciary in the United States implemented a reformed evidentiary instruction about eyewitness identification evidence, often referred to as the Henderson instruction. Papiliou et al. (2014) put this instruction to the test in a study in which 335 mock jurors watched a murder trial for which the strength of the identification evidence was manipulated. The participants received either a standard evidentiary instruction or the Henderson instruction, which addresses the shortcomings of the standard evidentiary instruction by outlining the three stages of memory and by explaining how nine estimator variables and seven system variables, which are in the State's control, influence identification accuracy. Those who received the standard instruction were found to be more than twice as likely to convict. However, those who received the Henderson instruction discounted weak and strong testimony in a similar manner and thus did not display an improved ability to discern quality. In other words, the evidentiary instruction triggered scepticism instead of increased sensitivity. Dillon et al. (2017) reported similar findings when conducting a study in which 468 participants watched a trial simulation in which the strength of eyewitness evidence was manipulated. The participants either received a Henderson instruction prior to the eyewitness testimony, at the end of the trial, or not at all. The authors found that the Henderson instruction induced overall scepticism instead of sensitizing jurors.

Deficiencies in the forensic sciences have become particularly prevalent across feature comparison disciplines (in which a crime scene sample is compared with a reference sample) as such evidence has been introduced in court without meaningful scientific validation or reliability testing (President's Council of Advisors on Science and Technology, 2016). When it comes to feature comparison evidence, courts in England and Wales routinely admit the opinions of forensic experts (Edmond, 2015b). Eastwood and Caldwell (2015) examined whether opposing expert witness testimony and the provision of judicial instructions can mitigate the impact of invalid forensic science testimony about hair comparison evidence. The authors found that judicial instructions had no impact on verdict decisions and, although the opposing expert condition seemed more successful, both safeguards were relatively ineffective (Eastwood & Caldwell, 2015). The finding that the provision of opposing expert testimony reduces the rate of guilty verdicts is not so comforting when considering that, realistically, the first expert may not get challenged at all as the defence often does not employ an opposing expert. This implies that jurors faced with invalid feature comparison evidence may be forced to rely on their own determination to assess the quality of the testimony (Eastwood & Caldwell, 2015). Moreover, although the judge in this study provided similar information in the judicial instruction as the opposing expert witness did in their testimony, contrary to the expert's testimony no mention was made of the fact that this instruction was supported by experts or scientific findings (Eastwood & Caldwell, 2015). The inclusion of this information could have made the instruction more effective. The judicial instruction about hair comparison evidence was furthermore provided at the end of the trial and was embedded within general instructions about how jurors should reach their verdict. It is possible that the participants had already processed the evidence by then and did not expect new information (Eastwood & Caldwell, 2015). As of yet research on the efficacy of jury instructions, and the risk of such instructions inducing scepticism effects, thus seems inconclusive.

A Mock Jury Experiment

Research Design

An online mock jury experiment with a mixed methods design of closed and open survey questions was created in which the qualitative answers were converted into quantitative data for ease of subsequent analysis. The independent variables were central versus peripheral scientific evidence and evidentiary instruction versus no evidentiary instruction, whereas the dependent variables were verdict and confidence level. The dependent variables were assessed through a Likert scale with each Likert item providing a range of fixed-choice answer options.

Materials

After an analysis of over 100 established and alleged miscarriages of justice in England and Wales, two cases were selected based on the type of scientific evidence at stake and the significant role of expert testimony during the trial. The cases selected for this experiment centre on feature comparison evidence. This approach was adopted because feature comparison is a common forensic activity, science has clear standards to determine whether such methods are reliable, and it has become apparent that faulty feature comparison evidence has been at the heart of numerous miscarriages of justice (Adam, 2016; President's Council of Advisors on Science and Technology, 2016). Moreover, the problems with such evidence cannot merely be

attributed to poor performance by a few experts as the reliability of many feature comparison methods has never been meaningfully investigated. This is worrying as it has been hypothesised that the majority of jurors lack the ability to interpret the probative value of feature comparison evidence and that the prejudicial impact moreover may be high as jurors are likely to overestimate the value of a 'match' between samples (President's Council of Advisors on Science and Technology, 2016).

In the first case, *R* v *Dallagher* [2002] EWCA Crim 1903, ear print evidence was admitted because two experts proffered it to be reliable, rigorous and of the highest scientific quality (Ireland & Beaumont, 2015). The judge stated that the expert testimony should be considered sufficient evidence for a conviction if the jurors accepted the testimony provided by the first of the two experts, who reportedly was absolutely convinced that the ear print found on a window at the crime scene matched the suspect's ear. A DNA analysis of material recovered from the ear print later demonstrated that this DNA could not have come from the suspect. As the conviction was largely decided on the basis of expert evidence from a new area of expertise, inconsistent with evidence from established sciences, it was concluded that the conviction had included over-focus on a single element of scientific evidence (Ireland & Beaumont, 2015). This high profile, controversial case gave rise to fundamental debates on the scientific basis and the nature of admissible opinion evidence and underscores that scientific validity must be assessed within the framework of a broader scientific discipline (Adam, 2016). The fact that an ear print examiner defends the validity of ear print examination means little and experience from previous casework cannot be considered informative because 'the right answer' is typically unknown as an examiner cannot accurately know how often he or she has erroneously declared matches (Executive Office of the President, 2016). Indeed, this case demonstrates the importance of establishing expertise through performance assessment as ear print pattern matching is relatively straightforward, quite probative and among the easiest to empirically test (Martire & Edmond, 2017). In the adapted case report created for this study, any other circumstantial or peripheral evidence was left out as the aim of the first case report was to establish the effect of an evidentiary instruction about central forensic evidence on verdict severity ratings returned by mock jurors.

The second case is based on *R* v *George (Barry)* [2007] EWCA Crim 2722, in which a single particle of gunpowder discharge residue found in the suspect's coat pocket a year after the crime was submitted as scientific evidence during the trial and claimed to link the defendant to the murder of Crimewatch presenter Jill Dando. The defendant was initially found guilty but was acquitted after seven years of imprisonment when it was successfully argued that the gunpowder residue could have been transferred from armed officers at the scene (Ireland & Beaumont, 2015). Although a first appeal had been dismissed, the CCRC ended up referring the conviction back to the Court of Appeal after commissioning a further report on the FDR evidence. The second (and successful) appeal revealed a fresh evaluation of the facts based on two competing propositions: Mr George is the man who shot Ms Dando vs. Mr George is not the man who shot Ms Dando. A group of experts formulated likelihood ratios and concluded that the value of the evidence was neutral and thus did not contribute to the legal debate (Adam, 2016). As this case demonstrates the importance of equipping jurors with the right means to assess the value of scientific evidence, it was selected for this article's mock jury experiment and presented to participants in a somewhat adapted format. In the adapted case report created

for this study, the circumstantial evidence described in the appeal judgment has been included to ensure that the second case is more complicated to assess.

Both cases drew attention to the risks associated with admitting forensic evidence in court without sufficient scrutiny. Although both verdicts were quashed on appeal, there is no reason to believe that miscarriages of justice due to the admission of unreliable forensic techniques will not occur again. In *Dallagher*, rather than excluding the ear print evidence the Court of Appeal ordered a new trial to enable the defence to challenge the expert opinion. The court quashed the conviction after the introduction of DNA evidence but remained satisfied that the expert evidence had been properly admitted (Roberts, 2008). Edmond and Roberts (2011) have since expressed concern that, as the Law Commission report did not assert unequivocally that *Dallagher* was mistaken and that ear print evidence should not be admissible, unreliable evidence could continue to be admitted in future trials. The Law Commission's report furthermore does not discuss existing research which suggests that lay people do not understand evidentiary expressions the way experts intent them to be understood (Edmond & Roberts, 2011). In appeals in which it is concluded that expert evidence should have been excluded the conviction may nevertheless be upheld if the court finds the verdict safe and the case persuasive based on the overall strength of the evidence (Edmond & Roberts, 2011). This is problematic as jurors may have misunderstood evidentiary expressions and would have acquitted if the expert evidence had not been admitted. Developing a case report based on *George* allowed for an examination of the impact of gunpowder residue evidence presented in conjunction with other circumstantial evidence. It moreover allowed for an exploration of the efficacy of educating jurors by expressing evidentiary probability in the form of a likelihood ratio. Finally, Ward et al. (2017) have expressed the concern that, although the *Crown Court Compendium* summarizes the law related to expert evidence and provides guidance on how to direct the jury, it does not provide a clear distinction between evidence with a scientific basis and evidence that is scientifically meaningless or subjective. If the jury is not adequately instructed that opinion evidence based on experience is not scientific and should be treated with caution, questionable evidence similar to the ear print evidence in *Dallagher* can remain influential (Ward et al., 2017).

