

Handle with care: Jury deliberation and demeanour-based assessments of witness credibility

The International Journal of
Evidence & Proof
2022, Vol. 26(4) 381–406
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13657127221120955
journals.sagepub.com/home/epj



James Chalmers 

School of Law, University of Glasgow, Glasgow, UK

Fiona Leverick 

School of Law, University of Glasgow, Glasgow, UK

Vanessa E. Munro 

School of Law, University of Warwick, Coventry, UK

Abstract

It is unclear how effectively jurors perform their task of assessing witness credibility. Drawing on evidence from a mock jury study involving 863 mock jurors deliberating across 64 juries, and building on existing research, this paper explores juries' reliance on demeanour. While jurors make use of factors which the research literature suggests are often appropriate credibility markers, for example external consistency of accounts, there is cause for concern over the nuance with which jurors apply those assessments in high stakes contexts. The manner in which jurors look to manner of delivery as evidence of credibility is also problematic. The paper makes the case for a more circumspect approach towards jurors' use of demeanour assessments. At a minimum, this requires that judicial directions no longer advocate their reliability, but remind jurors of the complexities associated with such assessments and the need to treat any conclusions grounded on presentational cues with caution.

Keywords

demeanour, juries, jury directions, sexual offences, witness credibility

Corresponding author:

James Chalmers, School of Law, University of Glasgow, 5–8 The Square, Glasgow G12 8QQ, UK.

Email: james.chalmers@glasgow.ac.uk

Introduction

Within an adversarial criminal trial, one of the jury's key roles is to assess witness credibility.¹ There will, of course, be trials in which 'what happened' is not in dispute and where the role of the jury lies primarily in evaluative judgments about, for example, how much force is reasonable in self-defence. Most criminal trials will, however, involve at least some element of competing factual testimony. It will be for the jury to evaluate such accounts (and the witnesses providing them), with the aim of delivering an accurate verdict.

Nonetheless, it is unclear how – or how effectively – jurors perform this task. Restrictions on research into the substance of jury deliberations in many jurisdictions mean we lack insight into the ways in which jurors evaluate witness credibility in concrete cases, and the reliability and robustness of the factors underpinning those assessments. Since, as a former Justice of the Supreme Court of British Columbia put it, 'in the courtroom, as in daily life, we are unequipped with the Lasso of Truth' (Smith, 2012: 13), jurors are likely to rely on a range of considerations and cues when assessing witness credibility. Some of these will relate to the *content* of accounts, such as the extent to which claims made by one witness are corroborated by other evidence. Others may relate to the *delivery* of those accounts, particularly in cases where there is a lack of such additional evidence. Cues here might pertain to the demeanour and presentation of the witness, including vocal cues (such as the tone and confidence with which their testimony was given) or non-vocal cues (such as their body language or signs of emotional distress).

Indeed, in some jurisdictions, jurors are specifically directed by the judge to consider delivery cues in their assessment of the evidence (or at least are not advised against drawing inferences from them). In Scotland, for example, standard judicial directions advise jurors that in assessing credibility and reliability they can 'look at the content of witnesses' evidence, their body language in giving it, and compare what they say with other evidence in the case' (Judicial Institute for Scotland, 2021: 5.13). In England and Wales, the *Crown Court Compendium* suggests directing jurors that 'it would be better not to take so many notes that they are unable to observe the manner/demeanour of the witnesses as they give their evidence'.² The importance of demeanour is similarly emphasised by judicial cautions to the jury that where they hear evidence such as hearsay material, they should bear in mind the lack of opportunity that has been afforded to them to assess demeanour,³ and by case law holding that a witness can be required to remove a niqab when testifying, given the importance of the jury seeing a witness's face in evaluating their evidence.⁴

What these directions do not do, however, is tell juries *how* they are to go about assessing demeanour. The implication is that it is something to be left to the common sense of the jury, drawing on their 'knowledge of human nature',⁵ although in some jurisdictions it is recommended that the judge add the caveat that giving evidence is a stressful experience and that 'looks can be deceiving'.⁶ In contrast to this assumption that assessment of demeanour is a key – and typically relatively simple – part of the

1. Judges, of course, make these assessments too – in jurisdictions without juries and in jurisdictions that only use juries for the most serious cases. Our primary interest here, however, is in the assessments that jurors make.

2. The Crown Court Compendium Part I: Jury and Trial Management and Summing Up (2020) page 3-5. See to similar effect page 3-33 on the situation where a jury is provided with a transcript of an 'Achieving Best Evidence' interview: 'it is what they see and hear on the recording which is the evidence not what they read on the transcript. For this reason, they must take care to watch the video as it is shown, so that they can assess the manner/demeanour of the witness when giving evidence.'

3. See e.g. The Crown Court Compendium Part I: Jury and Trial Management and Summing Up (2020) page 14-3 and Judicial Institute for Scotland (2021: 32.3). It may be notable that in the recent case of *Al Megrahi v HM Advocate* 2021 SLT 73 at [57]–[63], when discussing the caution required in assessing a statement not given on oath, Lord Justice-General Carloway did *not* refer to the absence of an opportunity to assess demeanour.

4. *R v NS* 2021 SCC 72 (Canadian Supreme Court) (discussed in Snook et al. (2017)); *R v D(R)*, Crown Court, 16 Sep 2013, unreported (discussed in Nicolson (2014)).

5. *Thomas v Commissioner of Police of the Metropolis* [1997] QB 813 at 831.

6. A phrase found both in National Judicial Institute (Canada), *Model Jury Instructions: 9.4 Assessment of Evidence* (2018) [10] and Judicial College of Victoria, *Victorian Criminal Charge Book* (2021) 1.11.

jury's task, there is a sizeable body of research which has established that accurately discerning veracity through third-party observers' assessments of demeanour is in fact extremely challenging. The presentational cues most popularly presumed to be associated with deception often lack foundation.⁷ Indeed, a key finding of psychological research has been that cues grounded in body language and tone of voice are particularly poor indicators of truthfulness.⁸ While there is ongoing discussion in the literature regarding whether, and how, more accurate evaluations might be made, it is widely recognised to be something that cannot be successfully undertaken without considerable expertise. The skills required extend far beyond anything that lay jurors are likely to possess or be able to cultivate, even with dedicated guidance or training.

Reliance on some level of demeanour assessment may be inevitable in trials, given the performative dimensions of the adversarial courtroom and the human impulse to try to 'read' others. However, we draw here on existing psychological research and new mock jury study findings to illustrate why we should be concerned about the ways in which this currently informs deliberations, and to make the case for a more circumspect approach. The criminal trial involves high-stake consequences, necessary engagement with trauma reactions and complex orchestration of structural and interpersonal power imbalances. This, if anything, is apt to amplify concerns evidenced in the experimental context over the accuracy of, and confidence attached to, third-party demeanour assessments. As such, it is not only unhelpful but potentially unjust to encourage assessors who lack appropriate specialist skills to rely upon them in deliberating.

The paper has three parts. In the first, we give an account of existing behavioural science which demonstrates the difficulties with accurately assessing truthfulness by reliance on demeanour. We also reflect on the relevance of this research to the jury, considering whether in principle there is any reason to have greater confidence in its assessments than those of participants in experimental studies. In the second part, we present findings from a mock jury study, involving 863 members of the public deliberating across 64 juries. We explore the ways in which participants grounded and deployed demeanour assessments as part of the process of individually and collectively evaluating the evidence presented at trial. In the third and final part, we reflect on the implications of our findings in terms of fair and equal access to justice and make the case for a more circumspect approach towards jurors' use of demeanour assessments. At a minimum, we suggest this requires that judicial directions no longer advocate their reliability. On the contrary, we suggest that directions should remind jurors of the complexities associated with such assessment and the need to treat conclusions grounded on cues arising from a witness's delivery of testimony with caution.

What do we know from existing psychological research?

Experimental studies

It is important from the outset to distinguish between witness credibility and witness reliability. A witness could be making every effort to tell the truth, but be innocently mistaken, such as an eyewitness who honestly misidentified a wrongdoer. This is an issue of witness reliability and much has been written about the link between (genuinely) mistaken eyewitness identification and wrongful conviction (Ferguson, 2014; Leverick, 2016; Roberts, 2004). Our interest here is in witness *credibility* – which involves an assessment as to whether the witness is deliberately providing a false, or perhaps redacted, version of events. In that respect, existing research illustrates that – notwithstanding their own confidence in their abilities (Meissner and Kassin, 2002: 476) – people's accuracy in spotting when another person is

7. See the section 'What do we know from existing psychological research?' below.

8. See *ibid*.

lying is poor. Analysing findings across 206 studies, Bond and DePaulo (2006: 223) summarised that, while participants were slightly better at spotting truths than lies (61% v 48%), overall they made accurate deception judgments only 54% of the time (Bond and DePaulo, 2006: 219).

A key difficulty here is that people often hold inaccurate beliefs about the cues that might signify deception. A study by the Global Deception Research Team (2006)⁹ found that the most significant disconnect was the idea that deception can be detected through the speaker's body language – what has been termed the 'demeanor bias' (Vrij and Hartwig, 2021: 2). Seventy-two per cent of respondents in that study believed that liars avoid eye contact; 65% believed that liars frequently shift posture (GDRT, 2006: 68). Similarly, Vrij et al. found it was commonly believed that liars would avert their gaze during testimony, engage in facial fidgeting (e.g., touching their face and hair), and display leg and foot movements, head movements, posture shifts and eye blinking more than those who were telling the truth (Vrij et al., 2019: 309 (Table 2)). In fact, there is no reliable correlation between any of these behaviours and lying. DePaulo et al.'s meta-analysis, examining 120 samples from 115 studies, found there was not one single aspect of body language associated with deception: those telling the truth were just as likely as those who were lying to avert their gaze, fidget, make leg or foot movements or shift their posture (DePaulo et al., 2003: 92 (Table 4)). Indeed, a meta-analysis of 54 studies (Sporer and Schwandt, 2007) has indicated that liars may make *fewer* hand movements than truth tellers.

