



Royal Commission
into Institutional Responses
to Child Sexual Abuse

CHIEF EXECUTIVE OFFICER

21 December 2016

Professor Jane Goodman-Delahunty,
Professor Anne Cossins and
Ms Natalie Martschuk

By email

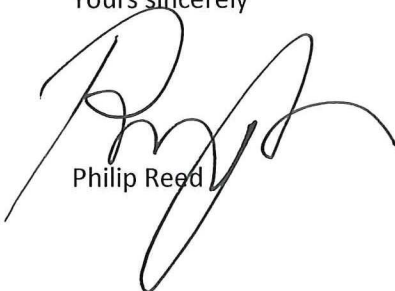
Dear Professor Goodman-Delahunty, Professor Cossins and Ms Martschuk

As you know, a number of submissions responding to the Royal Commission into Institutional Responses to Child Sexual Abuse's Consultation paper on criminal justice made comments on your research report, *Jury reasoning in joint and separate trials of institutional child sexual abuse: an empirical study* (Jury Reasoning Research). A number of witnesses in the public hearing in Case Study 46 also commented on aspects of the Jury Reasoning Research.

I enclose a document setting out relevant excerpts from submissions (the relevant part of one confidential submission is summarised rather than quoted) and from the transcript of the public hearing in Case Study 46.

I would appreciate receiving your response to these comments and your views on their implications for the Jury Reasoning Research. I would be grateful if you would provide your response and views in a form that could be published on the Royal Commission's website, whether by providing a separate written response or by inserting responses in the enclosed document.

Yours sincerely



Philip Reed

1. Extracts from submissions responding to the Royal Commission’s Consultation paper on criminal justice in relation to the Jury Reasoning Research

Below are extracts from four submissions responding to the Royal Commission into Institutional Responses to Child Sexual Abuse’s Consultation paper on criminal justice relating to concerns or queries about the Jury Reasoning Research. Relevant passages from a confidential submission are also summarised.

1.1 The Law Council of Australia

The Law Council of Australia’s submission is published on the Royal Commission’s website. It contains the following text in paragraphs 64 to 66 of the submission:

No inference can be drawn that juries are giving fair consideration to each count merely because they convict on some counts and acquit on others. There may be other reasons than lack of unfair prejudice why a jury might give an accused the benefit of the doubt in relation to some charges.

The Law Council has significant reservations with respect to the conclusions drawn by the researchers. In that regard, the following comments with respect to that research should be noted.

One form of ‘unfair prejudice’ is what the researchers call ‘character prejudice’ – a juror considers the accused a person of bad character and for that reason applies a lesser standard of proof. Such bad character might be established by previous incidents that the accused has admitted, or does not dispute. The evidence used in the mock jury research was not of that kind. There were simply multiple complainants. The fact that the mock juries do not appear to have adopted a lower standard of proof in those cases does not disprove the unfair prejudice hypothesis. Equally, the prejudice that a jury will over-value tendency evidence could not realistically be measured for the same reason – it is unlikely the jury were satisfied of one allegation and then used it to infer guilt in respect of others. It is more likely they engaged in coincidence reasoning (‘it is more likely one allegation is true because an independent person has made a very similar allegation’). As regards the danger of the jury over-valuing the evidence for coincidence reasoning, it is not apparent whether the research would be able to measure that. A juror saying, as some apparently did, that they needed more for proof beyond reasonable doubt in cases where tendency evidence was admitted may simply reflect the juror considering that there was in fact more evidence of guilt (because of the tendency evidence) and rationalising accordingly.

1.2 The Law Society of New South Wales

The Law Society of New South Wales’ submission is published on the Royal Commission’s website. The following text (references omitted) is extracted from pages 10 to 14 of the submission:

The Law Society urges caution with respect to reliance on the Jury Reasoning Study (JRS) report to justify reform, particularly with respect to joinder of counts and the reduction or removal of barriers to admissibility for tendency and coincidence evidence. For reasons we detail below, we strongly recommend the Royal Commission engage in significant peer review to allow legal and psychology experts to evaluate the findings of the JRS. The Royal Commission should take into account the large body of research within cognitive psychology relating to unconsciously biased reasoning. The absence of consideration of this research in the JRS is one of a number of reasons which calls for a broader review of this area of law.

...

The Jury Reasoning Study Report

The assertions made in the JRS report about a failure of justice because of failures in prosecutions raise questions about the utility of the JRS report referring to acquittals, convictions and the factual culpability of the defendant. This is because with no ground truth (because the mock trials are based on illustrative scripts) acquittals, convictions and culpability are potentially misleading indicators of efficacy. Even when used to measure comparative difference (whether a direction is given, a question trial is adopted, joinder or tendency appear etc.), these measures show a trend, but not whether support or otherwise for the trend has integrity.

No doubt a significant matter is the finding in the JRS "that verdicts were not based on impermissible reasoning or unfair prejudice to the defendant. These outcomes suggest that any fears or perceptions that tendency evidence - whether presented in a separate trial or a joint trial - is unfairly prejudicial to the defendant are unfounded". However, as factual culpability is a measure considered in relation to the dangers of impermissible reasoning, the [submitter] raises a concern that this measure may need further interrogation before reliance is placed on it.

Furthermore, the arguably narrow definitions of "impermissible reasoning" and "unfairness to the defendant" which the study adopts, create further concern in relation to the validity of the strong conclusions made in relation to the implications for the criminal justice system. Impermissible reasoning and unfair prejudice are much wider concepts.

Simplicity of case study trials

The Law Society notes that in each of the trials used in the JRS, the transcripts are very short, and do not reflect the actual length (and complexity) of many jury trials.

The Law Society notes that in the "limitations of the study" at p268 the authors state:

"Although the experiment was designed to replicate as closely as possible the experience and tasks of actual juries, we cannot exclude the possibility that the results obtained from this abbreviated experience may differ from those obtained in a real trial. For example, compared to the time available for juries to absorb, consider and discuss the evidence in a real trial, the mock jurors in this study performed under conditions that may have increased their cognitive load and made them more vulnerable to heuristic reasoning, confusion and errors than would be likely in a real trial, where the presentation of the evidence and deliberation typically proceed at a slower pace. Moreover, in an actual trial, a jury would have the opportunity to seek further direction or clarification from the judge, whereas that opportunity was not available in this trial simulation"

The Law Society is concerned that inadequate attention is paid to the impact of the abbreviated nature of the mock trials in the comments relating to the limitations of the study (i.e. significantly shorter, less complex and less emotionally charged than an actual trial). Given the relative brevity and mock conditions of the trials, the statements that the mock trials may have led to increased cognitive load and vulnerability to errors of confusion require peer review. We also note that the JRS's literature review omits reference to

significant research by Louise Ellison and Vanessa Munro, which focuses on jury reasoning issues in mock trial scenarios.

In addition, there appears to be an absence of acknowledging the effect of "unconscious bias". Jury reasoning towards "factual culpability" is treated as an indicator of a good strategy and underpins the authors' conclusions on "impermissible reasoning". This seems to differ from "unconscious bias", which arguably is the basis for the laws' resistance to admitting evidence of uncharged criminal allegations or joining charges.

It is for these reasons that the Law Society encourages caution with respect to changes to the law of evidence that challenge an accused's protections. There is a need to be alert to unintended consequences. Guidance on the effects of impermissible reasoning and unconscious biases within a ground truth environment would be particularly beneficial. Similarly, with respect to the operation of the presumption of innocence, we would encourage a study that could evaluate the impact of a defendant being from a particular category (e.g. priest) prominently featured in the Royal Commission.

Question Trails

We note the finding from the JRS that the use of question trails in the relationship evidence trial meant that 'the defendant was rated significantly less factually culpable'. However, the JRS indicates that the counts and the judges' instructions took over 'a significantly greater proportion of deliberation time' and 'the mock jurors perceived that they required less cognitive effort to evaluate the defence case'. These are encouraging signs and we urge the Royal Commission to explore further the benefits available through question trails, appropriately supported by judicial education. We understand that they have been used quite extensively in other jurisdictions, chiefly in New Zealand. If they improve jurors' ability to apply the presumption of innocence and reduce jurors' cognitive effort, they will be an asset.

However, we consider that any peer review should also address the implications of the finding that 'mock juries reported significantly more difficulty in understanding the charges in a joint trial when given a question trail than when they deliberated without one'; and that '[m]ock jurors who deliberated with the assistance of a question trail reported requiring significantly more cognitive effort to understand the charges than those jurors who deliberated without a question trail'. Question trails direct jurors to apply the prosecution's burden of proof to elements of the offence and focus on appropriate reasoning limitations. Self-reported indicators of increased cognitive effort by jurors given question trails raises a question as to whether there is an unidentified problem in the mock trials, given that normally we would expect that the use of question trails would improve understanding by jurors. We consider that a response to this from a peer review would be valuable.

...

Conclusion

...

The Law Society strongly urges a peer reviewed interdisciplinary report from legal and psychology experts who are disassociated from the researchers within or related to the Royal Commission to address:

- The strengths and limitations of the JRS' design, especially in terms of a ground truth basis;

- The strength and limitations of the JRS' assumptions;
- The strength and limitations of the conclusions, but also addressing specific questions, including:
 - the relationship of factual culpability to conclusions recommending, or founding recommendations for procedural or evidentiary change;
 - the relationship of unconscious bias to impermissible reasoning, and the extent to which the conclusions of the JRS acknowledge and address this;
 - the extent to which jurors' expressed assumptions or views may impact on their unconscious thoughts and impact on their deliberations and verdict;
 - whether the findings could be applicable to a trial in which the defendant did not give evidence, and
 - whether the conclusions would be valid in relation to trials where the plausibility of the defendant varied from that of the mock study. [References omitted.]

1.3 Bar Association of Queensland

The Bar Association of Queensland's submission is published on the Royal Commission's website. The following text (references omitted) is extracted from pages 6 and 7 of the submission:

The [Royal] Commission argues that the results of the jury reasoning research offer strong support for the view that the long held fears of prejudice to defendants from the admission of tendency evidence, or of allowing joint trials, is unfounded. We respectfully disagree. To the contrary, we contend that the results of the research demonstrate the opposite.

The results shown in the discussion paper at Figure 10.1 record the conviction rates for the four different trial types. The trial types were:

- (a) separate trial (where only a single complainant gave evidence)
- (b) relationship evidence (where a single complainant gave evidence but also gave evidence of uncharged acts)
- (c) tendency evidence (where the charge related only to a single complainant but evidence of other alleged victims was admitted), and
- (d) joint trial (where multiple charges relating to different complainants were tried together).

The results are recorded for the different types of offences, namely, non-penetrative offences and penetrative offences.

The results demonstrate that, where the jury considered only the evidence of the complainant, i.e., in the separate trial and relationship evidence trial, the conviction rates were low, namely: 11% (non-penetrative) and 0% (penetrative) in the separate trial; and 8% (non-penetrative) and 0% (penetrative) in the relationship evidence trial. The inescapable conclusion is that this resulted because of focus upon whether specific acts were proved beyond reasonable doubt.

By contrast, the conviction rates in cases where the jury heard evidence of the allegations of other complainants was high: 63% (non-penetrative) and 63% (penetrative) in the tendency.

In the tendency evidence trial, the charges considered by the jury remained the same as for the separate trial and relationship evidence trial. The tendency evidence (of similar conduct against two other boys) was not capable of proving the specific acts charged. The juries were instructed they could only use that evidence, if accepted, to reason that the defendant had a tendency to have a sexual interest in young boys, had a tendency to engage in sexual activity with young boys, and had a tendency to use his position of authority to access young boys in order to engage in sexual activity with them.

The inescapable conclusion is that the higher conviction rate was influenced by the other similar allegations led in evidence. Not unexpectedly, the conviction rates in the joint trials were similar. It is our view these results vindicate the fears of unfair prejudice to defendants expressed in the examples from the High Court set out above.

The discussion paper records the view of the researchers that the jury verdicts were logically related to the probative value of the evidence, that, as the inculpatory evidence was increased, conviction rates did too, that the credibility of complainants was enhanced by evidence from independent witnesses, and that little evidence was found that verdicts were based on impermissible or prejudicial jury reasoning. The additional evidence referred to was of course the evidence of other similar complaints against other complainants. As explained above, it is our view that, where none of the additional evidence could logically help prove the specific acts alleged by other complainants, it was simply the tendency reasoning which drove the convictions. That is, because they believed the defendant to be sexually attracted to boys, they were prepared to find specific acts were proved whereas, without knowledge of that attraction, the same allegations were not proved. [References omitted.]