Case 1 – Facts

- A woman was found murdered in her bed at home.
- The intruder used a jemmy or screwdriver to force open a small window above her bed. He suffocated her with a pillow after crawling through that window.
- Examination of the scene revealed ear prints on the glass of the window, which had been cleaned three or four weeks earlier.

The ear prints are examined by two experts who compare them with control prints provided by the suspect and others.

- The first expert is a police officer and lecturer at a police college.
    - He has no formal qualifications but has specialized in ear print comparison for over a decade and is convinced that no two ear prints are alike.
    - He has testified world-wide and has published a book on the topic.
- After examining the prints left at the scene of the crime, the expert reports being absolutely convinced that the prints of the suspect's ear are identical with the prints of the ear on the window.
    - The expert explains that he looks for five or six points when making a comparison.
    - He emphasises that what matters is the totality of the evidence, which he reviews by the use of overlays, choosing available control prints which appear to be set at an appropriate angle.
- The other expert furthermore concludes that it is very likely that the suspect made those prints.
- The judge directs you, the jury, that if you are sure that the evidence of the first expert (who reports being absolutely convinced) is correct, you are entitled to convict on the basis of this evidence alone.

Evidentiary Instruction

- Ear print identification cannot be regarded as generally accepted in the scientific community.
    - An article published in the Journal of Forensic Sciences has indicated that forensic scientists have reservations about the extent to which ear print evidence alone can safely be used to identify a suspect.
    - The validity of ear identification is unknown and the research necessary to say anything about the validity of ear identification has not yet been conducted.
- Neither the Forensic Science Service in the United Kingdom nor the Federal Bureau of Investigation in the United States carry out ear print comparisons.
- Two previous convictions based on ear identification evidence in other countries have been overturned on appeal.

Case 2 – Facts

- A woman was shot and killed as she was about to enter her home. Her death was caused by a single shot to the head.
- Almost a year later a suspect is arrested.
- Evidence identifies the suspect as being at the scene about four hours before the murder was committed.
- The prosecution characterises the suspect's interview as containing repeated lies.
- It is alleged that the suspect has attempted to create a false alibi for the time of the shooting.

Scientific evidence

- Among the findings by forensic scientists at the crime scene was firearm discharge residue in a bullet case and in the victim's hair.
- When searching the flat of the suspect almost a year after the crime, a coat is found which the suspect admits is his.
- The coat is subjected to forensic examination by a Senior Forensic Science Officer, who discovers a single particle (about one hundredth of a millimetre) of firearm discharge residue in the right pocket of the coat.
- The particle matches the essential elements of the firearm discharge residue found at the crime scene.
- Although the suspect admits owning the coat and being the only one who used it, the suspect cannot recall whether or not he was wearing it at the day of the crime.
- Two forensic scientists are called as witnesses in court:
    - The first expert states that, from experience, this type of residue would more often than not be found on the firer of the gun, but would not be found on ordinary members of the public unless they had been associated with firearms.
    - He states that it is most unlikely that the discharge residue finding is a result of innocent contamination.
    - Another well-qualified expert reviews these findings and agrees with them.

Evidentiary instruction

- The presence of a single particle in clothing gives no indication of how it got there.
- It is not possible to determine if the single particle is the last remainder of a prior association with firearms or whether it was deposited quite recently from a lightly contaminated source.
- The coat has been exposed to possibilities of innocent contamination during the police procedure.
- A significant number of particles may be more indicative of direct contamination than of secondary (innocent) contamination, but only one particle was found in this coat.
- The Forensic Science Service (FSS) reassesses the firearm discharge residue evidence that was provided at the suspect's trial and concludes that:
    - The particle found in the right pocket of the suspect's coat is indistinguishable from particles produced by the type of bullet used to shoot the victim, but a high proportion of bullets can produce such particles.
    - It would be just as likely that a single particle of discharge residue would have been recovered from the pocket of the suspect's coat whether or not he was the person who shot the victim.
    - The firearm discharge residue evidence is inconclusive. It provides no assistance to anyone asked to judge which proposition (*the suspect shot the victim* versus *the suspect did not shoot the victim*) is true.

Participants

The self-completion mock jury experiment, conducted through Bristol Online Survey which does not collect IP addresses, was distributed through convenience sampling on social media websites such as Facebook, LinkedIn, and Twitter. Furthermore, posters promoting the study were put up across various university buildings. Through this targeted sampling strategy, and in coherence with juror eligibility requirements, people from England and Wales between the age of 18 and 75 were considered eligible to participate in the study. As research has demonstrated that juror demographics such as gender, education, occupation and prior jury experience can influence a juror's evaluation of scientific evidence (Cramer et al., 2009), participants were asked for the following demographics: gender, age, highest obtained educational qualification, previous completion of a statistics or research methods course, previous jury duty, and previous or current employment in the field of criminal justice and/or law enforcement. Additionally, as both case reports were based on real miscarriages of justice, participants were asked whether they recognised either one or both of the cases. These questions were included at the end of the survey to ensure that participants could not use them to draw inferences about the aim of the survey.

Procedure

This mock jury experiment took approximately fifteen to twenty minutes to complete on a desktop or laptop. The design was selected for its ease of use, lack of spatial restrictions, time/cost-effectiveness and the potential for a high response rate. The latter enabled the potential acquisition of a representative cross-section of a wider population. No incentive was provided for participation in this study. In this classic experimental case-control design, the test group (Group 1) was exposed to an initial evidentiary instruction about the ear print evidence described in Case 1 whereas the control group (Group 2) did not receive such an instruction. In order to randomly allocate participants between the two groups, participants with last names beginning with the letters *A* to *K* (Group 1) were asked to click on a different link than participants with last names beginning with the letters *L* to *Z* (Group 2). This distinction was hypothesised to create two even groups based on an analysis of frequency distributions of last names in the United Kingdom, but the aforementioned requested demographics allowed for an extra control to check whether the groups were alike in every aspect other than the experimental manipulation. Furthermore, to establish that the design was indeed a randomised controlled trial, the verdicts and answers provided by both groups before the provision of the first evidentiary instruction could be compared.

In both surveys, the case facts were presented in the format of a written report and, before the presentation of each case, it was stated that to return a guilty verdict one had to be certain beyond a reasonable doubt. This specification, that the prosecution must prove its case against the defendant beyond a reasonable doubt, is referred to as the criminal standard of proof (Child et al., 2015). After reading the first case report, all participants were asked to return their verdict as follows: (1) Certainly Innocent, (2) Probably Innocent, (3) Probably Guilty and (4) Certainly Guilty. By granting four answers instead of a dichotomous verdict variable (i.e. Guilty or Not Guilty) it was possible to measure sophisticated changes in verdicts at later stages of the experiment without providing participants with an opportunity to remain neutral. An open question was included and so was a Likert scale for confidence: (1) Very Unconfident,

(2) Quite Unconfident, (3) Quite Confident and (4) Very Confident. Participants in Group 1 furthermore received an evidentiary instruction (see Case 1 – Evidentiary Instruction) and were asked for a reassessment of their verdict and confidence rating (*Figure 1*). When navigating through the survey participants could not return to previous pages to edit a response in retrospect.

All participants were subsequently presented with the second case report and were once more reminded that one must be certain beyond a reasonable doubt in order to return a guilty verdict. In this second case scenario, all participants were initially provided with peripheral evidence (suspect was at the scene four hours before the murder, the suspect's interview contained repeated lies, the suspect allegedly attempted to create false alibi) and asked for a verdict and an answer to an open question about the reasoning behind this verdict. This approach was adopted in order to tease apart the impact of the other circumstantial evidence and the – about to be provided - forensic evidence. After the provision of the scientific evidence, all participants were asked for a reassessment of their verdict and confidence rating. This was followed by an evidentiary instruction and a final request for verdicts and confidence levels. This design aimed to keep as many variables consistent across both test groups in order to control for potential confounding factors.