So too, the content or tone of speech does not necessarily provide a reliable guide. According to the GDRT, 62% of participants believed that liars tend to speak for longer and include more detail (GDRT, 2006: 68); other studies have found that people believe liars will hesitate more in their speech than truth tellers and make more use of fillers such as 'er' or 'um' (Vrij et al., 2019: 309 (Table 2)). Again, the DePaulo et al. meta-analysis established that these beliefs are inaccurate. Pausing or the use of speech fillers was as common for those giving truthful accounts as for liars; and while DePaulo et al. found that there was a relationship to the length of account and level of detail provided, this operated in the opposite direction to that commonly assumed, with liars tending to speak for a shorter amount of time and including fewer details (DePaulo et al., 2003: 91 (Table 3)). DePaulo et al. also reported that those telling lies were more likely than truth-tellers to narrate a story that had internal discrepancies and less likely to structure their account logically (DePaulo et al., 2003: 92 (Table 4)).

Of course, these studies are typically conducted under controlled conditions. Speakers in such experiments usually have no personal investment in the lie being told and face no significant penalties for failing to convince the observer they are truthful (Hartwig and Bond, 2014: 661).¹⁰ A small handful of studies have focused on 'high-stakes lies', suggesting these might be easier to detect because if liars are concerned about the consequences of their dishonesty, their stress will mean that more, or more pronounced, cues to deception will emerge (Porter and ten Brinke, 2010: 58). However, other studies involving footage of real criminal suspects have produced mixed results¹¹ and a meta-analysis of research on this issue has found the accuracy of lie detection varies little between low- and high-stakes settings (Hartwig and Bond, 2014: 665).

It is also important to note that these experimental studies have rarely been designed with the legal context specifically in mind. Where they have been, they have often lacked ecological validity, which restricts the ability to translate their findings into real courtroom settings. It is, of course, difficult to replicate the gravity of consequence associated with the criminal trial and the ways in which this might prompt a different array of behavioural cues that might be interpreted in distinctive ways by

9. Subsequently GDRT. The paper reports on two studies. The first involved 58 countries, 20 male and 20 female participants from each, 2,320 in total (at 62). The second involved 63 countries, 20 male and 20 female participants from each, 2,520 in total (at 67).

10. They also typically involve student samples, although it is less clear that this would itself affect the generalisability of the findings.

11. Compare Mann et al. (2004) with Vrij and Mann (2001).

those assessing credibility. That said, there is little *prima facie* evidence to suggest that these findings in the lab regarding the consistent lack of reliability of demeanour as a guide would not also hold within the courtroom. As we discuss below, the suggestion that the courtroom might in some way mitigate this lack of reliability lacks foundation. If anything, wider complexities associated with interpersonal and structural power dynamics, or cross-cultural norms of communication, within trials may amplify the problem.

From clinical trial to courtroom context

While a criminal trial clearly involves high stakes, it is a very specific type of high-stakes situation. Interactions are guided by evidential rules, and the conventions of evidence-in-chief and cross-examination, which are nothing like normal conversation (Meixner, 2012: 1451). Jurors will often hear evidence from multiple witnesses, some of whom may give directly conflicting testimony. Jurors must assess the credibility of this testimony, both on its own terms and in conjunction with other evidence. Studies that attempt to replicate these courtroom conditions are rare, and the vast majority of studies examining deception in the criminal justice context focus on police interviews and investigative interviewing (Denault et al., 2017: 102).

Nonetheless, it has been suggested that jurors might be well placed to assess credibility (Smith, 2012: 28) because of the structural setting of the trial and/or their own personal qualities. Jurors often have background information available to them and can consider whether witnesses have any motivation to lie. This sets jurors apart from assessors in most laboratory studies who have only a very brief interaction with the testimony-giver, and no background information that might shed light on any possible motivation they might have to lie. Jurors can also look to the internal and external consistency of a witness's account, by comparing it to earlier parts of that witness's testimony or other evidence (Hutchins, 2014: 526). Thus, Minzner has suggested that experimental studies, in requiring participants to assess credibility devoid of context, prime participants to an over-reliance on demeanour cues; but that this may be less of a concern in the jury context (Minzner, 2008: 2567). Grigel also suggests that jurors might be better at detecting deception than participants in experimental studies since jurors have the benefit of immersing themselves in the case for days or weeks and can discuss credibility assessments with fellow jurors during deliberations (Grigel, 2019: 481).

These are important considerations, but their impact should not be over-estimated. The contextual information that jurors have will be limited – often substantially so – by issues of discovery as well as evidence and procedure. Moreover, the trial's adversarial nature can skew the ways in which a witness's account is bolstered or challenged by competing narratives. There are differential levels of access amongst trial participants to advance knowledge of lines of enquiry and opportunities for preparing one's testimony, which can have an impact on how accounts are delivered (Vrij and Hartwig, 2021: 2).

To date, such research as has been conducted directly on jurors' approach to, and skills regarding, detection of witness deception has produced mixed findings. Work on the causes of wrongful conviction does not inspire great confidence, since juries' failure to identify witness perjury has been consistently identified as a leading cause (Gross and Shaffer, 2012: 40 (Table 12)),¹² and anecdotal evidence from US jurors interviewed post-trial suggests inappropriate reliance on behavioural cues. In the Ronald Cotton trial, for example, jurors revealed they had been influenced in their decision to (wrongly) convict by the defendant's facial expressions (Denault and Dunbar, 2019: 921). Meanwhile, McKimmie et al. found – albeit in an experimental study far removed from trial reality¹³ – that mock jurors were significantly less likely to convict when an eyewitness averted their gaze while giving

12. Indeed, in Gross and Schaffer's study, witness perjury was *the* leading cause of wrongful conviction, being implicated in 51% of the 1350 cases they examined.

13. Participants were given a written case outline and watched a video or listened to an audio tape of a single witness speaking for three minutes in response to questions from a lawyer. There was no cross-examination.

evidence, (wrongly) viewing this as a sign of deception (McKimmie et al., 2014: 303 (study 1) and 307 (study 2)). At the same time, though, there is some evidence that jurors might rely on cues to assess credibility in more nuanced and holistic ways. For example, Diamond et al. (2006: 1949 (Table 1)) analysed written questions asked by jurors in fifty civil cases in Arizona state courts and found that almost half of the questions asked were ‘cross-checking’ questions to help judge the plausibility of witness accounts, particularly where credibility was in doubt. Similarly, Young et al., who interviewed 575 jurors across 48 real criminal cases in New Zealand, found that while a minority were ‘unduly influenced by discrepancies in the complainant’s testimony, which led them to reject the prosecution case without regard for independent evidence about the nature of, or reasons for, the accused’s actions’, on the whole jurors assessed witness credibility carefully, taking a measured approach informed by the content of the evidence provided (Young et al., 1999: [3.23]).

One area of particular concern has been the extent to which jurors can apply nuanced assessments of credibility in contexts where they have more ingrained expectations regarding behavioural norms. On the one hand, the fact that jurors might be able to use their own life experience to assess the plausibility of a witness’s account (Smith, 2012: 34) has been cited by some as a key advantage over use of professional judges as fact-finders (Redmayne, 2006: 101). Jurors are drawn from the community and may indeed have experienced the sorts of situations that are the subject of criminal trials (Grigel, 2019: 481). That said, this can be both advantageous *and* problematic. Jurors – individually, or collectively within their juries – will not always have life experience relevant to a witness’s claims (Uviller, 1993: 783). They may also bring false beliefs and biases to bear on their assessments, something that has been identified as a particular concern in relation to sexual offences (Chalmers et al., 2021a; Leverick, 2020).

Indeed, the questionable link between the (actual or perceived) emotional demeanour of a victim-witness and others’ assessments of her credibility within sexual offences trials has been well-documented. There have been two meta-analyses of mock juror studies in this area – one of rape (Nitschke et al., 2019) and one that included rape and other sexual offences (van Doorn and Koster, 2019). Both studies found that complainers¹⁴ who were emotionally distressed when giving evidence were significantly more likely to be regarded as credible than those who were emotionally controlled, although the relationship between emotion and credibility was a complex one and complainers who seemed angry or ‘too emotional’ were also judged as less credible (Nitschke et al., 2019: 956; van Doorn and Koster, 2019: 84).¹⁵ This is supported by a qualitative study of mock jury deliberations, where members of the public – having watched a live performance of a fictional rape mini-trial – were asked to deliberate.¹⁶ Across the trials, the emotional demeanour of the complainer when testifying was varied. The researchers reported that when she was more calm in presentation, jurors were often ‘perplexed’, describing her as ‘cool’, ‘cold’ or ‘calculating’ and drawing negative implications about her credibility (Ellison and Munro, 2009a: 211). When she gave exactly the same account in a more observably distressed manner, this impacted positively on her credibility for most jurors (Ellison and Munro, 2009a: 213).¹⁷ This effect occurred notwithstanding a strong body of evidence indicating that survivors who experience symptoms of posttraumatic stress disorder (PTSD) (Foa, 1997: 25; Kline et al., 2021: NP3154) may suppress emotion to avoid becoming overwhelmed or induce a state of disassociation (Dancu et al., 1996: 258), particularly when – as required in testimony-giving – a traumatic memory is re-activated (Schauer and Elbert, 2010: 118–119).

14. ‘Complainer’ is the legal term in Scots law for an alleged victim. As our study was conducted in Scotland it is the term that we use throughout this paper.

15. In the Nitschke et al. study, it made no difference whether the participants were police officers/law students/judges compared to those with no relevant training (at 965).

16. The extensive steps that the researchers took to make the trial as realistic as possible are outlined in Ellison and Munro (2009a: 204–206).

17. That said, some still doubted her story, remarking that she could simply be a ‘good actress’ (213), reflecting the general scepticism of some jurors to rape claims.

This in turn maps to a wider concern about the ways in which demeanour assessments in the courtroom may be particularly prone to inaccuracy. Studies have shown that people's assumptions about deceptive behaviour cues can reflect a culturally monochrome view that might discriminate against particular groups. For example, research has established that some members of minority groups may be especially nervous when giving evidence for fear of being judged against a negative stereotype (Qureshi, 2014: 259), or may fail to make eye contact as a mark of respect (Rand, 2000; Uono and Hietanen, 2015),¹⁸ but this can be unfairly interpreted by others as suspicious. So too, there is evidence that adults with autism are more likely to be (wrongly) judged as deceptive because they are more likely to display behaviours people *think* are associated with deception, such as a lack of eye contact or fidgeting (Lim et al., 2022: 11).