1.4 The Law Society of New South Wales Young Lawyers Criminal Law Committee

The Law Society of New South Wales Young Lawyers Criminal Law Committee submission is published on the Royal Commission's website. The following text (references omitted) is extracted from pages 9 and 10 of the submission:

Jury Reasoning Research

One of the assumptions underpinning our legal system is that jurors obey directions. It is apparent from the research conducted by Goodman, Delahunty, Cossins and Martschuk for the Royal Commission that courts and legislators have consistently underestimated the ability of jurors to separate counts and evaluate evidence. While the Committee recognises the special dangers attaching to tendency and coincidence evidence, this research suggests that these dangers have been perhaps overstated. However, the Committee submits that further research in this area, and in particular a thorough peer review of the study, is appropriate to ensure that any changes to the law in this respect have a sound empirical basis.

Of particular interest are the findings by the study on the insignificance of the 'joinder effect'. That the mock jurors' definition of 'beyond reasonable doubt' was a certainty of under 90% in separate trials and over 90% in joint trials indicates that rather than lowering the threshold for conviction, joint trials increase the difficulty for the prosecution of securing a conviction. Of further note was the finding that jurors were more likely to engage in

impermissible reasoning in separate trials without tendency evidence, than they were in separate or joint trials with tendency evidence. In light of this, the Committee agrees with Counsel Assisting that there may be opportunities for reform in this area. The Committee notes that this is a complex area of law and recommends that any proposals to amend the Evidence Act be referred to the Australian Law Reform Commission.

1.5 Confidential submission

A confidential submission, which has not been published on the Royal Commission's website, expressed a number of concerns about the Jury Reasoning Research, which can be summarised as follows.

- The accumulation prejudice or effect arises from the effect of the jury being made aware of multiple allegations, not simply how many charges or counts there are or the number of witnesses called. Comparing trials with the same evidence but different numbers of charges (the tendency trials with two counts and the joint trials with six counts) and versions of the joint trial with different numbers of witnesses (four or six) does not test this.
- The report states that 'the factual culpability and conviction rates for the count based on the weakest evidence were not significantly elevated in the joint trial compared to the tendency evidence trial', but there were no conviction rates in the tendency trial for the weakest claim as it was not the subject of a charge. Factual culpability ratings were obtained in the post-trial survey of jurors. It is not surprising that there was little difference in the ratings relating to the 'weak' complainant between the two trials because there was no difference between the evidence in the different trials other than the leading of some further witnesses relating to another complainant in one version of the joint trial. There was no separate trial data for the 'weak case' to provide a valid point of comparison.
- The findings in relation to character prejudice are based on responses to the question in the post-trial questionnaire about how convincing the defendant was. The mean response in the basic, tendency and joint trials was the same and the authors conclude that this suggests the jurors were not engaging in impermissible reasoning on the basis of character prejudice, although the results cannot rule out this possibility. The result is surprising given the significant differences in conviction rates between the trials. One possibility is that the question was perceived to be about presentation rather than the weight the juror gave to the defendant's evidence.
- In relation to the analysis of juror comments and deliberations, the study found no juries in the tendency or joint trials impermissibly reasoned on the basis of character evidence. There is a severe limitation in the analysis because it considers only statements made during the course of deliberations. It was not uncommon for the jury to take an initial vote on each charge and, where there was unanimity, there was no deliberation and the reasoning process was not exposed. The study also found few instances of explicit permissible tendency reasoning, indicating that the analysis of the deliberations are not revealing the reasoning process.
- While the question asked of jurors as to the main reason for their verdict is designed to counter the limitations of the analysis of deliberations, a single quick response to the open ended question is unlikely to reveal a prejudicial reasoning process. The underlying rationale for the limits on the admission to tendency and coincidence evidence are the potential unconscious effects which are difficult to consciously correct for. While the findings clearly show the lack of an overt impermissible reasoning process in most cases, they do not provide evidence about the risk of impermissible reasoning processes at the level they are generally considered to occur.

2. Extracts from the transcript of Case Study 46

Below are extracts from the transcript of the public hearing in Case Study 46 in which certain expert witnesses gave evidence in relation to concerns or queries about the Jury Reasoning Research.

2.1 Mr Tim Game SC

Mr Tim Game SC gave evidence concurrently with Mr Peter Morrissey SC on 29 November 2016.

The following exchange occurred between Mr Game and the Chair:

MR GAME: Can I ask a question? I'm asking you a question, Commissioner. How does one test, in an exercise like that, the dangers of impermissible reasoning without feeding something in to the study that involves impermissible reasoning?

THE CHAIR: What was done was that their reasoning was analysed, and it was done, as you know, very thoroughly, and it didn't show problems. And I don't know how you would do the other.

MR GAME: But as I understand it, the distinction was between introduction of tendency evidence and introduction of other counts. That can be material of the same kind, but how do you test feeding in material that involves impermissible - that is to say, wrong - reasoning and then test that?

THE CHAIR: You mean irrelevant material?

MR GAME: Material that involves an impermissible train of thought towards reasoning as to guilt, because you have to test the false reasoning.

THE CHAIR: I'm not sure I'm understanding you.

MR GAME: Well, the assumption behind the questioning is that the tendency evidence all involves permissible reasoning. I'm positing the position that one introduces something that involves impermissible reasoning, that is to say, that doesn't pass the test.¹

2.2 Mr Peter Morrissey SC – 29 November 2016

Mr Peter Morrissey SC gave evidence concurrently with Mr Tim Game SC on 29 November 2016.

Mr Morrissey said:

I wanted to add - if I may take 30 seconds on this? The efficiency issues are one thing. The prejudice is another. What you can't test for in a mock trial where they know that they're not actually dealing with a real, damaged individual or a potentially dangerous accused is that you'll never get that emotional hijacking, which is what we're concerned about, because they will just simply know: I know I'm doing an intellectual exercise and that's what I'll do. Their heart will

¹ Transcript of T Game SC, Case Study 46, 29 November 2016, T24009:16-46.

never be troubled by the realities of a courtroom, which can be harrowing, and, if not properly managed, hijacking.²

2.3 Mr Stephen Odgers SC – 2 December 2016

Mr Stephen Odgers SC gave evidence concurrently with Mr Arthur Moses SC on 2 December 2016.

Mr Odgers said:

I would make the observation that it seems to me that one has to approach that research with some caution. For example, as I understand it, the research that was done with mock juries involved multiple allegations of child sexual abuse. I'm not aware that in any of the scenarios mock juries were actually informed that, for example, the accused had been convicted of an earlier offence or admitted that he had committed or that it was not in dispute.

So that highlights one point, which is that in those research scenarios, one of the greatest concerns about tendency evidence - that is, that a jury will be informed that the accused has, in fact, done that kind of act before - was not present in these scenarios, so that one risk of prejudice was not present. It was much more likely that the jury would be engaging in coincidence reasoning rather than in what I will call tendency reasoning. So that is an important qualification to the conclusions that have been drawn from that research report.

Another point to be made is - and I think Mr Morrissey made this on Tuesday - that the mock juries would have known that these were not real people; that, in fact, when they were being told about allegations it was a situation where it was unlikely that it would generate a kind of emotional response from awareness that a real person in front of you was in fact somebody who had engaged in child sexual abuse undoubtedly in the past.

So the concerns about emotional reactions, about undercutting the standard of proof as a result of awareness of somebody's previous significant misconduct, concerns about tendency to overweigh or give too much weight to such material - I have great concerns that the research would not, in the way it was conducted, have thoroughly elucidated those issues and that great caution should be taken in relying on the conclusions from that.³

Mr Odgers also said:

I am sorry, I accept that nothing was disclosed in the way they reasoned to show those kinds of prejudice. But what I'm saying to you is the information they were given was of a certain kind which, necessarily, in my view, meant that you wouldn't expect certain kinds of prejudice, because, for example, they were not told that the accused had, in fact, engaged in child sexual abuse on other occasions, which is one of the greatest concerns in this area; secondly, they weren't confronted by real world, as I've already pointed out and Mr Morrissey pointed out, so, therefore, you wouldn't expect an emotional response generated by such information; thirdly, just because juries don't, when they reason, appear to be engaged in prejudicial thinking or giving too much weight to material - one should be careful about this. One of the concerns is subconscious responses to information and that a juror might, with the best will in the world, be affected in a way which is prejudiced by information but then, in order to justify their conclusion that the person should be

² Transcript of P Morrissey SC, Case Study 46, 29 November 2016, T24010:23-33.

³ Transcript of S Odgers SC, Case Study 46, 2 December 2016, T24353:37-T24354:27.

convicted, will advance reasons explaining it which seem, on the face of it, entirely appropriate, and may not even be aware of the extent to which they've been prejudiced. I don't think studies of this kind will necessarily reveal those kinds of concerns.⁴

⁴ Transcript of S Odgers SC, Case Study 46, 2 December 2016, T24357:13-36.

Responses to Submissions to the Royal Commission’s Consultation Paper on Criminal Justice in Relation to the Jury Reasoning Research

Jane Goodman-Delahunty, Natalie Hodgson, Natalie Martschuk and Annie Cossins

Overview

1. A number of submissions to the Royal Commission’s Consultation Paper on Criminal Justice related to the Jury Reasoning Research (‘JRR’). We are very grateful to the parties who made these submissions, and thank the Royal Commission for this opportunity to respond to them. This document provides the authors’ responses to those comments. We will first make some general comments about common themes in the submissions, before turning to address each submission individually. All comments from the authors are contained in shaded boxes.

Peer review of the JRR

2. A number of submissions asserted that the JRR should be peer reviewed.
3. The Royal Commission published a response to this concern, stating the following:

The jury reasoning study was peer reviewed by three eminent law and social science academics, two of whom are from outside Australia. All three have published works on jury decision making and jury reform and were selected to provide a robust critique of both the study design and its findings.

All three assessed the research as suitable for publication. One reviewer noted that there are other ways of examining jury decision-making, and the limitations involved in only testing decision-making in the specific context of child sexual abuse offences, but accepted that the report’s approach was a valid approach. All comments received were passed onto the authors, including on issues such as:

- More comprehensively reflecting the results of prior research on the topic
- Better defining key terms, such as inter-case evidentiary conflation and character prejudice
- More clearly defining the different assessments of juror responses, such as self-reported cognitive effort
- Adding some relevant references to case law and other academic writings
- Typographical issues and format.

The authors amended the final report to reflect these comments where they considered it appropriate, and provided the Royal Commission with a response where they did not agree with the comments or did not consider amendments were

appropriate. The report was also subject to a review of the quantitative methods and conclusions by the Royal Commission's internal research team.

Distinguishing relevant prejudicial evidence from unfair prejudice

4. In a scholarly review of the psychology of evidence, in relation to the balancing test that judges must perform regarding potential unfair prejudice that may arise from relevant, probative, admissible evidence, Professors of Law Michael Saks and Barbara Spellman noted that “all relevant evidence is prejudicial in that it supports an argument for one side's interpretation of the facts.”¹ Their point is important to acknowledge at the outset because the JRR involved variations in the relevant prosecution evidence admitted at trial, all of which was prejudicial to the defendant. Accordingly, in interpreting the results of the JRR, we anticipated that the incriminating force or weight of the evidence against the defendant would increase in trial versions testing the admission of relevant and probative evidence of uncharged acts of abusive conduct by the defendant in a separate trial of a single complainant, or of relevant and probative evidence of acts of abuse against multiple complainants in a joint trial.
5. In the time available before publication of the JRR, we provided as comprehensive a set of analyses as was feasible, reporting the most critical outcomes addressing the key research questions. We acknowledge that there are other approaches and methods that can be used to analyse the online and the deliberation data, and we are continuing to analyse them. The challenge in interpreting the research outcomes is to distinguish the effects of these prejudicial and legitimate anticipated increases in the weight of the evidence against the defendant from those that are unfairly prejudicial, as it is only when relevant and probative evidence is unfairly prejudicial that a judge must exclude it from trial. Some submissions about the JRR did not appear to have taken this difference into account.

Types of unfair prejudice

6. The balancing test regarding the probative value versus the unfairly prejudicial quotient of relevant evidence applies generally in all types of criminal and civil cases, although this topic has not been as widely researched as have many other psychological aspects of the rules of evidence. In considering all types of cases and all possible types of evidence, Saks and Spellman provided three examples of modes of improper jury decision making that might arise and comprise unfair prejudice, for example, when a jury decision is reached on unreliable or emotional grounds rather than a permitted logical basis related to the relevant evidence. Although their review was published after the JRR research concluded, the framework they applied matches

¹ Michael J. Saks and Barbara A. Spellman, *The Psychological Foundations of Evidence Law*, (New York University Press, 2016) 63.

the approach we used in the JRR, and their perspective is helpful in considering the submissions in response to the JRR.