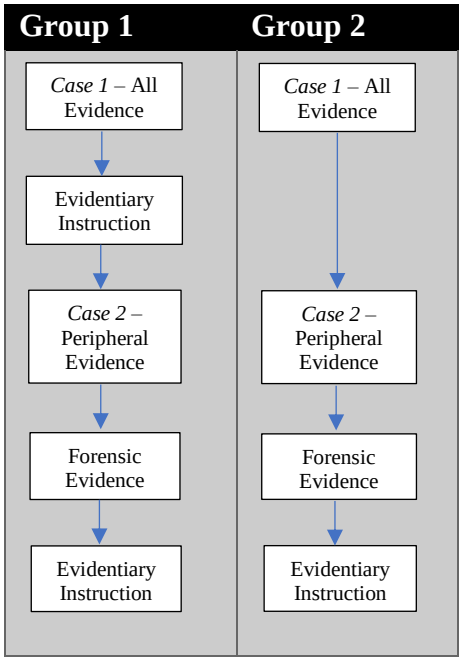

Figure 1 – Experimental Design

Coding and Scoring

The inclusion of open-ended questions allowed for an exploration of the full range of responses obtained as it did not constrain the respondents' opinions to predetermined categories (Sapsford & Jupp, 2006). After identifying multiple common themes, those themes were grouped together based on similarity to form six categories of open answers about the first case report and five categories of open answers about the second case report. Every category was given a code and every participant was scored as either *Yes (1)* if they mentioned that category in their answer or *No (2)* if they did not. This approach was adopted to take into account that some participants

addressed multiple themes or categories in their answer. Although the inclusion of open answers provides respondents with a sense that their responses are not constrained, and can thus help improve the naturalism of the method, the potential for bias introduced by the coder is considerable (Sapsford & Jupp, 2006). As the analysis of the open questions by the first author could indeed give rise to role conflict and biased interpretations, the categorization of the answers to the open questions was controlled by an impartial observer. In case of disagreement between the experimenter and the impartial observer, a final decision was made by the second author.

Ethical considerations

This research gained ethical approval from the independent Ethics Committee of the University of Portsmouth. Although anonymized and modified to fit a report format, the reports remain based on real court cases and potential participants were therefore provided with a disclaimer about distressing case information in the Participant Information Sheet. Participation was voluntary, anonymous and no incentives were provided. Participants received informed consent forms and were informed of their right to withdraw.

Results

Analytical Approach

Parametric tests are typically used to analyse continuous data whereas nonparametric tests are used to assess ordinal or ranked data such as data generated from Likert scales. However, research shows that, for studies with a large sample size, it can be appropriate to analyse data generated from Likert scales with parametric tests (Fagerland, 2012). Indeed, most parametric tests depend on a normal distribution of means and the means of studies with a sample size larger than 10 respondents per group are approximately normally distributed (Norman, 2010). If one were to use a parametric and a non-parametric test on the same data, the parametric test would have greater power to detect a true effect than the non-parametric test (Field, 2018). In other words, if there is a genuine effect a non-parametric tests is less likely to detect it than a parametric test (Field, 2018). To summarize, because treating Likert data with high response rates as ordinal prevents one from benefiting from the powerful and nuanced understanding parametric tests produce, restraining the analysis to non-parametric tests can lead to a loss of information (Carifio & Perla, 2008; Mircioiu & Atkinson, 2017). Non-parametric test equivalents were conducted nonetheless and indeed provided similar results.

Sample Characteristics

In Group 1, three participants were excluded for failing to answer either one or both of the open questions and two participants were excluded for indicating to have recognised either one or both of the cases. One respondent in Group 2 was excluded for responding to the open questions in Dutch and one respondent who indicated to have recognised either one or both of the cases was excluded. After exclusion of seven respondents, the final number of respondents was 174 (Group 1 = 100, Group 2 = 74). As can be seen in **Table 1**, the final sample consisted predominantly of higher educated, young, female participants.

Table 1

Frequency Table of Demographics.

		Group 1		Group 2		Totals	
		N	%	N	%	N	%
Gender	**Male**	21	**21.0**	18	**24.3**	39	**22.4**
	Female	78	**78.0**	55	**74.3**	133	**76.4**
	Other	1	**1.0**	1	**1.4**	2	**1.1**
Age	**18 – 24**	51	**51.0**	40	**54.1**	91	**52.3**
	25 – 34	16	**16.0**	14	**18.9**	30	**17.2**
	35 – 44	10	**10.0**	6	**8.1**	16	**9.2**
	45 – 54	13	**13.0**	5	**6.8**	18	**10.3**
	55 – 64	9	**9.0**	6	**8.1**	15	**8.6**
	65 – 75	1	**1.0**	3	**4.1**	4	**2.3**
Highest Obtained Qualification	**None**	0	**0.0**	1	**1.4**	1	**0.6**
	Lower secondary school	1	**1.0**	0	**0.0**	1	**0.6**
	Upper secondary school	20	**20.0**	10	**13.5**	30	**17.2**
	University/ college below degree	32	**32.0**	22	**29.7**	54	**31.0**
	University or college degree	47	**47.0**	41	**55.4**	88	**50.6**
Completion of Statistics Course	**Yes**	46	**46.0**	40	**54.1**	86	**49.4**
	No	54	**54.0**	34	**45.9**	88	**50.6**
Previous Jury Duty	**Yes**	6	**6.0**	0	**0.0**	6	**3.4**
	No	94	**94.0**	74	**100.0**	168	**96.6**
Employment in Criminal Justice or Law Enforcement	**Yes**	6	**6.0**	3	**4.1**	9	**5.2**
	No	94	**94.0**	71	**95.9**	165	**94.8**
	Totals	100	100.0	74	100.0	174	100.0

Impact of Age, Education, Statistical Knowledge and Gender on Verdict Severity

No significant differences in verdict severity ratings (Certainly Innocent, Probably Innocent, Probably Guilty, and Certainly Guilty) were established between various age categories, level of obtained qualification or completion of a statistics course. Research has suggested that female jurors may initially be tougher on defendants but are more susceptible to persuasion in deliberations than male jurors (Thomas, 2010). An independent-samples t-test conducted to examine gender differences in verdict severity for the initial verdicts about ear print evidence in Case 1 demonstrated no significant difference in verdicts between males (M = 2.69, SD = .73) and females (M = 2.92, SD = .59; t (53.39) = -1.76, p = .08, two-tailed).

A mixed between-within subjects analysis of variance was conducted to assess the impact of gender on verdict severity across three types of evidence (circumstantial evidence, gunpowder residue evidence, and an evidentiary instruction) in the second case (**Table 2**). The two participants who had selected 'Other/Prefer not to say' when asked about gender were excluded from the gender analysis. There was no significant interaction between gender and type of evidence, Wilk's Lambda = .99, F (2, 169) = .55, p = .58, partial eta squared = .01. The main effect when comparing male and female verdict severity was not significant, F (1, 170) = .22, p =.63, partial eta squared = .001 suggesting no gender differences in verdict severity across different types of evidence.

Table 2

Gender Comparison of Verdict Severity.

	Male			Female		
Type of Evidence	**n**	**M**	**SD**	**n**	**M**	**SD**
Circumstantial Evidence	39	2.33	.66	133	2.47	.64
Gunpowder Residue Evidence	39	2.90	.75	133	2.89	.73
Evidentiary Instruction	39	2.33	.77	133	2.35	.60

Case 1

Verdict Severity and Confidence

At this point in the experiment, both groups had read a report about the first case, in which ear print evidence was central to the case. As there had been no manipulation at this stage, no significant difference was expected in the independent samples t-test. Indeed, there was no significant difference in verdict severity between Group 1 (M = 2.79, SD = .70) and Group 2 (M = 2.95, SD = .55). Furthermore, there was no significant difference in reported confidence levels between Group 1 (M = 2.87, SD = .76) and Group 2 (M = 2.81, SD = .72) either.

Open Answers

The open question posed after the first verdict encouraged participants to elaborate on the reasoning behind their verdict. The ear print evidence was presented in a manner which made it central to the case. In other words, it would be nearly impossible to return a verdict without actively assessing the value of this evidence because any other circumstantial evidence had purposely been excluded from the report. As the manipulation for Group 1 (an evidentiary instruction) was provided afterwards, one would expect the initial open answers to be relatively comparable amongst both groups. The open answers of both groups were relatively similar

(**Table 3**). However, it seems that in Group 2 a higher percentage of respondents discussed the first expert's testimony in their elaboration, whereas the comment that there is not enough evidence to establish guilt beyond a reasonable doubt was mentioned less. As the aforementioned independent samples t-test demonstrated no significant difference between the verdicts of both groups, these findings are hard to explain. It is possible that the respondents mentioned the evidence because it was so central to the case report but did not let it sway them towards a significantly different verdict. If the participants in Group 2 had considered the evidence more convincing this should have been reflected in a significantly harsher verdict, which was not the case.

Table 3

Frequency Table of Coded Open Answers Case 1.