Particularly for victim-witnesses in criminal trials, these complexities can be compounded by misconceptions regarding memory recall and narrative reconstruction in the aftermath of trauma. Research exploring the impact of trauma on memory has highlighted that sequential narration or accounts without internal discrepancy should not be held out as reliable indicators of credibility. It is often assumed, for example, that a genuine victim would recollect a traumatic event in detail (Haskell and Randall, 2019: 18). Some of the laboratory research above would support this to the extent that one of the more reliable cues to deception has been found to be the level of detail in the account, with liars including fewer details than truth tellers. The vast majority of this research has, however, been undertaken with volunteers recounting unremarkable facts or events. By contrast, work with victims of trauma has clearly indicated that – as a result of the ways in which the brain encodes traumatic memories (Haskell and Randall, 2019: 23) and the complex effects of PTSD (Mason and Lodrick, 2013: 31) – it is very common for memories of events (such as a sexual assault) to be fragmented. This can make it difficult to recall or recount details in a complete or linear way (Haskell and Randall, 2019: 20; Mason and Lodrick, 2013: 27). It is not clear that those who assess credibility take this adequately into account (Baillot et al., 2014).

Overall, then, while there has been little detailed exploration to date of the extent to which, and ways in which, jurors rely on demeanour cues within their deliberations, existing research provides a robust counter to any intuitive self-confidence in jurors' ability to 'read' others accurately. It also highlights the need for particular care in the courtroom context, given the potential for high-stake lies, prevalence of traumatic narratives, effects of power dynamics and communication barriers, and constraints of adversarial testimony-giving. Improving our understanding of credibility assessment in the courtroom is vital since, despite their importance to verdict outcomes, jurors' credibility assessments are 'virtually unreviewable' (Bennett, 2015: 1334). Doing so, however, requires more than simply asking for jurors' individual evaluations. The jury is, after all, a dynamic site of discursive exchange in which assessments are challenged and collective narratives forged. The individual (pre-deliberation) preferences of jurors do not always determine the final jury verdict (Hastie et al., 1983: 63–65), and verdict destinations can be reached through myriad routes. Thus, studies that include realistic stimuli and collective deliberations situated in the wider framework of evidential and substantive legal tests jurors must consider are key. In the remainder of this paper, we provide an account of the design and staging of, and findings from, one such large-scale study.

Jurors' deliberations on demeanour

The Scottish jury research

The results we report here are drawn from the Scottish Jury Research, which was funded by the Scottish Government and conducted by the authors in conjunction with Ipsos MORI Scotland (Ormston et al., 2019). This study was not specifically designed primarily for the purposes of exploring jurors' use of

18. There was some evidence of this in the GDRT, where gaze aversion was significantly less likely to be associated with deceptive behaviour by respondents in Middle Eastern countries such as the UAE (GDRT, 2006: 66).

demeanour cues to assess witness credibility. Scottish juries have 15 members, make decisions by simple majority and have available to them – in addition to guilty and not guilty – a third ‘not proven’ verdict (see Chalmers et al., 2022). The primary aim of the Scottish Jury Research was to explore the impact of these features on jury decision-making, and to examine how jurors understood and used the not proven verdict in particular. The study does, however, provide valuable insight into how jurors approached the task of assessing witness credibility, and we have drawn out those strands of analysis in what follows.

The research involved two trial reconstructions (of a non-sexual assault and a rape case), scripted with input from legal practitioners, performed by actors in the High Court in Edinburgh, and professionally filmed and edited. The resultant videos (of approximately an hour) were shown to 64 mock juries, involving a total of 863 participants. The use of videos ensured that the content of the trials (including the manner of delivery of evidence) was held constant across the study. Jurors were drawn from jury service eligible members of the public to reflect a representative sample of the local population. Half of the juries watched the assault trial and half the rape trial. After observing the relevant trial video, participants deliberated in jury groups towards a verdict. The deliberations – which lasted up to 90 min – were audio- and video-recorded.

To reflect the primary aims of the study, jurors deliberated in groups of either 12 or 15.¹⁹ Aside from this difference, the procedure that the juries went through was identical, save for the directions they were given at the end of the trial about how they should make their decision. Half of the juries could return a verdict by simple majority; half had to strive to reach unanimity. Half of the juries could choose between the three verdicts of guilty, not guilty and not proven; the other half only had the options of guilty and not guilty. An analysis of which substantive aspects of the evidence the juries discussed found no statistically significant differences related to these variations (see further, Ormston et al., 2019: 30).

Previous work drawing on data arising from this study has explored the impact of the unique features of the Scottish jury system on verdict choices (Chalmers et al., 2020); the case against the not proven verdict (Chalmers et al., 2022); and jurors’ reliance upon misconceptions about sexual violence in verdict decision-making (Chalmers et al., 2021a). This paper focuses – against the background of the psychological research discussed above and the fact that judicial directions in Scotland specifically instruct jurors to consider the body language of witnesses – on the cues jurors used to assess credibility. It explores in particular the ways in which witnesses’ demeanour was relied upon to inform assertions about the believability of their claims.

In both trials, jurors were required to make credibility judgments to assess competing witness testimony. The rape trial involved a female complainer and male accused who had previously been in a relationship and who met in the complainer’s flat. Their accounts diverged, with the complainer alleging that the accused had raped her and the accused testifying that they had engaged in consensual intercourse. A forensic examiner testified that the complainer’s injuries were consistent with her account but that other explanations could not be excluded. In the assault trial, the complainer and accused, who were both male, were involved in an altercation outside a bar, in which the complainer was stabbed. The accused claimed he had acted in self-defence. The complainer’s partner witnessed the later stages of the altercation and gave an account during the trial that was consistent with the complainer’s. Further details of both trial scenarios can be found in Appendices 1 and 2 respectively.

The choice of the mock jury method was necessary for the wider project,²⁰ but also provided a window into credibility assessment which would not otherwise have been

19. In one jury, a juror viewed the trial but did not take part in deliberations as they felt unwell. That jury continued with 11 rather than 12 jurors. This is why the total number of jurors in the study is 863, not 864.

20. This aimed to investigate whether the specific features of the system made any difference to juror verdict preferences. To shed light on this, it was necessary to keep everything else about the case constant while varying the conditions in which jurors reached their verdict.

possible.²¹ Research undertaken in New Zealand has asked jurors how they made their credibility judgments in real cases (Young et al., 1999: paras 3.20–3.25).²² This would not have been permissible within current restrictions on researching the content of the deliberations of real jurors in Scotland, and in any event asking jurors to report retrospectively on what went on in the jury room has its own difficulties. It relies on accurate recollection and there is a danger that jurors may leave out of their account aspects that might be regarded as socially undesirable.

At the same time, it is important to bear in mind the limitations of the mock jury method. For one thing, there is an inevitable element of role-play involved, as well as streamlining of court processes to render reconstruction feasible. While this has led some to argue that there are ‘fundamental differences between real jurors and volunteers’ (Thomas, 2020: 1006) that severely limit the reliance that can be placed on any findings, that conclusion can be disputed (Chalmers et al., 2021b). Mock and ‘real’ jurors are not fundamentally different populations. Our jurors were all eligible for jury service and the compulsory nature of jury service means they could easily end up as real jurors in real cases. Moreover, while it is true that our jurors knew that they were not making a decision that had real-world consequences, we took steps to maximise the solemnity of the process and the authenticity of the trial. These included working with legal professionals in the scripting and delivery of the trial, incorporating in the videos appropriate legal directions delivered by a real judge and requiring all participants to take the standard juror affirmation before viewing the trial (Ormston et al., 2019: ch 2). There was evidence that participants took their deliberative task seriously. Deliberations often became heated, with participants reporting afterwards that they had found the process stressful at times. In addition, participants made comments throughout the deliberations regarding the behaviour of the witnesses that indicated they had suspended disbelief and forgotten they were viewing actors.

Though participants also completed pre- and post-deliberation questionnaires that posed specific questions tied to the primary focus of the project, we focus in the remainder of this discussion only on their deliberations, where their individual and shared approaches to assessing credibility were articulated, evaluated and defended. Audio recordings of these deliberations were transcribed and imported into NVivo for thematic coding and analysis. The scale and timelines for completion of the overarching jury project necessitated a collaborative approach to that coding. To facilitate this, and ensure inter-coder consistency, a sub-sample of deliberations was selected, and each member of the team ($n = 5$) undertook initial coding on an open basis (Glaser and Strauss, 1967), in order to identify top-level themes. The outputs of this initial coding were shared by and to all coders, synergies identified and any points of inconsistency or duplication remedied. This generated an agreed ‘tree node’ structure with headings tied to, amongst other things, jurors’ discussion of the parties’ injuries, their application of legal tests for consent and self-defence, their understanding of relevant evidential thresholds, their evaluation of the behaviour of the parties before and after the incident, and deliberation dynamics including use of votes. Thereafter, the team proceeded on the basis of this more structured coding frame in order to maintain consistency, although always with sufficient flexibility to flag, discuss and add newly emergent themes.

Amongst the headings in this ‘tree node’ structure that guided the primary project report was a theme around jurors’ assessments of the ‘credibility and reliability of the parties’. This was broken down further into jurors’ comments regarding the *demeanour* during testimony-giving of the accused, complainant and any other witnesses (with 162, 162 and 43 references respectively), and wider observations regarding the believability or credibility of the accounts provided by parties (with 269 and 448 references respectively for the accused and complainant). In preparation for this article, these references were further excavated by the authors to generate a more fine-grained analysis, informed both by the literature review above and by prior familiarity with the deliberation data. We created a coding frame in relation to credibility cues,

21. In Scotland, s. 8 of the Contempt of Court Act 1981 precludes asking about ‘statements made, opinions expressed, arguments advanced or votes cast by members of a jury in the course of their deliberations’.

22. New Zealand does not have the absolute restriction on asking about jury deliberations found in the UK jurisdictions.

which was divided into three broad categories: (1) non-vocal cues and body language, (2) vocal cues tied to the delivery of evidence and (3) cues tied to content. Within each of these, further sub-divisions were created to capture specific manifestations – for example, eye contact or fidgeting as examples of non-vocal cues; appearing emotional, or nervous or hesitating/stuttering as examples of vocal cues tied to delivery; and level of detail, internal/external consistency of testimony and use of jurors' personal experience as examples tied to the content. The prevalence and parameters of discussion around these themes, both within and across different juries, were then considered. This informed our analysis below of the ways in which these cues were relied upon to ground juror assessments, and how they influenced collective deliberations.