7. First, they described a reasoning process evoked by a sense of outrage and sympathy for the victim in response to inflammatory evidence, which they paraphrased as “Something terrible happened, so someone must pay.”² Next, they described an inappropriate reasoning process inflamed by emotional responses to negative information about a particular defendant, paraphrased as “This guy is terrible; he should pay.”³ The third type of impermissible reasoning that they distinguished is a form of mental contamination of the jury reasoning by means of which evidence relevant for one purpose is (sensibly) but impermissibly applied for another purpose, paraphrased as “The (inadmissible) evidence suggests he did it.”⁴
8. A number of submissions about the JRR appear to have misconstrued the scope of this research as encompassing any and all forms of impermissible reasoning in response to any and all forms of evidence, and in particular, evidence of the prior criminal history of the accused.

Unfair prejudice in joint trials

9. In the JRR, we did not examine all three types of potential impermissible reasoning suggested by Saks and Spellman because this was not a study of all possible types of unfair prejudicial reasoning in response to all possible types of prejudicial evidence. In other words, this study centred on *impermissible reasoning that might arise in joint trials*.
10. In cases of joinder, where multiple complainants testify against a single defendant, the concern that “This guy is terrible; he should pay” is the concern that judges must address. Based on specific hypotheses expressed by judges in cases in which the risk of unfair prejudice in joint trials was under consideration, we focused on the second of the three types of prejudice identified by Saks and Spellman. To examine the presence of this type of impermissible reasoning against the defendant in a joint trial, we differentiated three possible ways in which the additional negative information about the defendant’s previous abusive conduct might engender impermissible reasoning, namely (a) inter-case conflation of the facts; (b) accumulation prejudice; and (c) character prejudice.
11. In discussing rules of evidence that allow courts to exclude relevant evidence when the danger of unfair prejudice outweighs its probative value, Saks and Spellman listed three key questions:

² Ibid 64.

³ Ibid 68.

⁴ Ibid 69.

- (a) What is the actual probative value of the evidence?
 - (b) How probative will the jury think it is?
 - (c) Did the rulemakers, whether courts or legislatures, “get it right” when creating the rule?⁵
12. In relation to the actual probative value of the evidence, Saks and Spellman were careful to point out that “the probative value of a piece of evidence depends on what other evidence exists in a case.”⁶ Accordingly, the expectation about the actual probative value of the same evidence, when presented in the context of a different trial, is that its weight or value would shift as the context within which that evidence is assessed changes, not that it will remain invariant across all trials. In interpreting the JRR results, we took this principle into account.
 13. A number of submissions criticised the experimental design of the JRR because the key manipulated variable in the simulated trials was the presence or absence of tendency evidence. The JRR was not about the admissibility of this evidence. The tendency evidence presented in all simulated trials was admissible, relevant, probative and prejudicial. Curiously, none of the submissions engaged with the first question listed by Saks and Spellman regarding the actual probative value of the admitted evidence, for example about the focal complainant, and its appropriate weight as the context in which it was presented changed.
 14. Since the tendency evidence included in the simulated trials was admissible, the juries were entitled to use that evidence in reaching a verdict, and received guidance from the judge on how to use it. The focus of the JRR was the second question specified by Saks and Spellman, i.e., how juries respond to tendency evidence, and in particular, whether juries used this evidence in a permissible versus and impermissible way.
 15. The broader purpose of the JRR was to address the third question specified by Saks and Spellman, i.e., to generate information to determine whether judges’ views of jury use of tendency evidence in a joint trial were accurate. In other words, the value of empirical research such as the JRR is to inform and refine legal policy.⁷
 16. The baseline trial was a separate trial of a single complainant without any evidence of uncharged acts by the defendant or evidence from other witnesses about uncharged acts (Trial 1). In the relationship evidence trial, evidence from a single complainant about uncharged acts (Trial 2) was accompanied by a jury direction that uncharged acts could not be used as tendency evidence in another trial (Trial 3, Report pp. 359-364). In the separate trial with tendency evidence (Trial 5) and in the joint trial (Trial 10), allegations made by multiple witnesses and/or co-

⁵ Ibid 59.

⁶ Ibid 61.

⁷ Ibid 241.

complainants were accompanied by a judicial direction on the appropriate use of the tendency evidence (Trials 6, 7; Report, pp.365-371). In those trials, the juries received a direction to avoid improper reasoning based on the character of the defendant, i.e., to avoid use of this evidence as propensity evidence to establish guilt for the crime charged (Report, pp. 362-363, 370).

Unconscious bias

17. Several submissions suggested that the JRR failed to address unconsciously biased reasoning. The research design employed in the JRR including multiple different versions of a trial of institutional child sexual abuse took into account the potential that unconsciously biased reasoning might arise in response to those trial variations. This approach is the standard and most reliable scientific procedure used in experimental cognitive psychology to assess the presence of unconscious bias.⁸ The reason cognitive psychologists know about unconsciously biased reasoning is by conducting studies applying the experimental method applied in the JRR to expose such reasoning.
18. The presence of unconscious bias was assessed primarily by means of the random assignment of mock jurors to different experimental conditions in which they viewed different versions of the simulated trial, so that the outcomes in these different groups could be statistically compared. To avoid cueing the mock jurors to the issues of interest to the researchers, the mock jurors were not informed of the research questions of interest, nor were they aware that other trial versions existed, or how those other trials differed. Using this method, if reasoning was unconsciously biased, differences would have emerged in the series of common measures that compared ratings provided by groups of juries and jurors who were exposed to different versions of the trials.
19. The responses to the dependent measures set out in the Report, Appendices G-L, pages 324-342, were systematically compared to explore the extent of any unconscious bias inherent in the variations in the evidence presented in the different types of trials. Throughout the results sections, we reported the effect sizes, the results of significance tests and conducted multi-level modelling analyses to distinguish the influence of the individual jurors from the influence of the jury groups on the jury reasoning and decisions.
20. A second method used to explore unconscious bias in this study was the administration, 1-2 weeks in advance of the simulated trials, to all mock jurors, of three psychometrically validated scales designed assess the extent of individual juror attitudinal pre-dispositions or biases, namely, the Pre-trial Jury Attitude

⁸ Steven D. Penrod, Margaret Bull Kovera and Jennifer Groscup, 'Jury Research Methods' in Barry Rosenfeld and Steven D. Penrod (eds), *Research Methods in Forensic Psychology* (John Wiley & Sons, 2011).

Questionnaire (PJAQ), the Forensic Evidence Evaluation Bias Scale (FEEBS), and the Child Sexual Assault Knowledge Questionnaire (CSA-KQ). These measures formally assessed individual proclivities, attitudes and biases that particular jurors brought to the trials and the specific jury groups to which they were assigned. For example, one subscale of the PJAQ is *Conviction-proneness*, of particular interest because it is related to the fairness of a trial to the accused. Another sub-scale of the PJAQ is the belief in *Innate criminality*, of particular interest because of its relation to character prejudice.

21. Throughout the research analyses, we examined the extent to which these individual attitudes contributed to the decisions reached by individual jurors and by juries. By including these measures of individual juror bias in the JRR, we were able to analyse their influence as predictors of jury verdicts, their interactions with the evidence, and we were also able to control for their influence to ensure that these biases did not overwhelm or mask the effects of other independent manipulated variations in the trial evidence (see, e.g., Report, Table N, pp. 348-350).
22. We suspect that some submissions about unconscious cognitive biases may be referring to the potential influence of particular characteristics of a defendant. Numerous prior jury studies have researched the unconscious influence of many such factors, which have been designated as “estimator” variables in psycholegal research.
23. An “estimator” variable in the context of jury research is a factor that may influence the rate of conviction, but is a given or immutable feature in a case, such as the age, race, gender, socio-economic status, attractiveness, or number of prior convictions of the accused.⁹ For example, the scope of influence on jury decision making of five characteristics of defendants, namely their race, attractiveness, socioeconomic status, gender, and prior criminal convictions, was assessed in a substantial meta-analysis of 272 jury simulation studies.¹⁰ Of those five factors, only two, prior criminal record and socioeconomic status, had a small but significant effect. In other words, the vast number of studies on individual characteristics of the defendant yielded few unconscious biases from these sources of any magnitude, and any unconscious bias due to prior convictions and the socio-economic status of the defendant exerted far smaller effects than is often presumed. In planning the JRR, we took this body of research into account and cited the results of that meta-analysis (Report, pp. 61-62).

⁹ Margaret B. Kovera, ‘Evaluating the Validity and Importance of Juror and Jury Decision-Making Research (Paper presented at the Third International Conference on Empirical Studies of Judicial Systems: Citizen Participation Around the World, Taipei, Taiwan, 5-6 September 2014).

¹⁰ Dennis J. Devine and David E. Caughlin, ‘Do they Matter? A Meta-Analytic Investigation of Individual Characteristics and Guilty Judgments’ (2014) 20(2) *Psychology, Public Policy and Law* 109.

24. We also considered the advice of a leading jury scholar that research on variations that the court and litigants can control, designated as “system variables,” such as procedural decisions to allow a joint trial, are far more important to study than are variations in “estimator variables” that are beyond their control, and cannot be changed after the events in issue in a criminal case.¹¹
25. In sum, the suggestion that we failed to consider unconscious biases and considered only conscious jury reasoning processes is incorrect. Comments of this nature overlook and/or reflect a misunderstanding of the experimental method and design applied in the JRR. The findings and approaches used in the body of research on cognitive unconscious bias constituted the key premises underpinning the JRR design and procedures. In reaching their conclusions, the JRR researchers did not rely solely on oral comments made by jurors in the course of their deliberations and the written reasons for their verdicts provided by individual jurors on their written post-trial questionnaires.

The ecological validity of the research

26. A number of submissions centred on features of the JRR that did not replicate real world trials.
27. Responses to specific criticisms regarding these limitations of the study are provided below. Here, we offer a general comment about the framework for the assessment of the validity or integrity of experimental jury simulation research.
28. In experimental research, three different types of validity are relevant:
- Internal validity: The extent to which the effects detected in the study were caused by an independent variable in the study, rather than by biasing effects of unmeasured variables.¹²
 - External validity: The extent to which the findings of a research study will generalise to other people in similar situations.¹³ That is, the extent to which findings from this study will be replicated in trials with other complainants, with respect to other types of sexual abuse, and in other jurisdictions.
 - Ecological validity: The extent to which the findings of a trial simulation research study will generalise to real-life trials.¹⁴
29. The submissions about the JRR relate primarily to the third criterion, ecological validity.

¹¹ Ibid; Kovera, above n 9.

¹² American Psychological Association, *APA Dictionary of Psychology* (American Psychological Association, 2nd ed, 2015) 553.

¹³ Ibid 402.

¹⁴ Ibid 349.

30. Ecological validity and external validity are not the same thing. Professor Margaret Bull Kovera used the following example to distinguish the two:

Suppose, for example, that a jury researcher interviewed the jurors who served on the jury that heard evidence in the OJ Simpson murder trial about their evaluations of the DA evidence in the case and their beliefs about the influence of that evidence on their decisions. This hypothetical study has high ecological validity: the participants were real jurors, they heard real evidence presented live in court, they deliberated, and they made real decisions with real consequences. Yet, it is improbable that the findings from this hypothetical study would be generalizable because its descriptive methods would not provide any information about the causal relationships among variables. At its core, external validity is a question of whether the relationships among variables discovered in one setting, with one sample, at one time are the same relationships present in other settings, other populations, at other times.¹⁵

31. Professor Kovera has pointed out that “[a]uthors, reviewers and editors often conclude that the results of an unrealistic trial simulation will not generalize to real trial settings because of its low ecological validity.”¹⁶ However, “[t]hese statements mischaracterize the relationship between ecological validity and external validity.”¹⁷ Psychological research in other domains has provided “meta-analytic evidence that ecologically invalid research on aggressive behaviour generalizes to real world settings.”¹⁸
32. Therefore, ecological validity is just one criterion used to assess jury research, with the caveat that its importance is often exaggerated. It is improper to assume that the findings of trial simulation studies that do not precisely replicate the experiences of real-life jurors are invalid.