	Group 1		**Group 2**		**Total**	
	N	**%**	N	**%**	N	**%**
Based on the first expert's testimony	43	**33.3**	40	**43.9**	83	**37.7**
Not enough evidence for guilty beyond reasonable doubt	38	**29.5**	15	**16.5**	53	**24.1**
Based on the testimony of both experts combined	17	**13.2**	12	**13.2**	29	**13.2**
The expert lacks qualifications and/or ear print evidence is not reliable	15	**11.6**	12	**13.2**	27	**12.3**
Ear print on the window could have happened another time and/or someone else could be involved	12	**9.3**	8	**8.8**	20	**9.1**
Other	4	**3.1**	4	**4.4**	8	**3.6**
Totals	129	**100.0**	91	**100.0**	220	**100.0**

Impact of Evidentiary Instruction

A paired-samples t-test, used when collecting data from one group of participants on two different occasions, was conducted to evaluate the impact of an evidentiary instruction on the verdict severity ratings of participants in Group 1. There was a statistically significant decrease in verdict severity from $M = 2.79$, $SD = .70$ after the initial evidence to $M = 2.28$, $SD = .67$ after the provision of an evidentiary instruction, $t\ (99) = 8.58$, $p < .001$ (two-tailed). The mean decrease in the verdict severity score was .51 with a 95% confidence interval ranging from .392 to .628. The eta squared statistic (.43) indicated a large effect size. The guidelines for interpreting this value are .01=small effect, .06=moderate effect, .14=large effect (Pallant, 2013). In other words, after reading the evidentiary instruction, participants in Group 1 returned significantly less punitive verdicts.

Case 2

Open Answers

Participants were asked to respond to a second open question after the provision of the circumstantial evidence of Case 2 (**Table 4**). It was decided not to provide participants with both the circumstantial evidence and the central gunpowder residue evidence at once in order to remain capable of teasing apart exactly how much this gunpowder evidence impacted upon the verdict. It was hypothesised that gunpowder evidence, for it is perceived to be more scientific, should have a greater impact on juror verdicts than circumstantial evidence. Most respondents indeed considered the circumstantial evidence alone insufficient to establish guilt beyond a reasonable doubt, and the responses of Group 1 and Group 2 were comparable.

Table 4

Frequency Table of Coded Open Answers Case 2.

	Group 1		**Group 2**		**Total**	
	N	**%**	N	**%**	N	**%**
Not enough evidence for guilty beyond a reasonable doubt	58	**48.3**	47	**50.5**	105	**49.3**
Based on the available evidence	39	**32.5**	29	**31.2**	68	**31.9**
It has been a year since the crime	13	**10.8**	10	**10.8**	23	**10.8**
He may have lied for another reason	6	**5.0**	6	**6.4**	12	**5.6**
Other	4	**3.3**	1	**1.1**	5	**2.3**
Totals	120	**100.0**	93	**100.0**	213	**100.0**

Verdict Severity and Confidence

A one-way repeated measures ANOVA (which tests for significant differences in the mean of a dependant variable across two or more conditions) was conducted to compare the returned verdict ratings of all participants for Case 2 after the provision of circumstantial evidence, gunpowder residue evidence and an evidentiary instruction. The means and standard deviations are presented in **Table 5**. There was a significant effect, Wilks' Lambda = .60, F (2, 172) = 56.45, $p < .001$, multivariate partial eta squared = .39 which indicates a large effect size. In other words, there was a significant change in verdict severity ratings across the three different types of evidence.

Table 5

Descriptive Statistics of Verdict Severity after Circumstantial Evidence, Gunpowder Residue Evidence and Evidentiary Instructions.

Time Period	**N**	**Mean**	**Standard Deviation**
Circumstantial Evidence	174	2.44	.66
Gunpowder Residue Evidence	174	2.90	.75
Evidentiary Instruction	174	2.35	.64

A one-way repeated measures ANOVA was conducted to compare reported confidence ratings after the return of verdict ratings about circumstantial evidence, gunpowder residue evidence and the evidentiary instruction. The means and standard deviations are presented in **Table 6**. There was a significant effect for confidence level, Wilks' Lambda = .93, F = (2, 172) = 6.32, p = .002, multivariate partial eta squared = .068.

Table 6
Descriptive Statistics Confidence after Verdict Circumstantial Evidence, Gunpowder Residue Evidence and Evidentiary Instruction.

Verdict Confidence	N	Mean	Standard Deviation
After Circumstantial Evidence	174	2.75	.78
After Gunpowder Residue Evidence	174	2.84	.74
After Evidentiary Instruction	174	2.64	.79

Scepticism Effect

An independent samples t-test demonstrated no significant difference in verdict severity between Group 1 (M = 2.36, SD = .64) and Group 2 (M = 2.55, SD = .67) after provision of the circumstantial evidence of Case 2. This could indicate that the assessment of the previous evidentiary instruction about ear print evidence by Group 1 did not cause a generalized scepticism effect which extended to circumstantial evidence. Furthermore, no significant difference between the reported confidence ratings of Group 1 (M = 2.81, SD = .76) and Group 2 (M = 2.89, SD = .71) was found.

An independent-samples t-test was conducted to compare the verdict severity scores of both groups after the provision of gunpowder residue evidence. There was a significant difference in scores for Group 1 (M = 2.79, SD = .73) and Group 2 (M = 3.05, SD = .72; t (172) = -2.38, p = .02, two-tailed). However, the magnitude of the differences in the means (mean difference = -.26, 95% *CI*: -.48 to -.04) was small (eta squared = .032). An independent-samples t-test was conducted to compare the verdict severity scores of Group 1 and Group 2 after the provision of an evidentiary instruction about the gunpowder residue evidence. There was a significant difference in scores for Group 1 (M = 2.23, SD = .66) and Group 2 (M = 2.51, SD = .58; t (172) = -2.94, p = .004, two-tailed. The magnitude of the differences in the means (mean difference = -.28, 95% *CI*: -.47 to -.09) was quite small (eta squared = .048).

A mixed between-within subjects analysis of variance was conducted to assess the impact of (not) receiving an evidentiary instruction in Case 1 on participants' subsequent verdicts severity ratings across three types of evidence presented in Case 2 (circumstantial evidence, gunpowder residue evidence and evidentiary instruction) (**Table 7**). This analysis, at times referred to as a split-plot ANOVA, tests both whether there are main effects for the two independent variables and whether the interaction between the two variables is significant (Pallant, 2013). In this case the between-subjects variable was *Group* (Group 1 received a previous evidentiary instruction while Group 2 did not), the within-subjects variable was *Type of Evidence* (circumstantial evidence, gunpowder evidence, evidentiary instruction) and the continuous dependent variable was *Verdict Severity*. There was no significant interaction between group and verdict severity, Wilks' Lambda = .996, F (2, 171) = .354, p = .702, partial

eta squared = .004. Perhaps counterintuitively this non-significant interaction is favourable as it indicates that the changes in verdict severity across three types of evidence are not the same for both groups. There was a substantial main effect for type of evidence, Wilks' Lambda = .610, *F* (2, 171) = 54.730, p < .001, partial eta squared .390, with both groups showing an increased verdict severity rating after the gunpowder residue evidence and a subsequent reduced verdict severity rating after the evidentiary instruction (*Figure 2*). When comparing the two groups the main effect was significant, *F* (1, 172) = 9.874, p = .002, partial eta squared = .054, suggesting a difference between the verdict severity ratings of Group 1, who received a previous evidentiary instruction, and Group 2 who did not receive such an instruction.

Table 7
Group Comparison of Verdict Severity.

	Group 1			**Group 2**		
Evidence	**n**	**M**	**SD**	**n**	**M**	**SD**
Circumstantial Evidence	100	2.36	.64	74	2.55	.67
Gunpowder Residue Evidence	100	2.79	.73	74	3.05	.72
Evidentiary Instruction	100	2.23	.66	74	2.51	.58