Though we thus charted the number of references in each sub-theme as part of the process of our analysis, the aim of doing so was not to make statistical claims regarding prevalence, but rather to provide contextual grounding and an increased sense of depth to the qualitative analysis with which we were primarily concerned. Given the complicated relationship between individual and collective jury outcomes, and the fact that sometimes a single interjection can hold far greater discursive sway than a series of contributions on the same theme, we have focused any discussion around prevalence primarily at the jury level. Even one mention of a theme within a jury deliberation (which varied in length but did not exceed 90 min) would suffice for it to be included in the relevant tally, but for the most part, such singular contributions were rare. Where a theme was raised, it tended to be returned to by the same or other participants subsequently, although this could be for very different reasons. For example, sometimes it was because it became a prominent focus of debate between participants with competing perspectives and other times it was because it was widely deemed to be convincing and so it impacted the subsequent discursive trajectory of the group. In that sense, numbers of references tell a very partial tale, and while operating at jury rather than individual juror level assists with a more overarching lens of analysis, the relatively small numbers of juries involved across the two trial types precludes detailed statistical claims.

Which credibility cues did jurors rely on?

Turning to our findings in respect of jurors' approach to credibility assessment, Table 1 shows the cues that were most commonly present within jury discussions. Although there is not a clear line of delineation between the two dimensions, we have categorised these in terms of the *content* of testimony and its *delivery*. While jurors focused on the content of testimony in all 64 trials, they also looked to issues of delivery in 54 trials, with this being slightly more common in the rape trials than in the assault trials. It is immediately apparent that many of the cues relied upon in that respect are dubious and run contrary to the experimental evidence. In particular, there was a heavy focus on body language, and juries often drew inferences from the extent to which they felt a witness appeared nervous whilst giving their testimony. Though neither of these are reliable indicators of veracity, as we will discuss below, they were sometimes asserted with substantial confidence in the jury room, and they frequently played a role in justifying individual and collective verdict choices. Different expectations also appeared to be at play among jurors in the different trial types, to the extent that there was a pronounced focus in rape juries on the complainant's emotional presentation and its conformance to preconceived expectations of 'genuine victimhood'.

Content of testimony

Level of detail. In 39 of 64 juries, jurors looked to the level of detail the witness provided in their testimony to assess their credibility. There was no notable difference between the frequency with which this was done in the rape and assault trials, but the way it was done differed.

In the rape trial, there was evidence of bruising to the complainant's inner thighs and chest and scratches to her breasts that were consistent with considerable force being applied. The complainant was able to provide a detailed account of how – she claimed – those injuries were sustained, whereas

Table 1. Credibility cues used most frequently by jurors

	Number of juries (n = 64)	Assault juries (n = 32)	Rape juries (n = 32)
Content of testimony	64	32	32
Motivation to lie	56	29	27
External consistency	50	19	31
Level of detail	39	18	21
Internal consistency	29	17	12
Use of personal experience	24	12	12
Delivery of testimony	54	24	30
Body language	48	23	25
Level of emotion	27	3	24
Level of nervousness	27	15	12
Calmness of delivery	12	7	5
Confidence of delivery	11	4	7
Hesitation (e.g., stuttering, using ‘um’/‘er’)	11	6	5

the accused’s account was vague about how this might have happened. Jurors sometimes contrasted the two accounts, citing the accused’s vagueness as a factor pointing towards guilt, as the following passage from one deliberation illustrates:

F: What do you think she said that makes you sure he is guilty?

F: The way that he was not being specific as to what he can remember, he is a fully grown man and he can’t remember where he put his hands upon her leg in that moment and she can remember every single detail.

F: Uh-huh.

F: Why is that not picked up on that he can’t remember what he did? If he wanted to present himself as a credible defence for this, he should be able to say to the letter what he did and what he didn’t do. (M07F)²³

That said, jurors did not always see the detail of the complainer’s testimony as indicating that she was telling the truth. On the contrary, some jurors perceived her answers as *too* detailed and a sign, therefore, that her account was fabricated. One juror, for example, commented that ‘I felt because like everything [the complainer] said, she knew exactly what was happening, I kind of felt like it was rehearsed’ (M01H). Another exchange went as follows:

F: Aye, but what I’m trying to get at is if she was lying she would not have that story consistent at all like, you know, if it was a really traumatic event she would still remember it, that’s what I’m trying to get at. Like he couldn’t even think what even happened, if you know what I mean.

F: Sometimes they say though, when people are lying, they give *more* information. (M05F, emphasis added)

Relatedly, while other jurors appeared to be aware that trauma can affect recollection, they used this ‘knowledge’ to cast doubt on the complainer’s account, by suggesting that *because* she could give a detailed account, she could not possibly be traumatised. One juror, for example, commented that ‘I would have thought if something like that happened and you were traumatised and whatever you would be a bit kind of hazy about it, a bit vague’ (M05F). Another stated:

23. Each jury was given a unique code, used here to demonstrate that views were expressed across a wide range of juries.

F: I've been in situations, and no matter how much you try, it's very difficult to remember every single thing that happens in the order, so for that reason I thought that it had been kind of engineered a bit by her, you know, she seemed to have all the answers ... I just felt that she remembered so much detail and fortunately I've never been in a situation like that, but I've been in situations that have been quite traumatic and trying to recall it is really, really, difficult. (M02G)

In the assault trial, the role that detail/vagueness played in deliberations almost always centred around the witnesses' recollections about how much they had drunk. In this trial, events took place outside a bar, in which all parties had been drinking alcohol. The accused provided a detailed account of how much he had drunk.²⁴ The complainant and his partner did not, and were vague in their responses to cross-examination on this issue. Jurors often used this to cast doubt on the credibility of the rest of their evidence, with one juror preferring the accused's account of self-defence to the complainant's account of an unprovoked attack on the basis that 'they never actually said how much drink they had, they were all sketchy about that, whereas the accused was sure he had four pints' (M01C). Another exchange was similar:

M: One thing about [the complainant's] account, which was sort of prevalent all the way through it and of his girlfriend, was the fact that they couldn't really recall the amount of alcohol consumed in an evening, and I find that quite...they couldn't even give us a rough estimate of how many pints they had or frequency of pints or how much they spent and I find that...

M: Suspicious.

M: ...yes, very suspicious. (M02C)

Internal consistency. Internal consistency refers to the extent to which (if at all) the witness contradicted themselves in giving evidence. The internal consistency of a witness's account was raised by participants as a basis upon which to judge credibility in 29 of 64 juries. As Table 1 shows, this was more common in the assault trial than the rape trial. This was not surprising as it was directly raised in the assault trial, where cross-examination of the complainant's partner prompted an admission that she had lied when she said she had seen her partner being pushed to the ground, although she insisted all other aspects of her evidence were, nonetheless, truthful. Jurors suggested that her discredited claim 'called into question her other statements' (M03B) and that 'if she was lying about something you can't really trust her' (M05D). Internal consistency was also used to assess credibility in the rape trial, where the consistency of the complainant's account was sometimes contrasted to the accused's, as illustrated by the following juror's comment:

F: The other thing was with him I found that he contradicted himself, so he made his statement and said, at first he said, oh, her arms were wrapped tightly round me, but then when he was pressed about the bruising he said, oh well, maybe I'm holding her wrists and that's how she got it. Then, so I was like, well earlier you said that, and now you're saying that, and she was consistently saying this happened, that happened, that's how it all came to be. He didn't have such a clear vision of how that would happen, so I went with guilty. (M03F)

Occasionally jurors did question whether the fact that a witness had been shown to lie about one aspect of their evidence necessarily meant that their entire account could not be believed. Indeed,

24. He testified that everyone in his party had purchased a round of drinks, so he was able to state confidently that he had consumed four pints of beer.

they had been directed by the judge not to make this assumption.²⁵ But this was rare. It was far more common for jurors to assume that a single untruthful statement by a witness cast the rest of their evidence into doubt.²⁶

External consistency. External consistency refers to the extent to which the witness's account is supported by other evidence in the case. As discussed earlier, this is one factor that is generally thought to be an effective way of assessing witness credibility. As part of a standard direction on assessing credibility, jurors in the study were told by the trial judge 'to compare what the witness said to other evidence in the case'. External consistency was used as a measure of credibility in 50 of 64 juries, although this was more common in the rape trial, where it was referenced in 31 of 32 juries (compared to 19 of 32 assault juries).

In the assault trial, discussion generally focused either on inconsistencies between the evidence of the complainer and his partner or on whether the accused's testimony that he had grabbed the knife from the complainer was consistent with evidence that he was personally uninjured. These were pertinent questions and jurors spent some time weighing them up. However, jurors did sometimes place more weight than would seem to be warranted on minor and relatively inconsequential inconsistencies, as illustrated here:²⁷

F: The reason that I said I don't think he is guilty, I have reasonable doubt, is because going with what the judge said about credibility and reliability and corroboration of two witnesses, I don't feel that [the complainer] and his girlfriend's accounts tie up very well. So, that leaves me...well he said that they went out... I know it's maybe a bit of a stickler point they [the complainer] said they went at seven, and she said they went out at seven-thirty. (M08D)

In the rape trial, jurors who regarded the complainer's account as credible often pointed to the fact that it was consistent with her injuries (for which, they felt, the accused was unable to provide a convincing alternative account). One juror commented, for example, that '[the complainer's] story was a lot more consistent, she had evidence for every injury she sustained, she could say how she got injured' (M01H). Another stated that '[the complainer] gave a description about how she got these bruises' whereas '[the accused] did not give an adequate description of how she might have got these bruises'. This juror continued: 'Now, I can imagine she might have had bruises after passionate sex, but it seems strange that they would be there in the places she mentioned exactly how, they were exactly as to how she described herself being attacked' (M01G).

One theme that did arise in the rape trial, however, was that the external consistency of the complainer's account with the physical evidence was sometimes used *against* her, as a sign of fabrication. For example, jurors commented as follows:

F: See the other thing is right, and I don't believe this, but if she has got bruises, she can then make the story up to match the bruises, and she was very, she knew, she was very, like I would have thought if something like that happened and you were traumatised and whatever you would be a bit kind of hazy about it, a bit vague, you would be a bit, oh, he was doing this, and doing that. But she was very factual.

25. This direction was as follows: 'It is perfectly possible for you to accept part of a witness's evidence and to reject part; or you may accept it all or reject it all. If you reject a piece of evidence because you think a witness is lying, it doesn't mean the opposite must be true. It only shows that the witness does not want to tell the whole truth on the matter. Beyond that, you should lay that evidence aside and proceed on the evidence you accept.'