1. Extracts from submissions responding to the Royal Commission’s Consultation paper on criminal justice in relation to the Jury Reasoning Research

Below are extracts from four submissions responding to the Royal Commission into Institutional Responses to Child Sexual Abuse’s Consultation paper on criminal justice relating to concerns or queries about the Jury Reasoning Research. Relevant passages from a confidential submission are also summarised.

¹⁵ Kovera, above n 9, 23-24.

¹⁶ Ibid 22.

¹⁷ Ibid 23.

¹⁸ Ibid 23.

1.1 The Law Council of Australia

The Law Council of Australia's submission is published on the Royal Commission's website. It contains the following text in paragraphs 64 to 66 of the submission:

No inference can be drawn that juries are giving fair consideration to each count merely because they convict on some counts and acquit on others. There may be other reasons than lack of unfair prejudice why a jury might give an accused the benefit of the doubt in relation to some charges.

33. One manifestation of impermissible reasoning that might culminate in unfair prejudice in a joint trial is the uniform treatment of multiple counts against the defendant because, for example, jurors may be prone to reason that evidence of other bad behaviour by the accused establishes that he is "that kind of person." Our jury reasoning research centred on detecting the presence of unfair prejudice, and on assessing its influence on verdicts and other dependent measures. Thus, how juries responded to multiple counts, and in particular whether they treated them uniformly or not, was a key topic of interest.
34. The JRR did not explore in general whether fair consideration was given to each count. We agree that there may be reasons other than a lack of unfair prejudice for verdicts to acquit or to convict. We were seeking indications within juries and from individual jurors of unfair prejudice arising from impermissible reasoning. One of the indicators we examined to inform this question was the extent to which counts were discussed and evaluated separately in the four types of trials that we tested compared to the extent to which counts were treated uniformly in those types of trials.
35. Our research findings that the juries and individual jurors did not treat all counts in a uniform manner demonstrated that the verdicts reached were unlikely to be motivated by impermissible reasoning inflamed by emotional views that "This guy is terrible; he should pay."

The Law Council has significant reservations with respect to the conclusions drawn by the researchers. In that regard, the following comments with respect to that research should be noted.

One form of 'unfair prejudice' is what the researchers call 'character prejudice' – a juror considers the accused a person of bad character and for that reason applies a lesser standard of proof. Such bad character might be established by previous incidents that the accused has admitted, or does not dispute. The evidence used in the mock jury research was not of that kind.

36. The JRR did examine bad character in the form of previous incidents involving the accused that were not the subject of charges. Evidence of undisputed sexual acts by the accused is less likely to be included in joint trials.
37. For instance, in the relationship evidence trial, previous incidents between the complainant and the accused was provided in the form of photographs of the complainant taken by the defendant and uncharged acts of sexual indecency (Trials 2, 3 and 4). In another version of the separate trial of a single complainant, evidence was provided of previous incidents between the defendant and two other boys who were not complainants, namely four other incidents, three involving indecency and one involving penetration (Trials 5 and 6). The purpose of including these versions of the trial in the study was to test the extent to which evidence of previous abusive incidents prompted character prejudice, independently of any counts or charges regarding those previous incidents.
38. The assessment of the influence of other criminal behaviour in the form of prior convictions or admissions was a topic beyond the scope of the JRR. Moreover, as discussed above (para. 23), results of a recent meta-analysis on the effects on juries of prior convictions highlighted that this effect is smaller and weaker than many legal professionals anticipate.¹⁹

There were simply multiple complainants.

39. This description of the manipulation of the evidence tested in our jury reasoning research is not accurate. In three versions of a separate trial, there was a single complainant (Trials 1-6). In a fourth version of the trial, there were three complainants (Trials 7-10). The effect of the additional evidence of prior abusive incidents was compared both with and without multiple complainants.

The fact that the mock juries do not appear to have adopted a lower standard of proof in those cases does not disprove the unfair prejudice hypothesis.

40. The research question as to whether character prejudice arises in a joint trial is separate from the question of the standard of proof applied in a joint trial, and these questions were tested separately in this study, although some judges and lawyers have assumed that character prejudice leads to a change in the standard of proof applied. In our view, a different standard of proof could potentially be applied in a joint trial compared to a separate trial without any evidence of character

¹⁹ Devine and Caughlin, above n 10.

prejudice, and character prejudice could potentially arise in a joint compared to a separate trial without any change in the standard of proof applied.

41. The conclusion of the researchers is not that the absence of a change in the standard of proof alone disproves reliance on character prejudice. The conclusion of the researchers is that, in this study, the results provided scant, if any support for the hypothesis expressed by judges in published legal opinions that juries exposed to multiple complainants are (a) prone to apply a lower standard of proof because of (b) the presence of character prejudice.
42. In reaching this conclusion, the researchers did not rely solely on the fact that the standard of proof applied by the juries was higher in joint than in separate trials. We used several dependent measures in addition to verdict to make this assessment by applying quantitative methods of analysis (Report, Appendices G-L, pp. 324-342). These included measures of individual jurors' attitudes and beliefs about justice, their views of evidence and their knowledge and misconceptions about child sexual assault, and their responses to a series of questions about the evidence and the key witnesses.
43. In addition to the perceived *Criminal intent* of the defendant, the convincingness of the defendant, the defendant's factual culpability and conviction rates (verdicts), character prejudice was assessed by asking to the mock jurors whether the defendant posed a risk to other boys. The results did not show significantly elevated scores or a 'ceiling' effect on a 7-point scale, which is what one might expect if the evidence had prompted reasoning biased by character prejudice (e.g., Report pp. 101-102).
44. Further assessment of the juries' reasoning was provided by coding the content of the deliberations and applying quantitative methods to assess the coded data. The codes applied to the deliberation content are listed in the Report, Appendix M. The presence of character prejudice was assessed by comparing the frequency with which negative comments about the character of the defendant were made in the course of jury deliberations and the extent to which comments of that nature were causally related to the ultimate verdicts of juries.
45. Our final assessment of the extent to which jury reasoning was driven by character prejudice came from the convergent measures or triangulation of the outcomes derived from qualitative methods applied to analyse the content of the deliberations, as reported in Case Studies. Similarly, individual juror reliance on character prejudice was assessed by analysing the content of individual jurors' statements, made only in anonymous private written notes to the researchers, stating the reasons for their verdicts.

Equally, the prejudice that a jury will over-value tendency evidence could not realistically be measured for the same reason – it is unlikely the jury were satisfied of one allegation and then used it to infer guilt in respect of others.

46. Precisely what is intended by this criticism is somewhat unclear. We have taken this comment to mean one of two things, and our respective responses to each possible issue are set for the below.
47. Firstly, we have interpreted this comment as a criticism of the manner in which tendency evidence is operationalised in the simulated trials. As we described above (para. 39), tendency evidence was included in the trials in a number of different ways: firstly, we assessed the impact of tendency evidence that was not the subject of any charges against the defendant, provided by two boys who were called as independent witnesses (Trials 5 and 6); secondly, we assessed the impact of tendency evidence from multiple complainants in the same trial (Trials 7, 8, 9 and 10).
48. Our review of the jury deliberations revealed that in the joint trial, a number of the juries were convinced beyond a reasonable doubt that the accused had committed the offences against complainant with the strong evidence (Report, p. 197), thus this evidence was available for their use as tendency evidence in assessing the allegations brought by the other two complainants, as instructed by the trial judge.
49. Secondly, we interpreted this comment as an expression of general concern about the manner in which this study assessed whether jurors were reasoning appropriately with tendency evidence.
50. In the deliberation analysis, we examined correct and incorrect uses by juries of the tendency evidence (e.g., Report, pp. 136, 155, 199). We found that when juries were given tendency evidence directions, *none of the juries engaged in impermissible propensity reasoning* (Report, p. 136, 199). While several juries had difficulty understanding the tendency evidence direction (e.g., Report, pp. 155, 199), this often worked in favour of the accused, for example, because the juries misunderstood the direction and concluded that they could not use the tendency evidence in reaching a verdict. There was no evidence from the transcribed deliberations, that juries over-valued or mis-used the tendency evidence admitted in those trials. While this observation was based on the analysis of the articulated reasoning of the juries in each joint trial, other dependent measures used in the jury reasoning research confirmed the finding from the deliberation analyses that juries did not mis-use the tendency evidence.
51. In the joint trials, the three complainants' evidence varied in terms of their evidentiary strength, i.e., the evidence in support of the respective complainants involved: a weak case, a moderately strong case, and a strong case. If juries were

engaging in impermissible tendency reasoning, then the conviction rate for the complainant with the weak case would be undifferentiated from those of the other two complainants with stronger cases. However, the conviction rate for the case of the complainant with the weak evidence remained significantly lower (Report, p. 197). This finding indicated that juries were not according undue weight to the tendency evidence from the complainants with the strong and medium cases. Rather, the significantly lower conviction rate for the complainant with the weak case compared to the other complainants with moderately strong and strong evidence indicated that juries considered the counts separately and avoided impermissible tendency reasoning.

It is more likely they engaged in coincidence reasoning ('it is more likely one allegation is true because an independent person has made a very similar allegation'). As regards the danger of the jury over-valuing the evidence for coincidence reasoning, it is not apparent whether the research would be able to measure that.

52. Analysis of the jury deliberations revealed that some individual jurors did remark on the coincidence of multiple complainants, as shown in this excerpt from the deliberations of Jury 64 (Report, p. 146):

JUROR 6: May I ask you why you thought guilty on Count 5?

JUROR 1: I just think he would have no motivation, like, no reason for him to make up that story.

...

JUROR 3: Three people that don't know each other; it is such a coincidence.

However, the fact that one juror remarked on the coincidence did not prompt the jury as a group to adopt this reasoning to reach a collective verdict based on coincidence reasoning.

53. As discussed above, if juries engaged in impermissible coincidence reasoning, one would expect over-valuing of this evidence to impact the conviction rate for the complainant with the weak evidence, by elevating it to a level comparable to the conviction rates for the complainants with moderately strong and strong evidence. The significantly lower conviction rate for the complainant with the weak evidence indicated that juries considered the counts separately, and avoided impermissible coincidence reasoning; i.e., they did not conclude that because the defendant had abused the complainants with moderately strong and strong evidence that he also abused the complainant with the weak evidence.
54. Further, the JRR did examine the impact of independent corroboration of evidence of abuse. When an independent witness gave evidence that corroborated one complainant's allegations, as was the case in the separate trial with tendency evidence (Trials 5 and 6), the assessments of the criminal intent of the defendant,

the likelihood of factual culpability and guilt increased with respect to the focal complainant to a greater degree than when evidence of additional prior incidents was provided by that complainant himself as relationship evidence (Trials 2, 3, 4). However, the conviction rate and other measures of the defendant's culpability in the latter trial exceeded those in the separate trial, reflecting logical increases based on the greater incriminating force of the relevant and prejudicial evidence of the uncharged acts even in the absence of the independent corroborative witness.

A juror saying, as some apparently did, that they needed more for proof beyond reasonable doubt in cases where tendency evidence was admitted may simply reflect the juror considering that there was in fact more evidence of guilt (because of the tendency evidence) and rationalising accordingly.

55. Mock jurors were asked a neutral question: "What number between zero and 100% represents 'beyond reasonable doubt'?" Comparisons of mean responses across experimental conditions yielded statistically significant differences, showing an unconscious bias caused by the form of the trial i.e., a higher standard of proof was applied in tendency evidence and joint trials than in the separate trial (Report, p. 248).

1.2 The Law Society of New South Wales

The Law Society of New South Wales' submission is published on the Royal Commission's website. The following text (references omitted) is extracted from pages 10 to 14 of the submission:

The Law Society urges caution with respect to reliance on the Jury Reasoning Study (JRS) report to justify reform, particularly with respect to joinder of counts and the reduction or removal of barriers to admissibility for tendency and coincidence evidence. For reasons we detail below, we strongly recommend the Royal Commission engage in significant peer review to allow legal and psychology experts to evaluate the findings of the JRS.

56. See above (paras. 2-3) for summary of the peer review process applied to the JRR. The research was peer-reviewed by three independent psychology and law scholars, two of whom were trained in law.

The Royal Commission should take into account the large body of research within cognitive psychology relating to unconsciously biased reasoning. The absence of consideration of this research in the JRS is one of a number of reasons which calls for a broader review of this area of law.