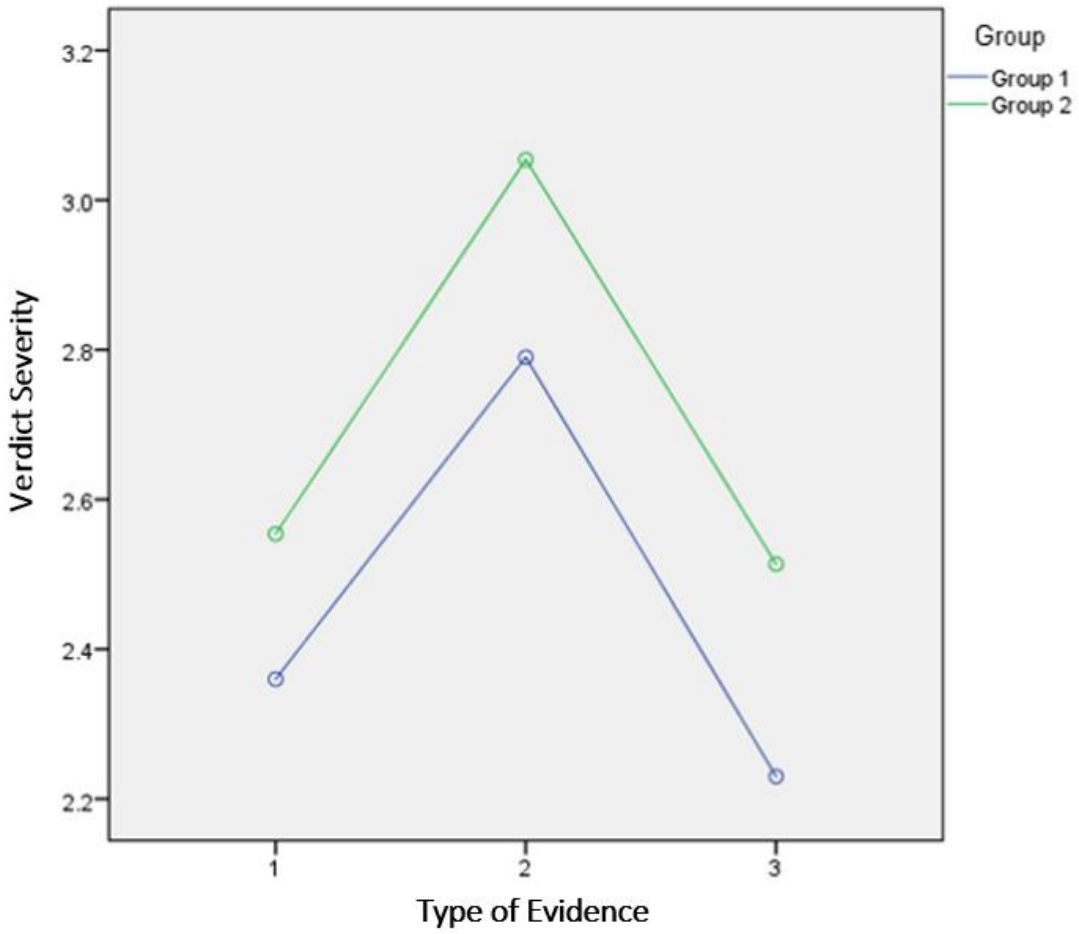

Figure 2 - Group Comparison of Verdict Severity after (1) Circumstantial Evidence, (2) Gunpowder Residue Evidence, and (3) Evidentiary Instruction.

Verdict - Central vs. Peripheral Evidence

As can be seen in *Figure 3*, both groups were less likely to return a guilty verdict based on circumstantial evidence alone compared to either scientific evidence alone or a scenario with both circumstantial evidence and scientific evidence.

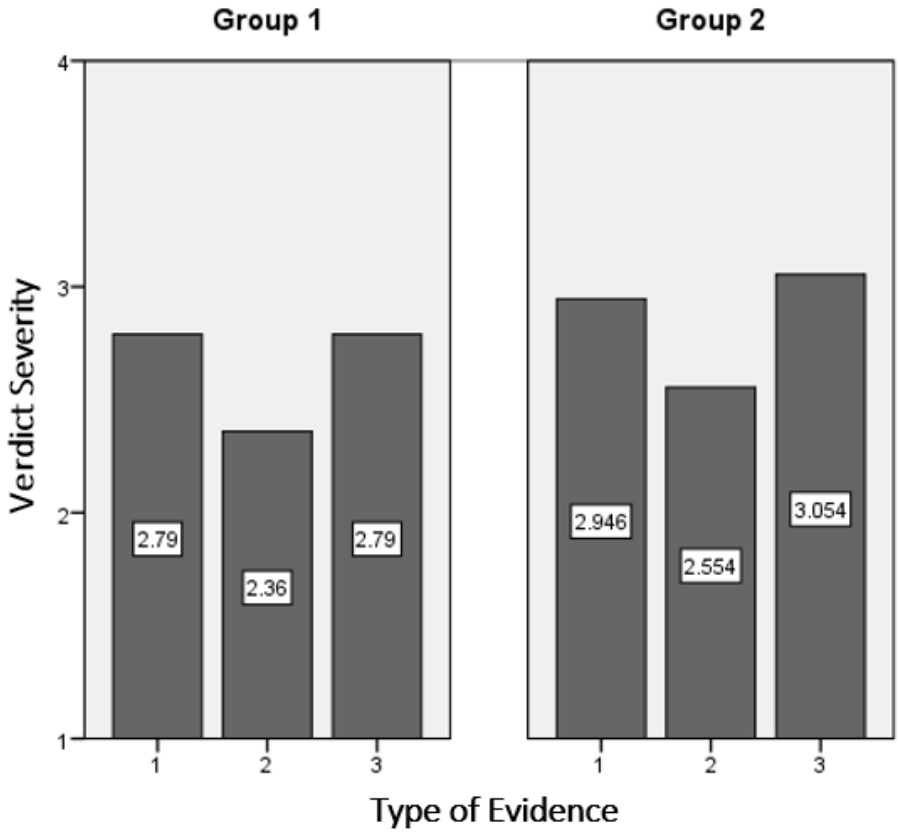

Figure 3 - Group Comparison of Verdict Severity after Scientific Evidence Case 1 (1), Circumstantial Evidence Case 2 (2), and Circumstantial + Scientific Evidence Case 2 (3).

Discussion

Main findings

This study demonstrates that both ear print and gunpowder residue evidence were considered convincing indicators of guilt regardless of the fact that neither type of forensic evidence provided direct or irrefutable proof of guilt. Previous research has indicated that jurors are inclined to trust expert witnesses. This was confirmed in Case 1, in which both the test and control group were predominantly willing to return a guilty verdict and mentioned the first expert's testimony as the main reason for this decision. Remarkably, jurors who reasoned that not enough evidence had been provided to establish guilt beyond a reasonable doubt did not always return an innocent verdict. A 'Probably Guilty' option which could be used if jurors were not convinced beyond a reasonable doubt, but nevertheless felt inclined to return a guilty verdict, may explain this finding. After the test group was provided with an evidentiary instruction, a significant decrease in verdict severity was observed in which mock jurors returned a 'Probably Innocent' instead of a 'Probably Guilty' verdict. This may indicate that

the evidentiary instruction successfully educated jurors about the questionable reliability of the ear print evidence.

For Case 2 it was hypothesised that the initial non-scientific circumstantial evidence would be considered insufficient to establish guilt beyond a reasonable doubt but that mock jurors would become swayed by the scientific aura of the gunpowder residue evidence provided afterwards. It was found that jurors attributed low levels of weight to non-scientific circumstantial evidence as the initial peripheral evidence was predominantly considered insufficient to return a guilty verdict. Indeed, the general theme in the open answers of both groups, before the provision of the gunpowder residue evidence, was that there was not enough evidence to establish guilt beyond a reasonable doubt. The content of the open answers indicated that jurors were both capable and willing to engage in systematic processing of the evidence. When comparing verdict severity scores across all types of evidence presented in Case 2 (circumstantial evidence, gunpowder residue evidence, and the evidentiary instruction) it was found that verdict severity increased significantly after the provision of gunpowder residue evidence. This could indicate that scientific evidence is considered a more persuasive indicator of guilt than other circumstantial evidence. The provision of an evidentiary instruction about the questionable reliability of gunpowder residue evidence led to a significant decrease in verdict severity, indicating that the instruction may have successfully educated jurors about the questionable reliability or probative value of that evidence. To summarize, the evidentiary instruction seemed to encourage a more cautious appraisal of the scientific evidence which resulted in a more lenient verdict. It furthermore impacted jurors' confidence levels as most jurors reported feeling less confident about the returned verdict.

There was no significant difference in verdict severity between the test and control group after the provision of the initial circumstantial evidence of Case 2. This could indicate that mock jurors were well aware of the fact that this peripheral evidence was not scientific, and that the initial evidentiary instruction about the ear print evidence thus did not cause a generalised scepticism effect. After the subsequent provision of gunpowder residue evidence, a significant difference in verdict severity could be observed as Group 1 became more lenient in their verdicts than Group 2. In other words, it seemed that the initial evidentiary instruction did not give rise to a scepticism effect in the assessment of general circumstantial evidence but did trigger a scepticism effect for new unrelated scientific evidence. This significant difference was also observed after the provision of the second evidentiary instruction, as Group 1 remained more lenient than Group 2 and predominantly returned a 'Probably Innocent' verdict.

Practical Implications

The present study underscores that the provision of an evidentiary instruction has the potential to educate jurors about the limitations of scientific evidence. This could be good news because an evidentiary instruction can be implemented with relative ease and has been implemented before. As aforementioned, in England and Wales judges provide guidance on the assessment of eyewitness identification evidence through the Turnbull directions. Surprisingly, there has been little empirical research into the efficacy of these directions, and it is thus unknown whether they successfully educate jurors about the reliability of eyewitness identification evidence. Moreover, the directions focus on event characteristics but say little about the procedural causes of false identification. Although the PACE Codes mandate many aspects of

identification procedures and, for example, state that care must be taken not to direct the witness' attention to any one individual, aspects such as the exact wording of instructions given to witnesses prior to the procedure are not included (Horry et al., 2013; Home Office, 2017). Therefore, an eyewitness who had decent viewing conditions but a biased procedure would most likely be considered reliable. In other words, even if the Turnbull direction is effective, it is not ideal and a revision would be recommendable.