26. This was compounded in the assault trial by the relationship between the complainer and his partner, which jurors viewed as giving the latter a motive to lie (see below).

27. The juror's reference to corroboration reflects the fact that Scots law continues to require corroboration in almost all criminal cases, and the jurors were directed accordingly by the judge.

F: Aye, the facts.

F: So, she could easily...if she had a bruise on her chest she could go, right, I need to get into this story to fit the bruises, which I don't believe, but it's again enough to put doubt in my mind. I don't think that's true, but...

M: Aye, it's enough to doubt it.

F: So, she could have made the story match the bruises as opposed to the bruises match the story. (M05F)

The implication here that the complainer had contrived her testimony was not one that arose in any of the assault juries. What we see here – as in relation to the level of detail provided by the complainer – is the rape complainer apparently being judged more harshly than the assault complainer. Factors which would normally be signs of credibility acted against her and were turned into evidence of deception.

Motivation to lie. Considering whether a witness has any obvious motivation to lie is clearly pertinent in assessing credibility and jurors did this in 56 of 64 juries. In the assault trial, discussion tended to centre around a plausible motivation for the complainer's girlfriend to lie – that she wanted to protect her partner. In the rape trial, discussion centred around rather less plausible motivations that the complainer might have for fabricating a rape allegation. In that scenario, the complainer and accused had a brief relationship, ending two months prior to the events in question. There had been little contact between them since, until the accused called round to the complainer's flat. According to his account, the two of them had consensual sexual intercourse, but he then stated he did not want to continue their relationship and left. In 27 of 32 rape juries, jurors speculated that the complainer might have made a false allegation to get revenge against him, as she had hoped to resume their relationship and was disappointed he did not wish to. For example, jurors commented as follows:

M: I mean the old adage, I mean it's a female thing of no hell like a woman scorned. So, she had been trying to get back with him, she had been phoning him and everything else. When he turns up what does she do, she just, you know, accepts it and has the sex and that and he says, no, I don't want to take it any further. That's the thing, she could be using it to get back at him. (M03G)

M: So, if you look at that, if the guy actually said that was a mistake and the girl said okay, all right, so I've been trying my best all along you've not been responding and now you come to my house, we have sex together and then you told me, you're now telling me it is a mistake. So, it is kind of a motive that to me, that's the way it looks like a motive for her to actually want to set the guy up that this is not going to happen, we are never going to be, but I'm going to bring you down. (M04F)

The plausibility of this revenge motivation was questionable. It was agreed in evidence that there had only been two short telephone calls between the complainer and accused since their relationship ended and there was no evidence presented that the complainer wished to resume it. Despite this, some jurors imputed this possible motivation into their credibility assessments, as the following quote illustrates:

M: According to the testimony of the guy he said at the end he told her that they still can't be together. Then it...round it there in my mind that she could have set it up or used this opportunity as an opportunity for them to come back together. If at the end he doesn't see the purpose, they have slept together, it is normal that she will be angry and look to do something. (M01F)

Jurors did occasionally question how realistic the revenge motive was, as shown in the following exchange:

F: I have come to the conclusion guilty, just on the evidence, so on both sides it's their relationship in terms of how she is speaking about him. Also the evidence for the defence was really flimsy. They kept saying about the

phone calls, but she only phoned twice, you know, in the period after they actually split up. Also why would she lie? She has got no reason.

M: Why would she put herself through that ordeal? (M01E)

However, most claims of an implausible revenge motivation went unchallenged within rape juries, and such speculation about the complainer's motives for lying was often accompanied by a general scepticism about the veracity of rape complaints and a belief that false allegations are frequent (Chalmers et al., 2021a: 245–246).

Use of personal experience. The use of personal experience to assess the plausibility of testimony is sometimes cited by commentators as an advantage of juries over professional judges (e.g., Redmayne, 2006: 101). While this may be so, it is important also to acknowledge that it might not always be advantageous. There is a risk that jurors generalise inappropriately from their personal experience, assuming that others would always react in the same way, or that other jurors might place too much weight on anecdotes when drawing conclusions about credibility.

There was plenty of evidence in the deliberations of jurors using their personal experience to assess witness testimony, with examples of this arising in 24 of 64 juries. In the assault trial, jurors used their own experience of being in similar situations to judge the credibility of various aspects of the witnesses' evidence, including the fact that the accused ran away and did not phone the police after the incident and the complainer's claim that he picked up the knife after being stabbed. These insights were generally positive and helpful to other jurors, leading them to reconsider their initial misplaced assumptions about how someone might react in a stressful situation. One example is the following exchange:

M: Right, I'll explain that one, I've been involved with knife crime, I have been hit with a wee cleaver in this hand, right, and you don't feel it at the time you don't feel it, whenever you're in a fight you don't feel it. So, you're getting up your adrenaline is going...

M: Your adrenaline kicks in.

M: ...so your adrenaline is kicking in, so that would explain why he has been able to [pick up the knife]. (M02B)

In the rape trial, the use of personal experience in deliberations was not always so positive. Here, jurors drew on their personal experience mostly to assess the complainer's account of the extent to which she tried to resist. In particular, they extrapolated from their own experience of trauma to suggest that if her claim were true, she would have been more forceful. This is illustrated by the following comment:

M: I can say that, I have been attacked by somebody that is bigger than me and I've tried to fight back and I knew he was stronger, I knew he was all the rest of it, but at the end of the day I still tried to fight back.

F: Aye.

M: That's good, that's good, but you've got that strength and all the rest of it. Not everyone is a hero or a macho person. (M07F)

At the same time, there were some jurors who also drew on personal experience to challenge such claims and lend weight to the idea that a victim might freeze and not fight back:

F: I've not been attacked like that but I had somebody at my door at half three in the morning trying to bash my door in and when I saw there was somebody at my door I absolutely froze. I literally just froze. (M01G)

In summary, reference to personal experience, while relatively common, was of mixed utility. In the assault trial, because knife crime and bar fights are not within everyone's general experience, it was sometimes useful for jurors to hear from others who had been in similar situations. Juries here did exactly what they are supposed to do, bringing knowledge to decision making that would – most likely – be lacking if a professional judge were the fact finder. In the rape trial, however, the reliance on personal experience was not always helpful. Biases about the extent to which a genuine victim would resist an attacker came into play, with (usually) male jurors drawing on their own experience of resisting attack to surmise that the female complainant could have resisted more forcefully, in contradiction to the extensive body of psychological literature establishing that 'freezing' or dissociative responses are common (Marx et al., 2008; Schauer and Elbert, 2010). This personal experience lent weight to more general prejudices and false beliefs already held by many jurors about the extent to which a victim of sexual assault would fight back (Chalmers et al., 2021a: 236–240), contributing to scepticism about the credibility of rape complaints in general and, by implication, this rape complaint in particular.

Delivery of testimony

Level of emotion. While both trials involved a witness testifying that they had been the victim of an attack, it was (as Table 1 shows) almost exclusively in the rape juries that inferences about credibility were drawn from the extent to which the complainant appeared emotional. When considering the extent to which a genuine victim would be emotional when testifying, jurors' expectations of a male assault victim (whose relatively unemotional delivery passed largely without comment) thus seemed to differ from those regarding a female rape victim.

It is important to say that the Scottish Jury Research did not *test* the impact of emotional delivery on credibility assessments. It ran only one version of the rape trial, in which the complainant was visibly emotional when giving her evidence – she became upset, for example, at various points and once had to be asked if she was OK to continue – but this was not tested against a scenario in which her delivery was unemotional.²⁸ Bearing that caveat in mind, it is nonetheless notable that jurors often pointed to the complainant's distress in testifying as indicating credibility. One juror commented that 'I don't think you could fake those kinds of emotions when she is going through it like that' (M05G). Another juror commented that he found her 'convincing' because 'she was emotional and upset right from the start' (M07E). Yet another noted that 'she was quite tearful on the podium there [and] I feel that her story was credible' (M02F).

The complainant's relatively emotional delivery was also sometimes contrasted unfavourably to that of the accused, who was seen as too 'calm' (M06F) or 'collected' (M07H) to be credible. As one juror put it:

F: There isn't DNA, but, credibility, reliability, she was quite emotional as I think you would be and he was just as cool as a cucumber, so I don't give him any credibility at all. (M06G)

Despite the actor playing the complainant having been asked to make her delivery somewhat emotional, she was still sometimes seen as *not emotional enough* to be a credible rape victim.²⁹ One juror stated that they did not believe her story as 'she didn't come across distressed enough personally for me' (M04E). Another saw her delivery as 'a bit tentative' and commented that 'I felt there would have been a bit more anger and rage if he had done it' (M03E). One juror doubted the complainant's account because 'I just felt sometimes when she was speaking it wasn't from her heart' (M01H). Another exchange was follows:

28. For two meta-analyses of studies in which this *has* been done, see Nitschke et al. (2019); van Doorn and Koster (2019). Both studies concluded that distressed rape/sexual offence complainants were significantly more likely to be perceived as credible.

29. See van Doorn and Koster (2019: 87).

M: I just think there wasn't enough emotion or anything to implicate that she was under...

F: She was shaking.

M: But I think there would be more panic, there would be more, especially through discussing and all of that again to bring it all back up. (M03E)

Conversely, some jurors reported that they saw the complainant as *too* emotional to be credible or doubted that her emotion was real,³⁰ describing her as 'putting on an act' (M08G) or suggesting that she only appeared 'really, really upset' because 'she might be a good actress'.³¹ One juror commented that she thought 'girls could do that [fake emotion] if they had an agenda' (M07H). Another suggested that:

F: You have to keep in mind that women show more emotions than men and court cases as well you misread body language from women, because they are better at that, presenting themselves as a victim. (M02E)

The level of emotion displayed by the accused in the rape trial was commented on during deliberations far less frequently, compared to that of the complainant, though there were some examples of this. Jurors occasionally pointed to behaviour in the witness box that they thought indicated the accused was angry at being accused of rape, seeing his anger as signifying credibility; conversely, some pointed to his relatively calm delivery as a sign that his account was not credible. There were also instances where jurors perceived the accused as too emotional to be credible, indicating that, in line with previous research (Wessel et al., 2012), there were expectations about the 'proper' level of emotion to be displayed by an accused, as well as a complainant. One juror commented that 'if a guy was being accused of rape whether he was guilty or not, I would think he would be a lot more distressed himself, and he didn't seem to be bothered' (M04E).