57. See above (paras. 17-25) about the research methods used to assess unconscious bias.

The Jury Reasoning Study Report

The assertions made in the JRS report about a failure of justice because of failures in prosecutions raise questions about the utility of the JRS report referring to acquittals, convictions and the factual culpability of the defendant. This is because with no ground truth (because the mock trials are based on illustrative scripts) acquittals, convictions and culpability are potentially misleading indicators of efficacy. Even when used to measure comparative difference (whether a direction is given, a question trail is adopted, joinder or tendency appear etc.), these measures show a trend, but not whether support or otherwise for the trend has integrity.

58. In detection of deception studies, mock jurors are asked to determine whether a witness is telling the truth or lying. In these studies, the researchers typically manipulate ground truth as a variable; some witnesses are told to lie while others are asked to tell the truth.
59. In a real trial, there is no ground truth as a standard for whether a verdict is error free or biased, as the ground truth is unknown to the court and jury. Thus, whether ground truth is a helpful criterion for assessing the validity of the JRR is questionable.
60. We agree that the specific proportions of convictions, acquittals, hung juries, or ratings on any other dependents measures, including factual culpability, may vary from one study to another and they are not intended as absolute predictors of the outcome in any given trial with similar features. The outcomes of comparative values or measures between experimental groups are of greater importance in interpreting the research findings than are the absolute numbers within any groups. In this respect we agree that the outcomes of the JRR show trends. These trends would be no different if the simulated trials were based exclusively on a script of a real case. Similarly, archival analyses of real cases can only show trends, and because those real cases all vary in multiple ways that are beyond the control of the researchers, no clear causal relationships can be determined. By contrast, in a controlled experiment such as the JRR, the causal relationship between the factors varied in the scripts can be empirically assessed, and the robustness of the observed trends can be discerned statistically. As noted in the comments on the experimental methods applied in the JRR, we included numerous indicators of the robustness of the observed trends in reporting the outcomes of the analyses that we conducted. Those standard quantitative indicators show the extent to which support or otherwise for a trend has integrity.

61. Research integrity, including that of trial simulations, is also formally assessed by means of the internal validity, external validity and ecological validity of a particular study (paras. 26-32).
62. The convergent results achieved using multiple methods and diverse measures substantiated the integrity of the research outcomes. In other words, the qualitative and quantitative methods applied in the JRR yielded the same trends and research outcomes. None of the variables pointed to a disparate trend in the research outcomes. The finding that the results from the multiple different measures employed in the JRR were consistent strengthened the force of the conclusions.

No doubt a significant matter is the finding in the JRS "that verdicts were not based on impermissible reasoning or unfair prejudice to the defendant. These outcomes suggest that any fears or perceptions that tendency evidence - whether presented in a separate trial or a joint trial - is unfairly prejudicial to the defendant are unfounded". However, as factual culpability is a measure considered in relation to the dangers of impermissible reasoning, the [submitter] raises a concern that this measure may need further interrogation before reliance is placed on it.

63. A cursory examination of jury simulation research will reveal that factual culpability is not a dependent measure unique to this study, but is the second most common measure used in research on jury decision making after verdict.²⁰ Factual culpability is used as a proxy for verdict. Analogous to a binary verdict, it allows a greater range of responses to be recorded on a scale than the less sensitive categories "guilty," "not guilty," and "hung". This measure was used to gain insight into whether the juries perceived that the accused committed the conduct in issue, separate and apart from the ultimate verdict reached. By gathering measures of both factual culpability and verdict we were able to assess the consistency of the two decisions as a method to cross-check jury reasoning. Critically, to compare ratings by jurors of the likelihood that the defendant committed the conducted described in uncharged acts versus charged acts when the same evidence was presented in a separate versus a joint trial, we used the factual culpability measures, as verdicts were gathered only for charged acts.
64. As is noted above, the JRR did not rely on a single measure to assess impermissible reasoning, but applied multiple measures and multiple methods with independent types of assessments to develop a coherent and triangulated picture of the juries' reasoning. The research conclusions are based on the degree of convergence that emerged from these diverse approaches.

²⁰ Devine and Caughlin, above n 10; Penrod, Kovera and Groscup, above n 8.

Furthermore, the arguably narrow definitions of "impermissible reasoning" and "unfairness to the defendant" which the study adopts, create further concern in relation to the validity of the strong conclusions made in relation to the implications for the criminal justice system. Impermissible reasoning and unfair prejudice are much wider concepts.

65. The JRR conclusions regarding impermissible reasoning and unfair prejudice are limited to the definitions applied in the study, and are not informative about other forms of unfair prejudice and other forms of impermissible reasoning.
66. The definitions of impermissible reasoning and unfair prejudice selected for the JRR were derived from what was reflected in judicial opinions in cases on joinder. We agree that impermissible reasoning and unfair prejudice have broader applications beyond those in joint versus separate trials.
67. All studies have limitations in terms of their scope. As we noted in our prefatory comments (para. 9), this study was not devised to address all conceivable forms of unfair prejudice that may arise through impermissible forms of jury reasoning.

Simplicity of case study trials

The Law Society notes that in each of the trials used in the JRS, the transcripts are very short, and do not reflect the actual length (and complexity) of many jury trials.

68. We note our earlier comments about ecological validity being only one type of validity to assess a study, and that it is improper to assume that the findings of trial simulation studies that do not precisely replicate the experiences of real-life jurors are invalid (paras. 26-32).
69. We agree that the overall length of the simulated trials was shorter, and that the evidence presented was less complex, than is the case in many actual trials, and we acknowledged this feature as a limitation of the study (Report, p. 268). We further note that in real life trials, jurors are not always in the courtroom. Jurors are given regular breaks throughout the day; they are also taken out of the room during a *voir dire*. By comparison, our trial simulations did not include such breaks.
70. Although abbreviated and presented in a condensed period, our simulated trials faithfully included all essential features of real trials (i.e., opening and closing statements, examination in chief and cross examination of witnesses, and judicial directions on the law), to reproduce the tasks and experiences of the mock jurors as faithfully as possible.
71. Few trial simulation studies achieve this degree of verisimilitude. For example, many simulation studies incorporate cognitively different jury tasks, by providing

a summary of evidence in place of direct and cross-examination of witnesses, or by omitting the judicial directions. Often, trial simulation materials are accessed online by individual mock jurors, thus the task of the mock jurors lacks all of the social aspects of jury decision making. Our trial simulations are at the upper end of the spectrum of experimental realism as they incorporated the key cognitive tasks of a jury as well as the social experiences of a real jury. These cognitive and social features of the jury experience are generally acknowledged by jury experts as more critical to replicate in a trial simulation experiment about jury reasoning than is the precise length of a particular trial.²¹

72. To place the realistic parameters of our trial simulation in perspective, we compared the trial length with features reported in a survey of trial characteristics of 62 jury trials conducted in the state of Queensland in a 12-month period.²² The duration of the shortest real jury trial in that sample of actual trials was a total of 2 hours 50 minutes, and the duration of the shortest real jury deliberation in the Queensland trial sample was a total of 24 minutes (following a trial that lasted one day). In general, the length of trials with expert witnesses exceeded the length of trials without expert witnesses. Since ours was a trial without expert witnesses, a shorter duration was realistic for the type of evidence tested.
73. Importantly, the results of the Queensland study of real trials showed that trial length was not predictive of jury verdicts to acquit or convict. This finding undercuts criticisms that the observed conviction rates or factual culpability rates in our JRR were an artefact attributable to the shorter duration of the simulated trials.
74. The duration of our research simulation matched the full experience of at least some jurors in real Australian criminal trials. As a consequence, we respectfully disagree that the overall experience of the mock jurors was so unrepresentative of that of real jurors due to the complexity and duration of the simulated trials that the research results were compromised.

The Law Society notes that in the "limitations of the study" at p268 the authors state:

"Although the experiment was designed to replicate as closely as possible the experience and tasks of actual juries, we cannot exclude the possibility that the results obtained from this abbreviated experience may differ from those obtained in a real trial. For example, compared to the time available for juries to absorb, consider and discuss the evidence in a real trial, the mock jurors in this study performed under conditions that may have increased their cognitive load and made them more vulnerable to heuristic reasoning, confusion and errors than would

²¹Dennis J. Devine, Laura D. Clayton, Benjamin B. Dunford, Rasmy Seying and Jennifer Price, 'Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups' (2001) 7(3) *Psychology, Public Policy and Law* 622; Penrod, Kovera and Groscup, above n 8.

²² Blake McKimmie, Erin Robson, Regina Schuller and Deborah Terry, 'Trial Characteristics Survey' (University of Queensland, 2008).

be likely in a real trial, where the presentation of the evidence and deliberation typically proceed at a slower pace. Moreover, in an actual trial, a jury would have the opportunity to seek further direction or clarification from the judge, whereas that opportunity was not available in this trial simulation"

The Law Society is concerned that inadequate attention is paid to the impact of the abbreviated nature of the mock trials in the comments relating to the limitations of the study (i.e. significantly shorter, less complex and less emotionally charged than an actual trial).

75. In the above comment, three criticisms are made of the ecological validity of the JRR.
76. First, the simulated trials were criticised as too abbreviated. As was discussed above (paras. 69-74), the duration of the trials used in this study was not impossibly different from the duration of a real life trial, as shown by the reported length of actual criminal jury trials conducted in Queensland, and the amount of time that juries spend outside of the courtroom on breaks or waiting in the courtroom.
77. Notably, the length of the simulated trials was comparable to, or exceeded, the length of all prior simulation studies examining the issue of joinder. Further, the experimental simulation used in this study provided the mock jurors with a jury experience that more closely approximates what occurs in real life than all other prior studies on joinder, since we used a videotaped professionally enacted trial (rather than written transcripts of evidence) and included all the critical features of a trial (opening and closing statements, witness examination in chief, cross examination, and judicial directions). Many prior studies lacked one or more of these features. As we mentioned in our review of the literature on joinder research, only one simulation study included deliberating groups, and the content of those deliberations was never analysed (Report, p. 67).
78. Secondly, the simulated trials were criticised as less complex than real trials. We relied on meta-analysis to devise the level of complexity required to produce and test a joinder effect. The meta-analysis indicated that a minimum of three similar counts was needed to be joined in a single trial to produce an increase in conviction rates.²³ To ensure that we created a strong likelihood of producing unfair prejudice attributable to joinder, we doubled the minimum number of similar counts in our trial simulations to six, i.e., our separate trial with tendency evidence and joint trial conditions involved six similar events, four alleging non-penetrative counts and two alleging counts of penetration or sexual intercourse.
79. Next, to place criticism about the parameters of the complexity of the simulated trials in the JRR into a more realistic context, we note that the analyses conducted

²³Sarah Tanford, Steven Penrod and Rebecca Collins, 'Decision Making in Joined Criminal Trials: The Influence of Charge Similarity, Evidence Similarity, and Limiting Instructions' (1985) 9(4) *Law and Human Behaviour* 319.

in the study of a sample of 62 jury trials in Queensland demonstrated that trial complexity, as rated by the presiding judges, was significantly related to the overall number of witnesses.²⁴ Further, this study reported that the modal (most frequently occurring) number of adult witnesses per trial was three, and that the modal number of police witnesses per trial was one (the modal number of child witnesses and expert witnesses per trial was zero). Thus, our simulated trials, which included between two to six prosecution witnesses and one defence witness, were comparable and sufficient in terms of the overall number of witnesses to generate complexity as assessed by Queensland judicial officers in real life jury trials. In other words, the total number of witnesses in our shortest baseline simulated trials, the separate trial and the relationship evidence trial (Trials 1 to 4) was three, which was realistically equivalent to the most commonly occurring number of witnesses in all 62 jury trials included in the Queensland study. Our simulated tendency evidence and joint trials (Trials 5 to 10) which included evidence from a total of seven witnesses exceeded the Queensland trial average by a factor of more than two.

80. To further assess the issue of the complexity of our simulated trials, all mock jurors who participated in our research study were asked a series of post-trial questions to measure their perceptions and experiences of the level of complexity of the trials (Report, Appendix L, p. 335). Their answers reflected the fact that their perceptions matched those of the Queensland judiciary. A comparison of their ratings across the different types of trials demonstrated that greater complexity was perceived as more witnesses were added, which increased both the trial length and the complexity of the evidence (see results on Report, p. 99).
81. Thirdly, the JRR was criticised based on the supposition that the mock jurors were not as emotionally engaged in the simulated trials as real jurors would be in an actual trial.
82. Our observations of the mock jurors disclosed that they were very emotionally charged by the trial simulation. Because we had an ethical obligation to protect the mock jurors from being harmed by their participation in the simulation, throughout the jury deliberations, research assistants were on duty to monitor the research procedures within each jury room. Two individual mock jurors, one male and one female, were released from the study partway through their respective jury deliberations because they became distressed about the case and/or the deliberations. Both received counselling from a psychologist on the research team before leaving the building, and were also referred to an independent counselling service. On at least two other occasions, juries had their deliberations interrupted by a research assistant with the suggestion that they take a short break because the level of emotional intensity among the jury members was highly charged. Other juries themselves initiated a request for a break from deliberation when the level

²⁴ McKimmie et al, above n 22, 7.

of emotional intensity became uncomfortable. These examples underscore the fact that while the trial itself was not one with real consequences for the defendant, the personal experiences and reactions of the juries who attended it were real and their engagement in the simulation had real, emotional consequences for them as individuals and as a group.