An important question relates to who would create, revise, and control the delivery of evidentiary instructions. The adversarial system assumes active fact-finding by defence and prosecution, but the reality is that the defence often does not employ their own expert. It is therefore desirable that the creation of evidentiary instructions is controlled by an independent party, in cooperation with the scientific community, and that the judge is enabled to enforce their implementation. Statistical experts have suggested that a summary of expert reports containing statistical or probabilistic evidence (e.g. likelihood ratios) should be made publicly available, a suggestion which could be extended to expert reports about all types of scientific evidence (Buchanan, 2007). In the USA, the President's Council of Advisors on Science and Technology (2016) has recommended that an independent scientific body should evaluate forensic science methodologies and explain their strengths and limitations with regards to their capability to provide accurate and reliable answers to specific and well-defined forensic questions. To ensure that scientific judgements are independent and unbiased, such evaluations must be conducted by a science agency with no stake in the outcome (Charlton, 2013).

The aforementioned scientific guides are a good start, but scientists are expected to volunteer their time to contribute to such guides. What has been disregarded in this discussion but may be just as relevant, is what Charlton (2013) refers to as a 'standard of expediency'. The reality is that improvements can only be introduced within certain fiscal boundaries, and financial constraints may ultimately determine the level and degree to which changes can be implemented. The standard of expediency is thus a balance between what is scientifically desired and what can be afforded (Charlton, 2013). It would be futile to develop technologies that increase juror understanding and minimise the effects of junk science if the cost of those tools is too prohibitive (Charlton, 2013). The question is whether the intrinsic risks involved in achieving a financial compromise are acceptable when considering the grave potential consequences of error in the criminal justice system (Charlton, 2013; Dodd, 2017).

Although it would be ideal to have a validation study addressing the reliability and consistency of a particular procedure or analysis for the precise conditions of each case, it is unlikely that resources would be made available for such testing and it is furthermore unlikely that case conditions can be faithfully reproduced (Martire & Edmond, 2017). This does not have to be incapacitating for the analysis of scientific expertise, as long as courts have access to validity data from other experts who have applied that particular methodology and can identify conditions under which performance is better or worse as well as whether those experts performed the task better than untrained people (Martire & Edmond, 2017). As aforementioned, evidentiary instructions could be more cost-effective than the provision of opposing expert testimony and may eventually reduce the number of costly appeal procedures for potential miscarriages of justice. In cases involving pattern matching sciences, the issue is not necessarily whether a procedure does or does not work. It is rather how well it works, in what conditions, and how the expert should express their opinion to accurately capture the

value of the evidence in a manner that facilitates juror comprehension (Martire & Edmond, 2017). Shapiro et al. (2015) found that, in cases with psychological expert testimony, the issues raised in court often centred on the weight the testimony should be given in the decision rather than the general admissibility. Indeed, in some circumstances judges may only need to moderate the strength of the expert's claim (Martire & Edmond, 2017). As the legal system is ill-equipped to correct the issues that come with the introduction of scientific evidence, the scientific community itself should be a crucial source for research and reforms (Garrett & Neufeld, 2009). To further enhance the contribution of expert witnesses and scientific evidence in court it is important to stimulate a closer partnership between academics and the judiciary, in which key stakeholders can work together to seek acceptable improvements so that research can contribute to a fair and accurate administration of justice (Dror et al., 2013).

Limitations

Although mock jury studies are the prevalent methodology for empirical research on juror comprehension of scientific evidence, such studies are often criticised for a lack of external validity (Lieberman et al., 2016). Written case summaries may not adequately simulate the actual courtroom experience (McAuliff & Duckworth, 2010). Nevertheless, comparisons of decision making by mock jurors and real jurors have failed to uncover major consistent differences as both types of jurors seem to reason in similar ways (Sklansky, 2013). Assessment of real case transcripts reveals that the information required for a rational evaluation of forensic evidence is not available to judges or jurors in the vast majority of cases (Edmond, 2015a). This observed lack of relevant information underlines a fundamental issue with observational studies and exit surveys, which is that it is of questionable value to study claims about comprehension made by decision-makers who were not provided with the necessary information required to assess the evidence to begin with (Edmond, 2015a). Therefore, as long as a mock jury study does not claim to replicate the actual courtroom experience and the severe consequences that come with a wrongful conviction, its findings can still have merit as such findings can provide an insight into juror reasoning, how this reasoning can be optimised, and can demonstrate the importance of regulations against the admittance of junk science in court.

The sample of this study consisted predominantly of higher educated young women and the findings may therefore not extend to real juries with more diverse demographics. Moreover, as jurors were assessing old court cases, they must have realised that their verdicts would not directly impact the lives of victims, suspects, and perpetrators. Participants also completed the survey in isolation, and the group dynamics of jury deliberation were thus not accounted for. This could be a disadvantage, although Devine et al. (2016) found that jury deliberation seems to have little impact on the final verdict. It was decided to adopt a Likert scale approach to assessing innocence in order to be able to examine more sophisticated verdict changes. Jurors in real courtrooms do not have this luxury as they must decide dichotomously between guilty or not guilty. However, as jurors in real trials undoubtedly experience uncertainty before reaching a dichotomous verdict decision as well, it would be interesting to assess whether mock jurors who returned probably innocent or probably guilty verdicts are more easily convinced to reassess that verdict during the deliberation than jurors who returned certainly innocent or certainly guilty verdicts.

Jurors in experimental settings may be more sensitive to evidence quality than real jurors because the provision of a written stimulus allows them to control the rate at- and the degree to which they process the evidence (McAuliff & Duckworth, 2010). A written report format nevertheless became the preferred choice of assessment as studies examining decision making differences between videotaped and written expert testimony have uncovered no effect on trial-related decisions (McAuliff et al., 2009). The reports presented in this study were derived from Court of Appeal judgments and designed to reflect the content of those trial summaries as accurately as possible. This undeniably remains a relatively impoverished stimulus, but this research aimed to examine central or systematic processing which is encouraged by ensuring that the decision maker is capable of meticulously analysing and evaluating the quality of the evidence. The report format therefore allowed for an attempt to avoid peripheral or heuristic processing by factoring out source-related cues such as likeability, displayed confidence level, and attractiveness (McAuliff et al., 2009). One could argue that, if justice ought to be blind, a study which factors such biases can stimulate a more factual assessment of the evidence. Nevertheless, it is undeniable that simulated case decisions do not carry the real-world consequences of an actual trial, and this realisation may impact the jurors' motivation and willingness to engage in the systematic processing of the evidence (McAuliff & Duckworth, 2010). To encourage systematic processing as much as possible, the length of the case reports and the duration of the entire experiment were kept relatively short.

Recommendations for further research

Trial safeguards do not consistently expose the potential weaknesses of scientific evidence (Martire & Edmond, 2017). Therefore, research aimed at evaluating evidentiary safeguards is greatly needed to better accommodate jurors' reasoning skills in trials with scientific evidence (McAuliff et al., 2009). Although the Turnbull directions are, as of yet, a rare example of the successful implementation of an evidentiary instruction in England and Wales, there is remarkably little knowledge about the effect of the directions on juror assessment of eyewitness identification evidence as this has not been evaluated in an experimental setting. Moreover, as the directions only inform jurors about witnessing conditions as opposed to procedural conditions, jurors could consider evidence obtained in a procedure with a biased officer reliable as long as the witnessing conditions were good. It is thus recommended that future research examines the efficacy of the directions both in its current state and in a revised format. The impact of the provision of evidentiary reports signed by multiple experts could also be examined as this requirement could serve as a gatekeeper for academic consensus.

A downside of evidentiary instructions can be found in their potential to elicit unwarranted scepticism in the assessment of scientific evidence. Further research should thus examine whether the impact of an evidentiary instruction depends on the timing of its provision. In other words, is there a difference in verdict severity if the instruction is delivered (a) before the scientific evidence, (b) immediately after the scientific evidence, or (c) before the jury deliberation? With regard to group deliberations, which were unaccounted for in this study, it would be interesting to examine whether more qualified or educated members of the jury with experience in a profession related to the science at stake could influence the ultimate verdict of the rest of the jurors.