All of this can be contrasted with the focus of discussion in assault juries where there were only three comments made about the emotional demeanour of any trial parties across all 32 jury deliberations.

Body language. Despite research demonstrating that body language is not a useful indicator of credibility, jurors in 48 of 64 juries in this study drew on the witnesses' body language in an attempt to discern whether or not they were truthful. That the jurors drew on body language in this way is perhaps unsurprising, as they had been specifically directed by the judge to do so.³² The confidence with which jurors (wrongly) pointed to particular aspects of body language as signifying veracity or deception was, however, remarkable.

Gaze aversion was one of the most frequently perceived signs of untruthfulness. One juror doubted the complainant's testimony in the assault trial because he 'couldn't make eye contact with anybody' (M02D). Another regarded the complainant as 'shifty' because he 'didn't make eye contact', contrasting that to the accused, who was seen as credible because he met the questioner's gaze (M06C). In the rape trial, the accused's credibility was doubted by one juror because he 'kept looking away' (M05F). Another claimed that 'something I'm quite familiar with is body language' and asserted that the accused's story should be questioned because he 'looked to the left quite a bit' (M05G). Similarly, one juror commented:

M: He kept looking to the side and kept looking to the side, and I know that that's a normal sign that you're not 100 per cent telling the truth [and] that you're kind of wandering to the right and like kind of fidgeting and things like that. (M06G)

30. See similarly Ellison and Munro (2009a: 212); van Doorn and Koster (2019: 87).

31. In these exchanges, the jurors appeared to have forgotten that the witnesses *were* actors.

32. Jurors were directed in line with the guidance in the Scottish Jury Manual that they could 'look at the content of witnesses' evidence, their body language in giving it, and compare what they say with other evidence in the case'.

Unusual hand movements were another commonly perceived sign of deception. Witnesses' credibility was questioned because, for example, they were 'twitching' (M03B), 'wringing' (M02D) or 'playing with their hands' (M02D) or, in one case, because the witness's 'hands were quivering' (M05F). One juror stated that they favoured a guilty verdict because the accused's 'hands started moving when he was trying to give details of what happened that night' (M02H). Another said of the point at which one witness started to move his hands 'that's when he started to lie' (M05G). One juror commented:

M: I went for guilty, purely because I noticed when he was giving evidence his body language was all over the place. Like one minute he is fidgeting his hands like this, he kept pulling his fingers, he kept looking to the right and things like that, which kind of made me feel as if he wasn't comfortable. That stuck out to me like a sore thumb. (M06G)

Another exchange went as follows:

M: You said that she was fidgeting when she was giving...the body language she has...

M: Aye, she was awful shaky...

M: Yes, some people, when they are lying, go that way, yes. (M02H)

Eye and hand movements were not the only aspects of body language used to judge credibility. Jurors doubted credibility because, for example, '[the witness's] neck was straining' (M02B) or '[the witness] kept licking his lips' (M03F). One doubted the witness because 'it's just the way he stood as if...he looks a bit dodgy' (M06B). Another asserted that 'when I was watching the video I was very observant, when [the accused] was asked if the knife was his, from his facial expressions you could tell he wasn't being honest' (M04D).

Fidgeting more generally was seen as suspicious, with one juror commenting about a witness that 'I think he is lying by the way he was fidgeting and didn't look right' (M06E). Bodily movements were sometimes seen as a sign of nervousness, which jurors commonly associated with untruthfulness. One juror favoured a not guilty verdict in the rape trial because the complainant 'was very sort of nervous, twitchy...which I kind of thought was a bit odd compared to [the accused's] mannerisms' (M07F). Another commented of the rape accused:

M: As it got further and further into his evidence, he got more and more nervous and more and more worried, he was looking, he was maintaining eye contact as the gentleman earlier on said, wringing his hands in some kind of almost biblical way, and as he went through the attack or whatever, consensual sex, he became more and more and more nervous. (M05G)

Such assumptions run contrary to the prevailing view in the research, which is that witnesses can be nervous when testifying for all sorts of reasons unrelated to their veracity. Jurors sometimes did suggest that a witness's nerves could simply be due to the stress of being in court, with one commenting that if they themselves were in court, their 'fingers would be twitching' as 'it must be hard going through a trial' (M02F). There was occasional reference to the trial judge's direction on this,³³ with one juror reminding the others that the judge 'said to bear in mind that people are nervous when they come to court so their body language may not be 100 per cent' (M03E). But the overall picture from the deliberations was one where false assumptions about body language were regularly used to bolster or cast doubt on credibility.

33. In line with Scottish Jury Manual guidance, jurors were told that when assessing credibility, they 'should, of course, make allowances for the fact that a witness may feel nervous and overawed by being in court'.

Where to next? Future directions for the role of demeanour in deliberations

As the above discussion illustrates, it was clear that in making credibility assessments, jurors in the present study drew on both the content of the evidence and the way it was delivered. In relation to content, discussions in the jury room considered the level of detail of the witness's account, its internal and external consistency and the extent to which the witness had any obvious motivation to lie. Jurors also sometimes used personal experience to assess the plausibility of a witness's claims. In these respects, the factors that jurors looked to were broadly in line with the factors research literature suggests are often appropriate credibility markers. In the assault trial, that use of personal experience sometimes appeared to help other jurors to evaluate credibility and 'put themselves in the witness's shoes'. However, the way in which jurors looked to the content of the evidence to assess veracity was not entirely unproblematic, especially in the rape trial. Here, jurors were sometimes reluctant to accept legitimate indicators of credibility, seeing the complainant's account as *too* detailed to be credible or the consistency of her account with the physical evidence as suspicious. Several jurors also attributed to the complainant a motivation to fabricate a rape claim – namely that she was seeking revenge for the accused's failure to resume their relationship – despite there being little or no evidence to support the plausibility of this in the circumstances of the case.

Moreover, in respect of jurors' assessment of the perceived manner of delivery (as distinct from its content), and its relevance to credibility, there was further cause for concern. Jurors often had clear – and misguided – expectations about the extent to which a rape complainant 'ought' to be emotional when testifying, expectations that were notably not placed on the male complainant in the assault case. In both trials, jurors looked to body language more widely, making false assumptions about the relationship between veracity and factors such as the direction of the witness's gaze and hand movements. In several juries, we observed these false assertions being made with marked confidence by jurors, who stated for example that the witness's body language 'stuck out' or was 'obviously' a sign of guilt. Another juror claimed they were 'quite familiar with body language', but proceeded to place weight on the fact that witness had 'looked to the left'. We also saw evidence of these false assumptions apparently being accepted by and influential upon peers within the deliberations.

Though extrapolation from the experimental to the trial context must be undertaken with caution, these findings raise crucial questions regarding the extent to which jurors' assessments of witness credibility, particularly where driven by cues relating to delivery, are robust and reliable. The implication of jurors placing unfounded weight on such cues is that we risk inaccurate verdicts. This can act in both directions, resulting in either wrongful conviction or acquittal, but the disproportionate focus on these factors in the rape trial suggests that, in that arena at least, expectations of emotionality interact with existing misconceptions about sexual violence and misplaced concerns about false reporting to reduce the prospects of conviction.

As noted above, this is not the only way in which jurors' intuitive reliance on dubious demeanour cues to assess credibility based on their own (non-specialist) assumptions about what truth-telling looks like can produce results that systematically impact upon particular cohorts. Members of ethnic minority groups, those required to navigate communication challenges, or those who have experienced traumatic events are amongst the participants that research has shown may face particular difficulty in being able to conform to observers' expectations in the witness box, and so be unfairly judged as lying. The significance of that in terms of access to justice in a context in which these are also often amongst the most disadvantaged and vulnerable populations within courtroom spaces should not be underestimated.

What can – or should – be done to better equip jurors to perform their task? The Scottish Jury Research did not set out to test the effect of particular interventions on jurors' approaches to credibility assessments, or the comparative weight given by jurors to different potential cues to credibility. Exploring these questions may prove a fruitful avenue for future research, but there is an existing research literature

which can be drawn upon in considering strategies for addressing the problems evidenced in these mock deliberations.

One approach might be to look to the growing body of literature demonstrating that the detection of deception can be improved to some degree using cognitive techniques. Some studies have achieved this by increasing the cognitive load of the speaker to make their task of maintaining deception more difficult, for example by encouraging them to elaborate, asking them to give their account in reverse chronological order or asking unexpected questions (Mac Giolla and Luke, 2021: 388; Vrij et al., 2017: 6). Others have focused more on training the observer (Driskell, 2012: 722; Hauch et al., 2016: 310), for example by introducing them to more specialist techniques³⁴ or giving them feedback on their performance (Driskell, 2012: 722; Hauch et al., 2016: 310). The difficulty with such initiatives is that it is hard to see how they could be incorporated into the courtroom. There is limited scope to use any of the cognitive techniques that focus on the speaker, given the evidential and procedural norms of testimony-giving. Those techniques that focus on increasing the specialism of observers and/or giving feedback to hone their skills are not feasible (or at least not without substantial investment) for jurors who do not have the opportunity to undertake detailed pre-trial training and who only sit on a single case (Denault and Dunbar, 2019: 924).

The impracticability of implementing any of these solutions has led some commentators to suggest that the best option may be to screen all witnesses from the jury so that they can hear but not see them, as a mechanism by which to reduce reliance on inappropriate behavioural cues (Blumenthal, 1993: 1202; O'Regan, 2017: 448). But logistical challenges aside, this approach would also not eliminate all of the concerns at issue. Some of the questionable cues that were relied upon in jury deliberations in our study, including a witness's speech hesitations and emotional 'tone', would not be addressed by placing witnesses behind a screen. This move could, moreover, have unintended side-effects such as preventing jurors with hearing disabilities from relying on lip reading. What is more practical, perhaps, is to look at whether suitable jury directions might help jurors to assess credibility appropriately.