83. While the mock jurors knew that their deliberations would not send someone to gaol, a number of jurors believed that they were viewing a videotape of a real trial, or a recreation of a real trial. Several jurors asked the researchers at the conclusion of the simulated trial what verdict had been returned by the jury in the original case, showing that they did not regard the trial as fictional.

84. The following excerpt from Jury 44, a jury which deliberated after viewing a separate trial, indicated how seriously the mock jurors took their task, despite knowing that they were not making the decision to send a real person to prison:

JUROR 4: To me, it sounds like a paedophile who goes free and I can't accept that.

JUROR 8: I know that you have a very strong opinion about this -----

JUROR 4: I have a very strong opinion.

JUROR 1: You know it is not a real case.

JUROR 4: Hey?

JUROR 1: We are not actually sending a person to gaol.

JUROR 4: I know, but child offence, child sexual abuse -----

85. Similarly, the first thing said in Jury 7 after the videotrial ended was "That was pretty full on." After the conclusion of the vote in Jury 10, one juror admitted that they "actually felt a bit stressed" by the deliberations. During the deliberations in Jury 90, one juror stated: "I am getting anxiety over - and this is a mock trial, but it's almost giving me anxiety over-----".

86. Further, despite the fact that jurors were not deliberating on the fate of an actual defendant, a number of juries referred to the consequences of their decision for the accused should they get it wrong. This excerpt from the deliberation of Jury 69 reflected this concern:

JUROR 10: So if you have a doubt, don't put a man basically in gaol-----

JUROR 6: Because last thing you want is someone being - some innocent person-----

87. Our observations of the jury deliberations in process confirmed that the juries were very engaged with the case and the group discussions, both intellectually, and emotionally. The picture below of one deliberating jury in action shows how they used the whiteboard, their own notes and copies of the judicial directions to work

systematically as a group to review the evidence pertaining to each count against the defendant.



88. In addition, the last question on the post-deliberation questionnaire asked mock jurors to report their feelings on a 5-point scale, using 20 adjectives, half of which were positive and half of which were negative (Report, p. 342), drawn from the *Positive Affect and Negative Affect Scale (PANAS)*. Results of analyses of these responses revealed that individual jurors reported a wide range of emotions across all types of trials. Overall, the mock jurors reported equivalent moderate levels of *Positive Affect*, and low to moderate levels of *Negative Affect* across all experimental conditions (Appendix N, pp. 348-350). These measures of their post-deliberation emotions were unrelated to credibility ratings of the focal complainant or the perceived factual culpability of the defendant (pp. 91-92). Given the lack of significant differences, in the time available before publication we did not undertake additional analyses.

89. We acknowledged the three above-mentioned limitations of the JRR (Report, pp. 268-270). Our mock juries did not have the precisely the same experience as they would have if they were seated in a courtroom, deliberating about a real trial. However, their experiences were not so dissimilar to those of real life juries that the study outcomes were compromised by a lack of ecological validity.

Given the relative brevity and mock conditions of the trials, the statements that the mock trials may have led to increased cognitive load and vulnerability to errors of confusion require peer review.

90. Firstly, we refer to our comments above regarding the peer review process applied to the JRR (paras. 2-3).
91. Our comment that the experimental procedures “may have led to increased cognitive load and vulnerability to errors of confusion” was an acknowledgement of the potential impact of the condensed pace of the trial presentation with a substantial amount of unfamiliar and technical information to the mock jurors, and the differences between this pace and the typical pace in a real life trial.
92. In the trial simulations, due to the logistical necessity to accomplish all experimental tasks requiring the participation of the mock-jurors within approximately four hours, the trial pace was condensed and more intense than is typical in a real trial. In other words, the pace of the sequence of discrete key stages in a trial, such as the opening statements, witness examinations, jury directions and deliberation, proceeded more rapidly than is typical in a real court trial. The mock jurors did not have the usual orientation and settling-in period that a real jury experiences, with time between empanelment and the start of the trial. Due to time constraints, the mock juries had only one brief bathroom break after the videotrial ended before group deliberation commenced.
93. The issue of the cognitive load on the mock jurors is related to the complexity of the trial materials and number of witnesses who testify, as was discussed above, and that point is incorporated in our response to this comment, by reference. We assessed the factual errors of the juries exposed to the simpler and shorter versus the more complex and longer simulated trials (Report, Appendix L, page 338). The results of these comparisons reflected that jurors who attended the longer trials made more errors than did their counterparts who attended the shorter trials, and also reported significantly more difficulty in performing various aspects of their duties as jurors. Separating witnesses and counts became more challenging for jurors in the longer experimental trials with more witnesses. Our analyses of errors of fact showed this number increased in relation to the overall number of witnesses whose testimony was presented at trial (Report, pp. 101, 191). In other words, the inference that flows from the above comment, that the trial simulations were so abbreviated and simplistic that no appreciable cognitive load or error rate could arise, is not supported by these research results.
94. The delay in a real trial between the point at which a jury experiences an emotion in response to evocative and potentially prejudicial evidence and the time when a verdict is rendered is often more protracted in a real trial than it was in our trial simulation. Because that emotional point in time is usually separated from the timing of the judgement about a verdict by “more testimony, breaks for rest or meals, closing arguments, jury instructions and sometimes nights of sleep or out-of-court experiences,” in a real trial, this has implications for the nature of jury

reasoning and jury decision making.²⁵ The view of leading empirical jury researchers is that the pace of real trials is likely to diminish the likelihood that juries will engage in improper reasoning motivated by inflamed emotions as the human “emotions tend to lose strength over time, as does their effect on decision-making.” As a consequence, those researchers suggested that in real trials, a prejudicial emotional state evoked by the evidence will not persist, unless re-invoked during deliberations.²⁶

95. Given that our trial simulation proceeded at a more intense pace than a real trial, and did not include extended break between the presentation of the evidence and the deliberation, we included the comment that the pace and procedures of the trial simulation experiment were more likely to capture and reflect any influence of jury prejudice evoked by the trial evidence than might be the case in a real trial. In sum, we respectfully disagree that the trial materials and procedures were inadequate to assess the impact of joinder on the reasoning process of the juries and individual jurors.

We also note that the JRS's literature review omits reference to significant research by Louise Ellison and Vanessa Munro, which focuses on jury reasoning issues in mock trial scenarios.

96. The number of jury studies listed in the Web of Science Social Science Citation Index in 2014 exceeded 3,700, and in Psycinfo, over 3,900.²⁷ In short, there are literally thousands of experimental and other jury simulation studies to which we omitted reference in our review of literature because our review was confined to research on the topic of jury reasoning in joint trials.
97. We are unaware of any published research by Ellison and Munro on jury reasoning in joint versus separate trials. They have conducted non-experimental descriptive studies of mock jury decisions in adult sexual assault cases, but their studies did not explore reasoning about unfair prejudice or joinder of multiple complainants. Nor does their research experimentally manipulate the number of complainants to test the causal relationship between these modifications and the nature and type of jury reasoning engaged in. i.e., their approach is not a robust method to assess the effects of unconscious biases and to examine the causal relationship between those biases and trial outcomes. We did not refer to their work because it did not inform the topic addressed in this study, namely the extent to which undue or unfair prejudice arises in response to the admission of otherwise relevant evidence presented in joint trials.

²⁵ Saks and Spellman, above n 1, 67.

²⁶ Ibid.

²⁷ Kovera, above n 9.

In addition, there appears to be an absence of acknowledging the effect of "unconscious bias". Jury reasoning towards "factual culpability" is treated as an indicator of a good strategy and underpins the authors' conclusions on "impermissible reasoning". This seems to differ from "unconscious bias", which arguably is the basis for the laws' resistance to admitting evidence of uncharged criminal allegations or joining charges.

98. Our response regarding the method we used to assess unconscious bias appears at paras. 17-25 above, and is incorporated here by reference.
99. The measure of “factual culpability” used in this research is a neutral dependent measure. As is noted above, measures of the verdict and of the probability that the defendant is guilty are common dependent measures in jury simulation research. Ratings on this scale do not themselves indicate good or bad reasoning strategies.
100. Just as verdicts can vary as a consequence of unconscious bias, so factual culpability ratings, or the assessments of the likelihood that the defendant is guilty, can vary as a consequence of the presence of unconscious bias. In other words, factual culpability is one measure of the presence of unconscious bias. For example, if mock jurors were unconsciously biased by unfair prejudice, one might expect uniformity to emerge in their factual culpability ratings for the allegations by the complainant with the weak evidence and those for the allegations by the complainant with the strong evidence, and for all non-penetrative and penetrative offences. This was not the pattern of findings that emerged. Because we could not ask jurors in the separate trial with tendency evidence to return a verdict for an uncharged act, we needed a proxy variable for verdict to make this assessment. Accordingly, we asked mock jurors what the likelihood was that the defendant committed the alleged acts, charged and uncharged (see Report, Appendix L, p. 338). A comparison of these factual culpability ratings showed that they were not uniform for counts based on weak and strong evidence, nor for all types of offences. What we were referring to as indicative of logical and permissible reasoning is the overall pattern of these results, which did not confirm the presence of impermissible reasoning associated with unfair prejudice.

It is for these reasons that the Law Society encourages caution with respect to changes to the law of evidence that challenge an accused's protections. There is a need to be alert to unintended consequences. Guidance on the effects of impermissible reasoning and unconscious biases within a ground truth environment would be particularly beneficial. Similarly, with respect to the operation of the presumption of innocence, we would encourage a study that could evaluate the impact of a defendant being from a particular category (e.g. priest) prominently featured in the Royal Commission.

101. The JRR did not aim to explore general issues related to the presumption of innocence as that is a separate research question and topic, outside the scope of

the research on joinder (and has been previously addressed in other studies). Nonetheless, some of the measures of individual juror pre-dispositions that we administered to all mock jurors, such as the PJAQ, included questions on this topic. For example, when mock jurors' scores on the PJAQ factor *Conviction-proneness* were analysed, we found that these individual juror attitudes and biases regarding justice did not predict their verdicts in the simulated trial (see, e.g., Report, p. 103).

102. As was noted above (paras. 22-24), researchers and courts alike are advised to exercise caution about allocating research resources to study the effects of estimator variables which are beyond the control of the court and the litigants, such as the characteristics or history of the defendant that cannot be changed. Moreover, in our report we cited a prior meta-analysis showing that these types of variables exert less influence on jury reasoning than is often anticipated. We noted that the socio-economic status of a defendant has been found to exert a small but significant effect on jury decisions. Generally, a defendant's higher socio-economic status would lead juries to be more lenient. However, where the status of the defendant is implicated in his criminal *modus operandi*, this may lead to a harsher evaluation of, for example, a defendant who is not a priest. For example, defendants who rely on their attractiveness to commit fraud are punished more severely than unattractive defendants facing the same charges because the defendant used that personal characteristic to perpetrate the alleged crime.
103. In this particular research program, to isolate factors associated with the joinder of trials of multiple complainants, and to avoid confounding the research questions about the effect of joinder with any separate potential effects that might arise if the defendant was a priest, we deliberately chose a non-religious institutional setting for the alleged abuse. For this reason, the defendant in the JRR was designated as a sports coach.

Question Trails

We note the finding from the JRS that the use of question trails in the relationship evidence trial meant that 'the defendant was rated significantly less factually culpable'. However, the JRS indicates that the counts and the judges' instructions took over 'a significantly greater proportion of deliberation time' and 'the mock jurors perceived that they required less cognitive effort to evaluate the defence case'. These are encouraging signs and we urge the Royal Commission to explore further the benefits available through question trails, appropriately supported by judicial education. We understand that they have been used quite extensively in other jurisdictions, chiefly in New Zealand. If they improve jurors' ability to apply the presumption of innocence and reduce jurors' cognitive effort, they will be an asset.