This study contained case summaries which were presented in a comprehensible report format in order to encourage systematic processing. However, real cases will involve more complex scientific evidence and expert testimony. It could thus be examined whether jurors are more willing to engage the systematic processing of evidence which is easier to understand or which is recognizable from, for example, crime television shows. Recent research has focused on neuroscientific evidence as neuroscience provides the law with opportunities and risks at the same time. The number of publications in the field of neuro-law has risen from approximately 100 publications in 2003 to nearly 1,200 publications by 2013 (Jones et al., 2013). An increased insight into the workings of the brain can contribute to an understanding of a suspect's behaviour, but academics fear that judges and jurors may fail to comprehend this type of evidence as a result of its complex and technical nature (de Kogel et al., 2013; Jones et al., 2013; Leonard, 2015). Research has indeed demonstrated that jurors are easily convinced by neuroscientific evidence, even if this is not merited (de Kogel et al., 2013). With numerous scientific breakthroughs on the horizon, this is reason for serious concern. As the use of neuroscientific evidence is on the rise but under-investigated, it could be examined whether jurors can be successfully educated on the relevance and reliability of such complex evidence.

Conclusion

Admissibility of forensic evidence has become a difficult concept in the law of criminal evidence (Charlton, 2013). Scientific conclusions are often expressed in probabilities and subject to continuous revision whereas the criminal law system adheres to a more dichotomous distinction of guilt versus innocence and expects disputes to be resolved swiftly (Ainsworth, 2001; Ormerod & Sturman, 2005). As science is a fluid concept and new techniques are constantly evolving it should not be perceived to hold definitive answers (Ireland & Beaumont, 2015). This observation may identify the core issue at stake, which is that guilt in the courtroom must be established beyond reasonable doubt while a scientist can never be entirely certain of findings (Canter & Alison, 1999).

Conventional admissibility criteria regarding, for example, formal qualification and training, experience, and assistance to the jury do not provide an adequate insight into the validity or limitations of forensic evidence. It is concerning that the party relying upon forensic evidence is not required to demonstrate that the underlying technique is reliable or that the expert is proficient in its use (Edmond, 2015a). Whether a defendant is convicted could depend on which expert examines the evidence as experts can vary in performance due to different training, subjective decision thresholds, risk tolerance and even eyesight, (Dror & Murrie, 2018). This is especially concerning when considering that forensic evidence is rarely contested in court. The current admissibility jurisprudence and legal practices, in particular in attempts to regulate feature comparison evidence, are misguided when compared to criteria promoted by scientific organisations (Edmond, 2015b). Moreover, recent developments reflect a greater concern for finality and efficiency with a reduced concern for establishing the truth (Anderson et al., 2005; Tully, 2018). A continuing focus on cost-reduction, which affects both commercial and government-funded forensic science practitioners, is eroding the potential for professional development and has left experts exposed to critique about failing to keep current on scientific developments and to provide the court with accurate information regarding the range of scientific opinions within a field (Tully, 2018).

To effectuate a proper functioning of the judicial decision-making process it is important to improve juror understanding of scientific evidence. This mock jury experiment demonstrated that feature comparison evidence is highly persuasive as it swayed mock jurors towards guilty verdicts despite the fact that neither the ear print nor the gunpowder residue evidence were direct indicators of guilt. Moreover, feature comparison evidence was considered convincing proof of guilt both when presented on its own and when presented in conjunction with other circumstantial evidence. The introduction of an evidentiary instruction successfully educated mock jurors about the limitations of such evidence and stimulated a verdict reassessment. The hypothesised scepticism effect applied to feature comparison evidence only insofar as no significant difference was observed between test and control group verdicts regarding the circumstantial evidence presented in the second case. One could argue that the observed scepticism effect may be desirable as compensation for the observation that jurors tend to overestimate the value of a match between samples. Indeed, if one were to agree with Blackstone's argument that it is better to have ten guilty persons escape than one innocent suffer, inducing scepticism about scientific evidence introduced in court is not necessarily undesirable for it may reduce false positives (Papailiou et al., 2014).

If properly executed and blind to contextual information, scientific evidence has the potential to enhance the quality of fact-finding. However, it should not be perceived to hold definitive answers and scientists should acknowledge potential exposure to biased information and ensure not to overstate the value of their findings. In the absence of the equality of arms upon which adversarial fact-finding is thought to depend, the introduction of evidentiary instructions can stimulate a more adequate evaluation of feature comparison evidence. Informative protocols which stimulate the critical assessment and adequate delivery of scientific evidence in court must be developed to improve juror comprehension of the reliability and potential limitations of scientific evidence. As judges and attorneys have expressed discomfort with the task of reviewing the validity of scientific evidence, an independent body should be responsible for the development of evidentiary instructions about various types of scientific evidence. The accruement of an extensive evidence base is desirable to successfully provoke the implementation of such instructions. Therefore, further research into the effective implementation of safeguards against the admittance of unreliable science in court is direly needed if we truly wish to improve the criminal justice system and prevent miscarriages of justice. Although mock jury study designs come with limitations, this type of research remains relevant as it provides insight into potential ways in which such objectives can be achieved.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Adam C. *Forensic Evidence in Court: Evaluation and Scientific Opinion*. Chichester: John Wiley & Sons, 2016.

Ainsworth P. *Offender profiling and crime analysis*. Devon: Willan Publishing, 2001.

Anderson T, Schum D, Twining W. *Analysis of evidence.* Cambridge: Cambridge University Press, 2005.

Baguley C, McKimmie B and Masser B. Deconstructing the simplification of jury instructions: How simplifying the features of complexity affects jurors' application of instructions. *Law and Human Behavior* 2017; 41(3): 284-304.

Boudreau C and McCubbins M. Competition in the Courtroom: When Does Expert Testimony Improve Jurors' Decisions? *Journal of Empirical Legal Studies* 2009; 6(4): 793-817.

Brants C and Field S. Truth-finding, procedural traditions and cultural trust in the Netherlands and England and Wales: When strengths become weaknesses. *The International Journal of Evidence & Proof* 2016; 20(4): 266-288.

Bromby M, MacMillan M and McKellar P. An Examination of Criminal Jury Directions in Relation to Eyewitness Identification in Commonwealth Jurisdictions. *Common Law World Review* 2007; 36(4): 303-336.

Benforado A. *Unfair: the new science of criminal injustice*. 1st ed. New York: Crown Publishers, 2015.

Buchanan M. Statistics: Conviction by numbers. *Nature* 2007; 445(7125): 254-255.

Canter D and Alison L. *Profiling in Policy and Practice*. Offender Profiling Series, Volume 2. Aldershot: Ashgate Publishing, 1999.

Carifio J and Perla R. Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 2008; 42(12): 1150-1152.

Charlton D. Standards to avoid bias in fingerprint examination? Are such standards doomed to be based on fiscal expediency? *Journal of Applied Research in Memory and Cognition* 2013; 2(1): 71-72.

Cicchini MD and White L. Truth or doubt? An empirical test of criminal jury instructions. *University of Richmond Law Review* 2016; 50: 1139-1167

Child J, Ormerod DC and Smith JC. *Smith & Hogan's essentials of criminal law*. Oxford: Oxford University Press, 2015.

Cooper J and Hall J (2000). Reaction of mock jurors to testimony of a court appointed expert. *Behavioral Sciences & The Law* 2000; 18(6): 719-729.

Cramer RJ, Brodsky SL and DeCoster J. Expert witness confidence and juror personality: Their impact on credibility and persuasion in the courtroom. *Journal of the American Academy of Psychiatry and the Law Online* 2009; 37(1), 63-74.

Cutler B, Penrod S and Dexter H. Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior* 1990; 14(2): 185-191.

De Keijser J, Elffers H. Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. *Psychology, Crime & Law* 2012; 18(2): 191-207.

De Kogel K, Van De Beek P, Leeuw F, et al. Themanummer 'Neurolaw in Nederland' De betekenis van de neurowetenschappen voor het recht. *Nederlands Juristenblad* 2013; 88: 3129-3161.

Devine DJ, Krouse PC, Cavanaugh CM, et al. Evidentiary, extraevidentiary, and deliberation process predictors of real jury verdicts. *Law and Human Behavior* 2016; 40(6): 670-682.

Diamond SS. How jurors deal with expert testimony and how judges can help. *Journal of Law and Policy* 2007; 16(1): 47-68.

Dillon MK, Jones AM, Bergold AN, Hui CYT & Penrod SD. Henderson instructions: do they enhance evidence evaluation? *Journal of Forensic Psychology Research and Practice* 2017; 17(1): 1-24

Dodd V. Forensic science cuts pose risk to justice, regulator warns. *The Guardian*, https://www.theguardian.com/science/2017/jan/06/forensic-science-cuts-pose-risk-justice-regulator-warns (2017, accessed 10 January 2018).

Dror I, Kassin S and Kukucka J. New application of psychology to law: Improving forensic evidence and expert witness contributions. *Journal of Applied Research in Memory and Cognition.* 2013; 2(1): 78-81.

Dror I, Morgan R, Rando C, et al. The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making. *Journal of Forensic Sciences* 2017; 62(3): 832-833.