Previous research has demonstrated that changes to judicial directions, particularly where designed to structure the deliberative process or to counter misconceptions through the provision of background information pertinent to jurors' assessment of evidence, can have a productive impact. Coyle and Thomson (2014: 484), for example, found – albeit in an experimental study far removed from trial reality – that directing mock jurors on the factors associated with veracity was successful in improving understanding. Ellison and Munro (2009b) found that giving educational guidance to mock jurors (either via judicial direction or expert testimony) impacted on the extent to which they focused on the apparent emotionality of the witness (a rape complainant) and the ways in which they interpreted this as bolstering or undermining her overall credibility. More specifically, they found that jurors who were informed about the range of reactions to victimisation and traumatic recall were less likely to refer to the complainant's emotional demeanour and – when the issue was raised – were more likely to offer explanations for what might account for her apparent calmness and lack of observable distress (Ellison and Munro, 2009b: 368). This was supplemented by jurors' responses to post-deliberation questionnaires, which demonstrated that after educational guidance they were less likely to report that it would have influenced their verdict decision if the complainant had been more obviously distressed when giving her testimony than those who had not been given this instruction (Ellison and Munro, 2009b: 370). Research in other – related – contexts suggests that jury directions can be effective in sensitising jurors to have regard to appropriate cues when assessing witness testimony, most notably in relation to eyewitness identification evidence (Leverick, 2016).

34. Especially the cognitive techniques discussed above – a meta-analysis found that training combined with the use of cognitive techniques improved the rate of deception detection to an average of 75% (Mac Giolla and Luke, 2021: 390).

The evidence relating to judicial directions is not, however, universally positive. Nitschke et al. (2022), for example, found that a judicial instruction about the likely effects of trauma on witness testimony resulted in jurors rating *all* complainer testimony as less credible, regardless of the degree of emotion the complainer displayed. The authors speculate that this was because the instruction left jurors uncertain about how they *should* evaluate credibility, resulting in a general scepticism towards the witness. It is difficult to assess this claim as the content of the direction is not included in the published paper. It should also be said that the experiment was conducted entirely online on the basis of written materials. Nonetheless, crafting judicial directions that usefully inform the processes and content of deliberations will be complicated. As we have already seen, detecting deception on the basis of behavioural cues is a highly complex and specialist task.

An obvious starting point is that we should certainly *not* be specifically directing jurors – as judges currently do in the Scottish courts – to pay attention to factors such as body language in assessing witness credibility. Beyond that, it is clear that simply directing jurors to ignore behavioural cues is unlikely to be productive (and may even, perversely, turn their attention more towards them). There are also some findings from experimental research that might assist us in embarking on the process of developing more effective and appropriate guidance. It has consistently been shown, for example, that jurors are more likely to follow instructions if it is explained why they are given (Leverick, 2016: 581). This suggests that directions should explain to jurors *why* they should not take particular factors into account in their credibility assessments (such as eye contact), by pointing to the research evidence that they have no relationship with veracity. Directions could also do more to point jurors towards factors that *are* more robustly established to be bases for assessing credibility (for example in respect of external consistency with other evidence),³⁵ and to draw their attention explicitly and directly to cultural biases and normative assumptions they might harbour that, left unaddressed, might lead them to misinterpret others' behaviours for various reasons.

The timing of such directions is also likely to be important. There is research demonstrating that giving directions at the start of the trial, rather than at the end, is more effective (Chalmers and Leverick, 2018: 27–31). This is especially pertinent in relation to credibility assessment because of the danger of tunnel vision (Porter and ten Brinke, 2010: 127). Jurors, rather than passively absorbing all the evidence as it is presented to them, instead settle on a 'story' that makes sense to them relatively early in the proceedings. Once an initial judgment about credibility has been made, it is very potent and affects the manner in which all remaining evidence is interpreted (Pennington and Hastie, 1992). Research has shown, moreover, that such judgments are often made quickly and if the outcome is negative, jurors will not pay sufficient attention to other, often more reliable, cues such as the external consistency of accounts (Porter et al., 2007). For example, Porter et al. (2010: 484) found that where the facial appearance of a defendant was considered to be 'untrustworthy', mock jurors required significantly more items of incriminating evidence for a guilty verdict than where the facial expression was rated as 'trustworthy'.

In conclusion, then, the timely introduction of well-crafted directions around credibility assessment will not render jurors experts at lie detection. It cannot rule out the possibility of inaccurate credibility assessments, with a resultant impact on verdict outcomes. Nor will it cure the broader problems evidenced by our mock jury deliberations in the rape trial, which require ongoing initiatives to combat the well-evidenced problem of rape myths (Chalmers et al., 2021a). However, such directions may significantly assist in generating the more cautious and circumspect approach from jurors that research clearly suggests would be merited in this context and may reduce the prospects of the most egregious and systemic denials of justice due to assessments of witness demeanour. Rigorous testing of the impact of any such directions in as realistic a setting as is possible – including collective deliberation – will, of course, also be important in assessing the extent to which this is the case and in identifying their appropriate style and content.

35. For two attempts to draft appropriate jury directions see Vrij and Turgeon (2018: 241–243); Denault et al. (2020: 7).

Acknowledgments

We are hugely grateful to our colleagues in Ipsos MORI Scotland with whom we conducted the Scottish Jury Research: Rachel Ormston (who led the project) and Lorraine Murray.


Declaration of conflicting interests


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


Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Scottish Government and undertaken by the authors and others (the Scottish Jury Research).

ORCID iDs

James Chalmers  <https://orcid.org/0000-0002-5423-8917>

Fiona Leverick  <https://orcid.org/0000-0001-5245-9874>

Vanessa E. Munro  <https://orcid.org/0000-0002-2552-9651>

References

- Baillot H, Cowan S and Munro VE (2014) Reason to disbelieve: Evaluating the rape claims of women seeking asylum in the UK. *International Journal of Law in Context* 10(1): 105–139.
- Bennett MW (2015) Unspringing the witness memory and demeanor trap: What every judge and juror needs to know about cognitive psychology and witness credibility. *American University Law Review* 64(6): 1331–1376.
- Blumenthal JA (1993) A wipe of the hands, a lick of the lips: The validity of demeanor evidence in assessing witness credibility. *Nebraska Law Review* 72(4): 1157–1204.
- Bond CF and DePaulo BM (2006) Accuracy of deception judgments. *Personality and Social Psychology Review* 10(3): 214–234.
- Chalmers J and Leverick F (2018) *Methods of Conveying Information to Jurors: An Evidence Review*. Edinburgh: Scottish Government.
- Chalmers J, Leverick F and Munro VE (2020) Three distinctive features, but what is the difference? Key findings from the Scottish Jury Project. *Criminal Law Review* 2020: 1012–1033.
- Chalmers J, Leverick F and Munro VE (2022) Beyond doubt: The case against ‘not proven’. *Modern Law Review* 85(4): 847–878.
- Chalmers J, Leverick F and Munro VE (2021a) The provenance of what is proven: Exploring (mock) jury deliberation in Scottish rape trials. *Journal of Law and Society* 48(2): 226–249.
- Chalmers J, Leverick F and Munro VE (2021b) Why the jury is, and should still be, out on rape deliberation. *Criminal Law Review* 2021: 759–771.
- Coyle IR and Thomson DM (2014) Opening up a can of worms: How do decision-makers decide when witnesses are telling the truth? *Psychiatry, Psychology and Law* 21(4): 475–491.
- Dancu CV, Riggs DS, Hearst-Ikeda D, Shoyer BG and Foa EB (1996) Dissociative experiences and post-traumatic stress disorder among female victims of criminal assault and rape. *Journal of Traumatic Stress* 9(2): 253–267.

- Denault V and Dunbar NE (2019) Credibility assessment and deception detection in courtrooms: Hazards and challenges for scholars and legal practitioners. In: Docan-Morgan T (ed) *The Palgrave Handbook of Deceptive Communication*. Cham: Springer International Publishing, 915–935.
- Denault V, Dunbar NE and Plusquellec P (2020) The detection of deception during trials: Ignoring the nonverbal communication of witnesses is not the solution – A response to Vrij and Turgeon. *International Journal of Evidence and Proof* 24(1): 3–11.
- Denault V, Jupe LM, Dodier O and Rochat N (2017) To veil or not to veil: Detecting lies in the courtroom. A comment on Leach et al. *Psychiatry, Psychology and Law* 24(1): 102–117.
- DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Charlton K and Cooper H (2003) Cues to deception. *Psychological Bulletin* 129(1): 74–118.
- Diamond SS, Rose MR, Murphy B and Smith S (2006) Juror questions during trial: A window into juror thinking. *Vanderbilt Law Review* 59(6): 1925–1972.
- Driskell JE (2012) Effectiveness of deception detection training: A meta-analysis. *Psychology, Crime & Law* 18(8): 713–731.
- Ellison L and Munro VE (2009a) Reacting to rape: Exploring mock jurors' assessments of complainant credibility. *British Journal of Criminology* 49(2): 202–219.
- Ellison L and Munro VE (2009b) Turning mirrors into windows: Assessing the impact of (mock) juror education in rape trials. *British Journal of Criminology* 49(3): 363–383.
- Ferguson P (2014) Eyewitness identification evidence. In: Chalmers J, Leverick F and Shaw A (eds) *Post-Corroborator Safeguards Review: Report of the Academic Expert Group*. Edinburgh: Post-Corroborator Safeguards Review, 44–66.
- Foa FB (1997) Trauma and women: Course, predictors, and treatment. *Journal of Clinical Psychiatry* 58(suppl 9): 25–28.
- Glaser BG and Strauss AL (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Global Deception Research Team (2006) A world of lies. *Journal of Cross-Cultural Psychology* 37(1): 60–74.
- Grigel KB (2019) Credibility interrogatories in criminal trials. *Stanford Law Review* 71(2): 461–511.
- Gross SR and Shaffer M (2012) *Exonerations in the United States, 1989–2012: Report by the National Registry of Exonerations*.
- Hartwig H and Bond CF (2014) Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology* 28(5): 661–676.
- Haskell L and Randall M (2019) *The Impact of Trauma on Adult Sexual Assault Victims*. Ottawa: Department of Justice Canada.
- Hastie R, Penrod SD and Pennington N (1983) *Inside the Jury*. Cambridge, MA: Harvard University Press.
- Hauch V, Sporer SL, Michael SW, et al. (2016) Does training improve the detection of deception? A meta-analysis. *Communication Research* 43(3): 283–343.
- Hutchins RM (2014) You can't handle the truth! Trial juries and credibility. *Seton Hall Law Review* 44(2): 505–556.
- Judicial Institute for Scotland (2021) *Jury Manual*. Edinburgh: Judicial Institute for Scotland.
- Kline NK, Berje DS, Rhodes CA, Steenkamp MM and Litz BR (2021) Self-blame and PTSD following sexual assault: A longitudinal analysis. *Journal of Interpersonal Violence* 36(5–6): NP3153–NP3168.
- Leverick F (2016) Jury instructions on eyewitness identification evidence: A re-evaluation. *Creighton Law Review* 49(3): 555–587.
- Leverick F (2020) What do we know about rape myths and juror decision making? *International Journal of Evidence and Proof* 24(3): 255–279.