However, we consider that any peer review should also address the implications of the finding that 'mock juries reported significantly more difficulty in understanding the charges in a joint trial when given a question trail than when they deliberated without one'; and that '[m]ock

jurors who deliberated with the assistance of a question trail reported requiring significantly more cognitive effort to understand the charges than those jurors who deliberated without a question trail'. Question trails direct jurors to apply the prosecution's burden of proof to elements of the offence and focus on appropriate reasoning limitations. Self-reported indicators of increased cognitive effort by jurors given question trails raises a question as to whether there is an unidentified problem in the mock trials, given that normally we would expect that the use of question trails would improve understanding by jurors. We consider that a response to this from a peer review would be valuable.

104. We welcome further peer review on these topics, and agree that further research on this topic is warranted. Fortunately, a study of the use of fact-based question trails has been conducted in Melbourne, Victoria and Wellington, New Zealand, with 90 deliberating mock juries, and may shed some light on these issues.
105. We agree that the outcomes of the self-reported cognitive effort in the joint trial were counter-intuitive. One contributing factor may be that the question trails used in this study were far more conservative in form than many of those used in New Zealand. For example, the question trail itself did not incorporate the judge's directions on the permitted use of tendency evidence (see Report, Appendix K, p. 331), so juries had to refer to the written transcript they were given of the judge's directions before they could respond to the questions about the elements of the offences listed on the question trail. In fact, the tendency evidence direction was long and intellectually challenging. The unexpected elevated scores may have resulted from the greater effort exerted by juries prompted by the question trail to understand and follow the written directions about the permitted use of tendency evidence, whereas juries without the question trail may have devoted less effort to that task. It may be that further research is needed to determine the impact of different types of question trails on juries' cognitive effort in relation to different types of jury directions (simple vs. complex).

Conclusion

...

The Law Society strongly urges a peer reviewed interdisciplinary report from legal and psychology experts who are disassociated from the researchers within or related to the Royal Commission to address:

- The strengths and limitations of the JRS' design, especially in terms of a ground truth basis;
- The strength and limitations of the JRS' assumptions;
- The strength and limitations of the conclusions, but also addressing specific questions, including:

- the relationship of factual culpability to conclusions recommending, or founding recommendations for procedural or evidentiary change;
- the relationship of unconscious bias to impermissible reasoning, and the extent to which the conclusions of the JRS acknowledge and address this;
- the extent to which jurors' expressed assumptions or views may impact on their unconscious thoughts and impact on their deliberations and verdict;
- whether the findings could be applicable to a trial in which the defendant did not give evidence, and
- whether the conclusions would be valid in relation to trials where the plausibility of the defendant varied from that of the mock study.

106. Responses to all of the foregoing matters were addressed above, with the exception of the question as to whether the research findings from the JRR will generalise to a trial in which the defendant does not take the stand and give evidence.

107. While that was not a variable that was manipulated in the JRR (the defendant gave evidence in all versions of the simulated trial), our expectation, based on findings of other jury studies that have examined this question, is that these findings will generalise to cases in which the defendant does not give evidence. This assessment is made because most typically, evidence of prior convictions is admitted in cases in which the defendant takes the witness stand, but in our study, no evidence of prior convictions or of any criminal history on the part of the defendant was entered into evidence when he took the stand.

108. Prior research has shown that when the evidence against the defendant is strong, verdicts against a defendant who takes the stand are not strongly associated with information about prior convictions. However, when the evidence against the defendant is weak, and information about prior convictions is admitted when the defendant takes the witness stand, the prior convictions exerted a greater influence on the verdicts.²⁸

1.3 Bar Association of Queensland

The Bar Association of Queensland's submission is published on the Royal Commission's website. The following text (references omitted) is extracted from pages 6 and 7 of the submission:

The [Royal] Commission argues that the results of the jury reasoning research offer strong support for the view that the long held fears of prejudice to defendants from the admission

²⁸ Theodore Eisenberg and Valerie P. Hans, 'Taking a Stand on Taking the Stand: The Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes' (2008) 94 *Cornell Law Review* 1353.

of tendency evidence, or of allowing joint trials, is unfounded. We respectfully disagree. To the contrary, we contend that the results of the research demonstrate the opposite.

The results shown in the discussion paper at Figure 10.1 record the conviction rates for the four different trial types. The trial types were:

- (a) separate trial (where only a single complainant gave evidence)
- (b) relationship evidence (where a single complainant gave evidence but also gave evidence of uncharged acts)
- (c) tendency evidence (where the charge related only to a single complainant but evidence of other alleged victims was admitted), and
- (d) joint trial (where multiple charges relating to different complainants were tried together).

The results are recorded for the different types of offences, namely, non-penetrative offences and penetrative offences.

The results demonstrate that, where the jury considered only the evidence of the complainant, i.e., in the separate trial and relationship evidence trial, the conviction rates were low, namely: 11% (non-penetrative) and 0% (penetrative) in the separate trial; and 8% (non-penetrative) and 0% (penetrative) in the relationship evidence trial. The inescapable conclusion is that this resulted because of focus upon whether specific acts were proved beyond reasonable doubt.

By contrast, the conviction rates in cases where the jury heard evidence of the allegations of other complainants was high: 63% (non-penetrative) and 63% (penetrative) in the tendency.

In the tendency evidence trial, the charges considered by the jury remained the same as for the separate trial and relationship evidence trial. The tendency evidence (of similar conduct against two other boys) was not capable of proving the specific acts charged. The juries were instructed they could only use that evidence, if accepted, to reason that the defendant had a tendency to have a sexual interest in young boys, had a tendency to engage in sexual activity with young boys, and had a tendency to use his position of authority to access young boys in order to engage in sexual activity with them.

The inescapable conclusion is that the higher conviction rate was influenced by the other similar allegations led in evidence. Not unexpectedly, the conviction rates in the joint trials were similar. It is our view these results vindicate the fears of unfair prejudice to defendants expressed in the examples from the High Court set out above.

The discussion paper records the view of the researchers that the jury verdicts were logically related to the probative value of the evidence, that, as the inculpatory evidence was increased, conviction rates did too, that the credibility of complainants was enhanced by evidence from independent witnesses, and that little evidence was found that verdicts were based on impermissible or prejudicial jury reasoning. The additional evidence referred to was of course the evidence of other similar complaints against other complainants. As explained

above, it is our view that, where none of the additional evidence could logically help prove the specific acts alleged by other complainants, it was simply the tendency reasoning which drove the convictions. That is, because they believed the defendant to be sexually attracted to boys, they were prepared to find specific acts were proved whereas, without knowledge of that attraction, the same allegations were not proved.

109. We accept and expect that different readers and reviewers may interpret the study findings differently. Our views regarding higher conviction rates in trials which included additional prosecution evidence were guided by principles such as those included in paras. 4, 5 and 12 above. The comments in this submission appear to endorse an alternative interpretation of the research findings and reflect a different interpretation of the law on what is permissible and what is impermissible reasoning. As such, thus submission appears to raise legal issues external to the scope and parameters of this research project.
110. We respectfully disagree with the submission's comments. The JRR tested cross-admissible evidence. Juries who were exposed to tendency evidence received an explicit jury direction from the trial judge that if satisfied beyond reasonable doubt that (i) the sexual acts against the one or both of the tendency evidence witnesses had been committed, and (ii) that the accused had a sexual interest in one or both of these two witnesses, they could then use those facts in determining whether the accused committed the offences against the complainant. Juries who used the tendency evidence in this way were engaged in *permissible tendency reasoning*. As instructed by the trial judge, they were not permitted to assume the accused was a person of bad character.
111. Any such findings by juries (that the acts in the separate trial with tendency evidence or the joint trial had been committed, and that the accused had a sexual interest in boys) do not amount to unfair prejudice because the tendency direction did not encourage the jury to engage in any form of prejudicial reasoning.
112. If prejudicial reasoning had been used in the tendency evidence and joint trials, it would have been detected in our deliberation analysis. The deliberations showed that juries did not make the assumption stated in the above comments: they did not assume that the accused had a tendency to be sexually attracted to boys.
113. Juries took pains to acknowledge and avoid bias or prejudicial reasoning, as is shown in the following excerpt from the deliberations of Jury 90:

JUROR 4: You know, I find myself just constantly pulling back and thinking just because somebody's done it, like, make sure that you don't accuse somebody else in something like that. And don't--kind of almost, like--be the torch-wielding villagers coming-----

JUROR 10: You are worried that the media has influenced-----

JUROR 4: Maybe, yeah.

JUROR 10: -----your - how you see things.

JUROR 4: You know, maybe. And it is just so - I can't actually - the fact that I can't give you a definite answer on why/what I am thinking, it's probably the answer, itself, I guess.

1.4 The Law Society of New South Wales Young Lawyers Criminal Law Committee

The Law Society of New South Wales Young Lawyers Criminal Law Committee submission is published on the Royal Commission's website. The following text (references omitted) is extracted from pages 9 and 10 of the submission:

One of the assumptions underpinning our legal system is that jurors obey directions. It is apparent from the research conducted by Goodman, Delahunty, Cossins and Martschuk for the Royal Commission that courts and legislators have consistently underestimated the ability of jurors to separate counts and evaluate evidence. While the Committee recognises the special dangers attaching to tendency and coincidence evidence, this research suggests that these dangers have been perhaps overstated. However, the Committee submits that further research in this area, and in particular a thorough peer review of the study, is appropriate to ensure that any changes to the law in this respect have a sound empirical basis.

114. We refer to our comments above (paras. 2-3) on the peer review process applied.

115. We agree that further research on these issues is desirable.

Of particular interest are the findings by the study on the insignificance of the 'joinder effect'. That the mock jurors' definition of 'beyond reasonable doubt' was a certainty of under 90% in separate trials and over 90% in joint trials indicates that rather than lowering the threshold for conviction, joint trials increase the difficulty for the prosecution of securing a conviction. Of further note was the finding that jurors were more likely to engage in impermissible reasoning in separate trials without tendency evidence, than they were in separate or joint trials with tendency evidence. In light of this, the Committee agrees with Counsel Assisting that there may be opportunities for reform in this area. The Committee notes that this is a complex area of law and recommends that any proposals to amend the Evidence Act be referred to the Australian Law Reform Commission.

1.5 Confidential Submission

A confidential submission, which has not been published on the Royal Commission's website, expressed a number of concerns about the Jury Reasoning Research, which can be summarised as follows.

The accumulation prejudice or effect arises from the effect of the jury being made aware of multiple allegations, not simply how many charges or counts there are or the number of witnesses called. Comparing trials with the same evidence but different numbers of charges (the tendency trials with two counts and the joint trials with six counts) and versions of the joint trial with different numbers of witnesses (four or six) does not test this.

116. In the JRR we defined accumulation prejudice as prejudice arising from multiple charges or counts and multiple witnesses. Although we did not define accumulation prejudice as arising from multiple allegations, this form of prejudice was nonetheless tested when we compared the case of the moderately strong complainant in the separate trial (two allegations) with the tendency evidence (six allegations; four uncharged) and joint trials (six allegations, all charged) (Report, pp. 94-154).

The report states that 'the factual culpability and conviction rates for the count based on the weakest evidence were not significantly elevated in the joint trial compared to the tendency evidence trial', but there were no conviction rates in the tendency trial for the weakest claim as it was not the subject of a charge. Factual culpability ratings were obtained in the post-trial survey of jurors. It is not surprising that there was little difference in the ratings relating to the 'weak' complainant between the two trials because there was no difference between the evidence in the different trials other than the leading of some further witnesses relating to another complainant in one version of the joint trial. There was no separate trial data for the 'weak case' to provide a valid point of comparison.

117. It is correct that there was no verdict measure to compare juror responses to the charges based on weak evidence in the separate trial with tendency evidence versus the joint trial. Instead, we relied on the factual culpability ratings, our proxy measure for verdict, which were not significantly different (mean score of 5.6 in the tendency evidence trial versus 5.4 in the joint trial). In other words, the defendant was not rated significantly more likely to have committed the act of indecency against the complainant with the weak evidence in the joint than in the tendency evidence trial.

118. To examine whether juries were susceptible to the accumulation hypothesis, we also examined whether they differentiated between the claims supported by weak versus strong evidence in the joint trial. The pattern of verdict and factual culpability ratings differed for these complainants, as did other mean scores on a 7-point scale, such as the likelihood that the complainant (a) misinterpreted events (2.64 vs 2.14); (b) was severely harmed (5.34 vs 5.65); should have recovered by now (2.21 vs. 1.86); was convincing (4.52 vs. 5.59). The composite picture from this range of quantitative analyses, supplemented by the deliberation analyses, showed differentiation of the weak versus strong claims, refuting the concern that unconscious prejudice arising from exposure to multiple allegations would result in the equivalent treatment of all allegations.