Dror, I and Murrie, D. A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law* 2018; *24*(1): 11-23.

Dunne D. Re-assessing Trial by Jury in Early Modern Law and Literature. *Literature Compass* 2015; 12(10): 517-526.

Eastwood J and Caldwell J. Educating Jurors about Forensic Evidence: Using an Expert Witness and Judicial Instructions to Mitigate the Impact of Invalid Forensic Science Testimony. *Journal of Forensic Sciences* 2015; 60(6): 1523-1528.

Ebisike N. *Offender Profiling in the Courtroom: The Use and Abuse of Expert Witness Testimony.* Westport: Greenwood Publishing Group, 2008.

Edmond, G. Forensic science evidence and the conditions for rational (jury) evaluation. *Melbourne University Law Review* 2015a; 39(1): 77-127.

Edmond, G. Legal versus non-legal approaches to forensic science evidence. *The International Journal of Evidence & Proof* 2015b; 20(1): 3-28.

Edmond, G. and Roberts, A. The law commission's report on expert evidence in criminal proceedings. *Criminal Law Review* 2011; 844-862.

Ellison L. and Munro V. Telling tales: exploring narratives of life and law within the (mock) jury room. *Legal Studies* 2015; 35(2): 201-225.

Fagerland MW. T-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Medical Research Methodology* 2012; 12(1): 78-84.

Field A. *Discovering statistics using IBM SPSS*. 5th ed. London: SAGE Publications; 2018.

Benforado A. *Unfair: the new science of criminal injustice*. 1st ed. New York: Crown Publishers, 2015.

Freckelton I. Admissibility of Expert Opinions on Eyewitness Evidence: International Perspectives. *Psychiatry, Psychology and Law* 2014; 21(6): 821-836.

Garrett BL and Neufeld PJ. Invalid forensic science testimony and wrongful convictions. *Virginia Law Review* 2009; 95(1): 1-97

Ghosh P. UK judges to get scientific guides. *BBC News,* http://www.bbc.co.uk/news/science-environment-42057009 (2017, accessed 06 January 2018).

Hans VP. Judges, Juries, and Scientific Evidence. *Journal of Law and Policy* 2007; 16(1): 19-46.

Hans VP. What Difference Does a Jury Make? *Yonsei Law Journal* 2012; *3*(1): 36-54.

Home Office. *Police and Criminal Evidence Act 1984 (PACE). CODE D – Revised. Code of Practice for the identification of persons by Police Officers*. London: The Stationary Office; 2017.

Horry R, Memon A, Milne R, et al. Video Identification of Suspects: A Discussion of Current Practice and Policy in the United Kingdom. *Policing: A Journal Of Policy & Practice* 2013; 7(3): 307-315 .

Henneberg ML. Admissibility Frameworks and Scientific Evidence: Controversies in Relation to Shaken Baby Syndrome / Abusive Head Trauma. *British Journal of American Legal Studies* 2015; 4(2): 555-584

Henneberg ML. Worlds apart: Cold case reviews and investigations into alleged wrongful convictions in England and Wales. *Journal of Cold Case Review* 2017; 3(1): 24-37.

Ireland J and Beaumont J. Admitting scientific expert evidence in the UK: reliability challenges and the need for revised criteria – proposing an Abridged Daubert. *The Journal of Forensic Practice.* 2015; 17(1): 3-12.

Jackson G, Aitken C and Roberts P. Practitioner Guide No. 4: *Case Assessment and Interpretation of Expert Evidence*. Royal Statistical Society's Working Group on Statistics and the Law, January 2015.

Jones A, Bergold A, Dillon M, et al. Comparing the effectiveness of Henderson instructions and expert testimony: Which safeguard improves jurors' evaluations of eyewitness evidence? *Journal of Experimental Criminology* 2017; 13(1): 29-52.

Jones O, Wagner A, Faigman D & Raichle M. Neuroscientists in court. *Nature Reviews Neuroscience* 2013; 14(10); 730-736.

Law Commission (Law Com No. 325). *Expert Evidence in Criminal Proceedings in England and Wales*. London: Stationary Office, 2011.

Leonard E. Forensic Neuropsychology and Expert Witness Testimony: An Overview of Forensic Practice. *International Journal of Law and Psychiatry* 2015; 42-43: 177-182

Leverick F. Jury directions. In: Chalmers J, Leverick F and Shaw A (eds*) Post-Corroboration Safeguards Review Report of the Academic Expert Group*. Edinburgh: The Scottish Government, 2014. pp. 101-117.

Leverick F. Jury Instructions on Eyewitness Identification Evidence: A Re-Evaluation. *Creighton Law Review* 2016; 49(3): 555-587.

Levett LM and Kovera MB. Psychological mediators of the effects of opposing expert testimony on juror decisions. *Psychology, Public Policy, and Law* 2009; 15(2): 124-148.

Lieberman JD, Krauss DA, Heen M, et al. The Good, the Bad, and the Ugly: Professional Perceptions of Jury Decision-making Research Practices. *Behavioral Sciences & The Law*. 2016; 34(4): 495-514.

Martire KA and Edmond G. Rethinking Expert Opinion Evidence. *Melbourne University Law Review* 2017; 40(3): 967-998.

McAuliff B and Duckworth T. I spy with my little eye: Jurors' detection of internal validity threats in expert evidence. *Law and Human Behavior* 2010; 34(6): 489-500.

McAuliff B, Kovera M and Nunez G. Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law and Human Behavior* 2009; 33(3): 247-257.

Mircioiu C and Atkinson J. A Comparison of Parametric and Non-Parametric Methods Applied to a Likert Scale. *Pharmacy* 2017; 5(26): 1-12.

Newirth KA. An eye for the science: Evolving judicial treatment of eyewitness identification evidence. *Journal of Applied Research In Memory And Cognition 2016;* 5(3): 314-317.

Norman G. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 2010; 15(5): 625-632.

O'Donnell C. and Safer M. Jury instructions and mock-juror sensitivity to confession evidence in a simulated criminal case. *Psychology, Crime & Law* 2017; 23(10): 946-966.

Ormerod D and Sturman J. Working with the Courts: Advice for Expert Witnesses. In: Alison L (ed) *The Forensic Psychologist's Casebook*. Devon: Willan Publishing, 2005.

Pallant J. *SPSS Survival Manual*. 5th ed. Maidenhead: McGraw-Hill Education, 2013.

Papailiou A, Yokum D and Robertson C. The Novel New Jersey Eyewitness Instruction Induces Skepticism But Not Sensitivity. *PloS ONE* 2015; 10(12): 1-16

President's Council of Advisors on Science and Technology. Report to the president. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Washington: Executive Office of the President, September 2016.

Roach K. Forensic Science and Miscarriages of Justice: Some Lessons from Comparative Evidence. *Jurimetrics* 2009; 50(1): 67-92.

Roberts, A. (Andrew J.) Drawing on expertise: legal decision-making and the reception of expert evidence. *Criminal Law Review* 2008; 6: 443-462.

Sapsford RJ and Jupp V. *Data collection and analysis*. 2nd ed. London: Sage, 2006.

Sense about Science. *Making Sense of Forensic Genetics – What can DNA tell you about a crime?* EUROFORGEN, January 2017

Shapiro D, Mixon L, Jackson M, et al. Psychological expert witness testimony and judicial decision making trends. *International Journal of Law and Psychiatry* 2015; 42-43: 149-153.

Sklansky D. Evidentiary instructions and the jury as other. *Stanford Law Review* 2013; 65(3): 407-456.

Stockdale M and Jackson A. Expert Evidence in Criminal Proceedings: Current Challenges and Opportunities. *The Journal of Criminal Law* 2016; 80(5): 344-363.

Thomas CA. *Are juries fair?* Ministry of Justice Research Series 1/10. London: Ministry of Justice, February 2010

Tully G. *Annual Report. November 2016 – November 2017.* Birmingham: The Forensic Science Regulator; 2018.

Valentine T and Fitzgerald R. Identifying the Culprit: An International Perspective on the National Academy of Sciences Report on Eyewitness Identification Evidence. *Applied Cognitive Psychology* 2015; 30(1): 135-138.

Ward T, Edmond G, Martire K and Wortley N. Forensic science, reliability and scientific validity: Advice from America. *Criminal Law Review* 2017; 5: 357-378.

Weiss K and Xuan Y. You can't do that! Hugo Münsterberg and misapplied psychology. *International Journal of Law and Psychiatry* 2015; 42-43: 1-10.

Wilson TJ, Stockdale MW, Gallop AC, et al. Regularising the Regulator: The Government's Consultation about Placing the Forensic Science Regulator on a Statutory Footing. *The Journal of Criminal Law* 2014; 78(2): 136-163.a