- Lim A, Young RL and Brewer N (2022) Autistic adults may be erroneously perceived as deceptive and lacking credibility. *Journal of Autism and Developmental Disorders* 52(2): 490–507.
- Mac Giolla E and Luke TJ (2021) Does the cognitive approach to lie detection improve the accuracy of human observers? *Applied Cognitive Psychology* 35(2): 385–392.
- Mann S, Vrij A and Bull R (2004) Detecting true lies: Police officers' ability to detect deceit. *Journal of Applied Psychology* 89(1): 137–149.
- Marx BP, Forsyth JP, Gallup GG, Fusé T and Lexington JM (2008) Tonic immobility as an evolved predator defense: Implications for sexual assault survivors. *Clinical Psychology: Science and Practice* 15(1): 74–90.
- Mason F and Lodrick Z (2013) Psychological consequences of sexual assault. *Best Practice & Research Clinical Obstetrics & Gynaecology* 27(1): 27–37.
- McKimmie BM, Masser BM and Bongiorno R (2014) Looking shifty but telling the truth: The effect of witness demeanour on mock jurors' perceptions. *Psychiatry, Psychology and Law* 21(2): 297–310.
- Meissner CA and Kassin SM (2002) 'He's guilty!': Investigator bias in judgments of truth and deception. *Law and Human Behavior* 26(5): 469–480.
- Meixner JB (2012) Liar, liar, jury's the trier? The future of neuroscience-based credibility assessment in the court. *Northwestern University Law Review* 106(3): 1451–1488.
- Minzner M (2008) Detecting lies using demeanor, bias, and context. *Cardozo Law Review* 29(6): 2557–2582.
- Nicolson D (2014) Truth and demeanour: Lifting the veil. *Edinburgh Law Review* 18(2): 254–259.
- Nitschke FT, McKimmie BM and Vanman EJ (2019) A meta-analysis of the emotional victim effect for female adult rape complainants: Does complainant distress influence credibility? *Psychological Bulletin* 145(10): 953–979.
- Nitschke FT, McKimmie BM and Vanman EJ (2022) The effect of trauma education judicial instructions on decisions about complainant credibility in rape trials. *Psychology, Public Policy, and Law*. Advance online publication. DOI: 10.1037/law0000353.
- O'Regan D (2017) Eying the body: The impact of classical rules for demeanor credibility, bias, and the need to blind legal decision makers. *Pace Law Review* 37(2): 379–454.
- Ormston R, Chalmers J, Leverick F, Munro V and Murray L (2019) *Scottish Jury Research: Findings from a Large-Scale Mock Jury Study*. Edinburgh: Scottish Government.
- Pennington N and Hastie R (1992) Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology* 62(2): 189–206.
- Porter S, McCabe S, Woodworth M, et al. (2007) Genius is 1% inspiration and 99% perspiration? Or is it? An investigation of the effects of motivation and feedback on deception detection. *Legal and Criminological Psychology* 12(2): 297–309.
- Porter S and ten Brinke L (2010) The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology* 15(1): 57–75.
- Porter S, ten Brinke L and Gustaw C (2010) Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law* 16(6): 477–491.
- Qureshi AM (2014) Relying on demeanour evidence to assess credibility during trial: A critical examination. *Criminal Law Quarterly* 61(2): 235–267.
- Rand JW (2000) The demeanor gap: Race, lie detection, and the jury. *Connecticut Law Review* 33(1): 1–76.
- Redmayne M (2006) Theorising jury reform. In: Duff A, Farmer L, Marshall S and Tadros V (eds) *The Trial on Trial Volume 2: Judgment and Calling to Account*. Oxford: Hart Publishing, 99–116.

- Roberts A (2004) The problem of mistaken identification: Some observations on process. *International Journal of Evidence and Proof* 8(2): 100–119.
- Schauer M and Elbert T (2010) Dissociation following traumatic stress: Etiology and treatment. *Zeitschrift für Psychologie/Journal of Psychology* 218(2): 109–127.
- Smith L (2012) The ring of truth, the clang of lies: Assessing credibility in the courtroom. *University of New Brunswick Law Journal* 63: 10–37.
- Snook B, McCardle MI, Fahmy W, et al. (2017) Assessing truthfulness on the witness stand: Eradicating deeply rooted pseudoscientific beliefs about credibility assessment by triers of fact. *Canadian Criminal Law Review* 22(3): 297–306.
- Sporer SL and Schwandt B (2007) Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law* 13(1): 1–34.
- Thomas C (2020) The 21st century jury: Contempt, bias and the impact of jury service. *Criminal Law Review* 2020: 987–1011.
- Uono S and Hietanen JK (2015) Eye contact perception in the West and East: A cross-cultural study. *PLOS ONE* 10(2): e0118094.
- Uviller HR (1993) Credence, character, and the rules of evidence: Seeing through the liar's tale. *Duke Law Journal* 42(4): 776–839.
- van Doorn J and Koster NN (2019) Emotional victims and the impact on credibility: A systematic review. *Aggression and Violent Behaviour* 47: 74–89.
- Vrij A, Fisher RP and Blank H (2017) A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology* 22(1): 1–21.
- Vrij A and Hartwig M (2021) Deception and lie detection in the courtroom: The effect of defendants wearing medical face masks. *Journal of Applied Research in Memory and Cognition* 10(3): 392–399.
- Vrij A, Hartwig M and Granhag PA (2019) Reading lies: Nonverbal communication and deception. *Annual Review of Psychology* 70: 295–317.
- Vrij A and Mann S (2001) Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology* 15(2): 187–203.
- Vrij A and Turgeon J (2018) Evaluating credibility of witnesses – are we instructing jurors on invalid factors? *Journal of Tort Law* 11(2): 231–244.
- Wessel EM, Bollingmo GC, Sønsteby C, et al. (2012) The emotional witness effect: Story content, emotional valence and credibility of a male suspect. *Psychology, Crime & Law* 18(5): 417–430.
- Young W, Cameron N and Tinsley Y (1999) *Juries in Criminal Trials, Part Two, Volume 2: A Summary of the Research Findings*. Wellington: Law Commission.

Appendix

Appendix I: rape trial synopsis³⁶

The complainer (C) and accused (A) had been in an eight-month relationship, which ended approximately two months before the alleged offence took place. Both parties agreed that the break-up was cordial and that, in the intervening period, the complainer had made two short telephone calls to the accused to ask if he would like to go for a drink with her and her friends (which he declined). There

36. Though modified as appropriate to the Scottish context, some elements of the rape trial simulation drew on scripts prepared for previous ESRC research projects carried out by Vanessa Munro and Louise Ellison [RES 000-22-2374 and RES 000-22-4277].

had been no other contact between them during the period between the break-up and the events of the night in question.

On the night in question, the accused called at the complainer's home (which they previously shared) to collect some possessions. He and the complainer each drank a glass of wine and some coffee as they chatted. A few hours later, as the accused made to leave, the two kissed. It was the prosecution's case that the accused then tried to initiate sexual intercourse with the complainer, touching her on the breast and thigh, and that the complainer made it clear that she did not consent to this by telling the accused to stop and pushing away his hands. The prosecution alleged that the accused ignored these protestations and went on to rape the complainer. The complainer testified that after the accused left her flat, she was shocked and traumatised. She initially telephoned her sister, but got no answer. After 40 min, she telephoned the police to report the rape.

When the accused was questioned by the police, he admitted that he had had sexual intercourse with the complainer, but maintained that all contact was consensual, and this was the approach taken by the defence. The accused testified that after he had consensual sexual intercourse with the complainer, he did not want to resume a relationship with her, but felt awkward about this and left the flat without speaking further to her.

A forensic examiner (W) testified that the complainer had suffered bruising to her inner thighs and chest and scratches to her breasts that were consistent with the application of considerable force, but that – as was not uncommon in cases of rape – she had sustained no internal bruising. The forensic examiner advised that the evidence available following her examination of the complainer was consistent with rape, but that she could not rule out alternative explanations for the injuries.

Sections of Video: Judicial Introduction to Jury; Prosecution Evidence (examination-in-chief of C; cross-examination of C by defence; and re-examination of C; plus examination-in-chief of W; cross-examination of W by defence); Defence Evidence (examination-in-chief of A; cross-examination of A by prosecution; and re-examination of A); Closing Speeches by Prosecution and Defence; Judicial Summing Up and Instructions to the Jury.

Appendix 2: assault trial synopsis

The complainer (C) and accused (A) had both been drinking in a bar in which they were regulars. They knew each other by sight. On the night in question, both had been drinking for several hours. The accused attended the bar with three friends. The complainer was with his partner (W). The accused left the bar to smoke outside. There was no-one else present. The complainer left the bar to go home with his partner, but she stopped to speak to a work colleague, meaning that he exited before her. Once outside, he collided with the accused, spilling the accused's drink.

What happened afterwards is unclear. The complainer's account (and the prosecution case) is that the accused took a knife from his pocket, lunged at the complainer and stabbed him in his left shoulder. The accused's case is that it was the complainer who produced the knife, which he managed to grab from the complainer during the course of a struggle and which he then used on the complainer in self-defence. The complainer's partner testified that she saw the accused push the complainer to the ground, but on cross-examination admitted that she could not have seen this as she had not left the bar at that point. She was adamant, however, that she saw the accused produce the knife.

It was agreed in evidence that both sets of fingerprints were found on the knife, which the complainer explained by testifying that he had picked up the knife after the accused had fled the scene. It was also agreed that the complainer sustained a knife wound to his left shoulder that required surgery, leaving a permanent scar of 10 cm in length and permanently impairing the movement of his arm.

Sections of Video: Judicial Introduction to Jury; Prosecution Evidence (examination-in-chief of C; cross-examination of C by defence; and re-examination of C; plus examination-in-chief of W and cross-examination of W by defence); Defence Evidence (examination-in-chief of A; cross-examination of A by prosecution; and re-examination of A); Closing Speeches by Prosecution and Defence; Judicial Summing Up and Instructions to the Jury.