119. The sole difference between the separate trial with tendency evidence and the joint trial was whether juries were required to render a verdict about all of the abusive allegations in the latter. Both trials were conducted with 4 Crown witnesses. The

number of prosecution witnesses was varied only in two of the joint trials (4 vs 6 witnesses).

120. In addition to the comparisons of factual culpability ratings for the allegations supported by weak evidence across trials, which yielded similar ratings in the tendency evidence and joint trial, a number of other measures yielded similar scores for the claim with weak evidence across trials, tending to show that this case was not treated differently whether a verdict was rendered on this allegation (more counts/charges) or not. Examples of other measures used that yielded mean scores that were statistically equivalent in both types of trial were the likelihood that the complainant (a) misinterpreted events (2.64 vs 2.64); (b) was responsible for what happened (1.64 vs 1.60); was severely harmed (4.83 vs 5.34); should have recovered by now (2.06 vs 2.21); and was convincing (4.43 vs. 5.53).
121. The composite picture which emerged from this range of quantitative analyses, supplemented by the deliberation analyses, showed differentiation of the weak versus strong claims, refuting the concern that unconscious prejudice arising from exposure to multiple allegations would lead juries to treat all allegations in a similar or equivalent manner.
122. The preliminary online study that we conducted did include a separate comparison of the trials for the complainants with weak, moderately strong and strong evidence (Report, Appendix E, pp. 306-314, 320). Those online mock jurors did not deliberate in groups but rendered individual verdicts.
123. In interpreting the results of the online study versus the deliberation study, we took into account that a previous meta-analysis examining the effects of deliberations on verdicts showed that, in joinder studies, when multiple charges against a defendant were joined in a single trial, the group deliberation increased the effect of joinder on verdict ($r = .33$ with deliberation versus $r = .25$ without deliberation).²⁹ For this reason, we expected that the conviction rates and factual culpability rates for the complainants following group deliberation would be higher than those obtained from the individual online mock jurors. We monitored differences in jury verdicts, factual culpability ratings, and conducted multi-level analyses that disclosed the extent of the group contribution to the verdicts.

The findings in relation to character prejudice are based on responses to the question in the post-trial questionnaire about how convincing the defendant was. The mean response in the basic, tendency and joint trials was the same and the authors conclude that this suggests the jurors were not engaging in impermissible reasoning on the basis of character prejudice, although the results cannot rule out this possibility. The result is surprising given the

²⁹ Michael T Nietzel, Denis M McCarthy, and Monica J Kern, 'Juries: The Current State of the Empirical Literature' in Ronald Roesch, Stephen D Hart and James R P Ogloff (eds), *Psychology and Law: The State of the Discipline* (Kluwer Academic/Plenum, 1999) 999; Penrod, Kovera and Groscup, above n 8, 207.

significant differences in conviction rates between the trials. One possibility is that the question was perceived to be about presentation rather than the weight the juror gave to the defendant's evidence.

124. If the critique of the convincingness ratings of the witnesses held water, i.e., this measure merely assessed the superficial presentational style of an actor and was unrelated to the weight of the evidence, then all convincingness ratings of all witnesses should be invariant in all trials. However, that was not the case. For example, we observed changes in the convincingness ratings of the moderately strong complainant across the four different types of trials. The fact that the ratings of the convincingness of the defendant were not significantly lower in the trial versions where evidence of multiple allegations from multiple witnesses was admitted was interpreted in light of the fact that ratings of convincingness for the complainant varied based on the weight of the relevant admitted prejudicial evidence. Those ratings of the complainant were positively and significantly correlated with other independent measures of the complainant's credibility (Report, p. 106), all of which increased when additional evidence was introduced by the prosecution (Report, p. 107).
125. Our findings on character prejudice did not rely solely on the convincingness ratings of the defendant across trials. In addition, we coded the comments made by jurors in the course of jury deliberations and at the comments they made on the post-trial questionnaire in writing, when explaining why they reached the verdicts they did.
126. Thus, the difference observed in conviction rates appeared to be due to the admission of more relevant evidence from independent sources, not character prejudice.
127. The inference that an early vote precluded disclosure of jury reasoning was not supported by our findings. Our findings match those of prior deliberation research showing that when an early vote is suggested and taken, it is often derailed by discussion.³⁰ In our study, where a vote of guilty followed limited deliberation, this was most often for the claim supported by the strong evidence, not the claims supported by moderately strong or the weak evidence. Juries continued to discuss the way the charges/complainants' evidence interacted after taking a vote. Mock jurors who agreed with each other nonetheless articulated reasons for their views.
128. Preliminary analyses of the time the juries in the JRR took their first vote indicated that the majority took their first vote on the charges after ten minutes of

³⁰ Shari Seidman Diamond, Mary R. Rose and Beth Murphy, 'The Jury: A Look at Deliberations' (Paper presented at the Third International Conference on Empirical Studies of Judicial Systems: Citizen Participation Around the World, Taipei, Taiwan, 5-6 September 2014) 6. An analysis of deliberations showed most calls to vote were unsuccessful: 18.

deliberation. Of those juries who took their first vote in less than ten minutes after starting the deliberation, the deliberation continued on average for 40 minutes, and up to 92 minutes. A similar range was found for juries that took their first vote after ten minutes, with an average deliberation time of 55 minutes, and deliberations lasting up to 96 minutes. Overall, juries had much the same opportunity to discuss the evidence before and after taking their first vote.

129. In addition, in the JRR, the jury reasoning processes for their verdicts on each count were quite clear on the face of the deliberations. As prior deliberation research has shown, while deliberations may not reveal every jurors' private views, they do reveal important features of the decision process, what jurors think is important to mention, how they justify their opinions, how they try to persuade each other, and how they resolve differences.³¹ As Diamond et al noted, these aspects of deliberation confirm to predictions based on the psychological theory of accountability showing that when people have to justify their views, as they do in the course of deliberation, they are motivated to engage in more complex thinking.³²
130. We respectfully disagree with the comment that the research procedure did not generate ample discussion about the tendency evidence and the directions and warnings about it. Nor was the deliberation analysis the problem. The jury discussions were considered as a whole.

In relation to the analysis of juror comments and deliberations, the study found no juries in the tendency or joint trials impermissibly reasoned on the basis of character evidence. There is a severe limitation in the analysis because it considers only statements made during the course of deliberations. It was not uncommon for the jury to take an initial vote on each charge and, where there was unanimity, there was no deliberation and the reasoning process was not exposed. The study also found few instances of explicit permissible tendency reasoning, indicating that the analysis of the deliberations are not revealing the reasoning process.

While the question asked of jurors as to the main reason for their verdict is designed to counter the limitations of the analysis of deliberations, a single quick response to the open ended question is unlikely to reveal a prejudicial reasoning process. The underlying rationale for the limits on the admission to tendency and coincidence evidence are the potential unconscious effects which are difficult to consciously correct for. While the findings clearly show the lack of an overt impermissible reasoning process in most cases, they do not provide evidence about the risk of impermissible reasoning processes at the level they are generally considered to occur.

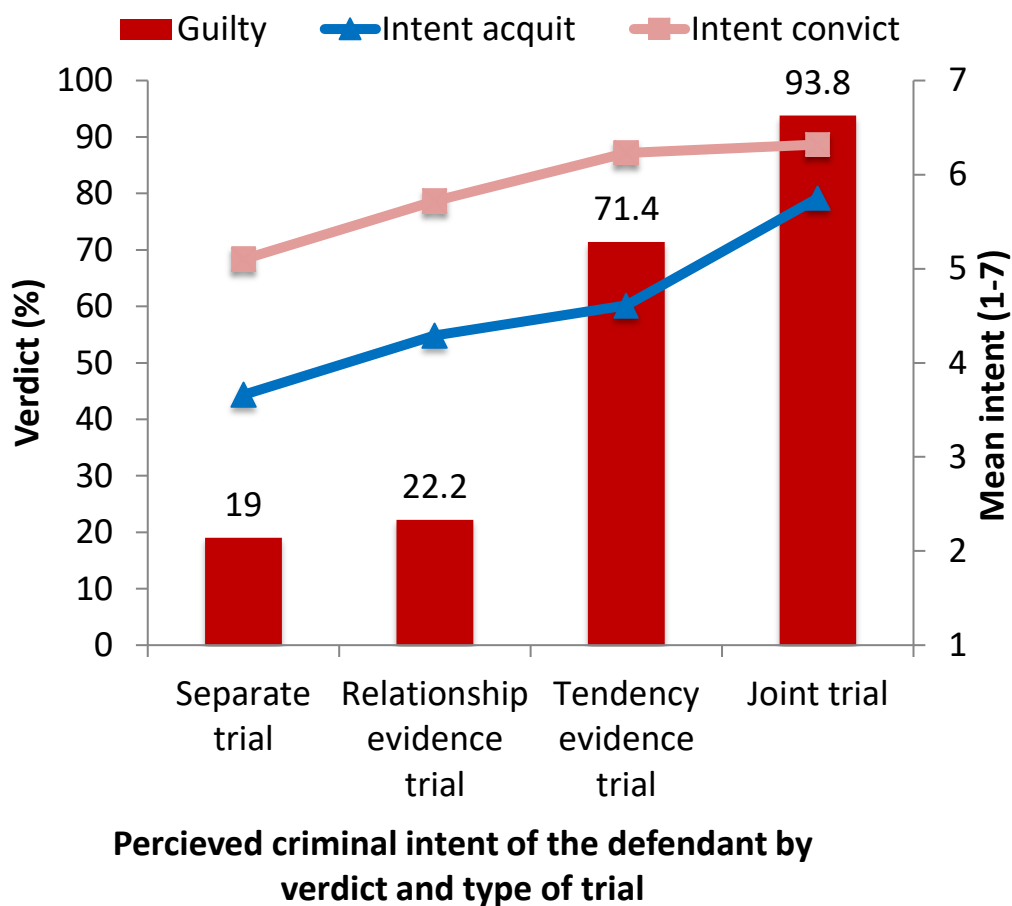
³¹ Ibid 12.

³² Ibid 30.

131. We assessed jury susceptibility to character prejudice using multiple methods. We did not rely exclusively on the deliberation analyses and/or individual juror responses to the post-trial open-ended question about the major reasons for the jurors' verdicts to test the presence of character prejudice, although those measures were helpful in confirming the results of other quantitative analyses of the potential unconscious influence exerted by character prejudice. Additional measures of character prejudice came from psychometrically validated measures of individual juror attitudinal and evidentiary biases, and most importantly, from the experimental design, allowing us to compare responses between trial types to discern the presence of any unconscious biases. Our responses at paras. 17-25 and 40-45 elaborated this point.
132. Secondly, we have more confidence in the jury deliberation analysis and in the open-ended research measure of individual jurors' reasons for their verdicts than is reflected in this submission. Jurors were not rushed into making their responses to the post-trial questionnaire; each juror took as much time as he or she wanted to complete these questions. In responding, the jurors had no insight into our research interest in character prejudice. Nor were they aware of the facts presented in other trials. When we compared response rates across trial types, the findings showed that the stated reasons matched the trial type. In other words, the reasons stated were related directly to the evidence admitted in the different trials. This finding refutes the notion that the content of the jurors' individual written responses were offhand, unconsidered and unreliable.
133. Thirdly, we assessed jurors' individual proclivity to resort to decisions based on character prejudice by calculating their scores on factors within the psychometrically validated Pre-trial Juror Attitude Questionnaire (PJAQ), specifically the factors *Innate Criminality* and *Conviction Proneness*. As we noted in the Report, p. 103, neither of those factors predicted verdicts. Had the verdicts and jury reasoning been motivated by character prejudice, we would expect different outcomes, showing a different relationship between those measures than was obtained.
134. Fourth the experimental research design included parallel measures to gather jurors' perceptions of the defendant in response to a total of six separate questions (Report, p. 340) so that unconscious character bias could be assessed across the four trial types, as well as the extent to which responses to those measures predicted the verdicts in the respective trials. In the experimental conditions in which the mock jurors were exposed to the evidence of multiple acts of abuse against multiple persons, a comparison of mean scores on each question showed no ceiling effects or statistically equivalent values, contrary to what we expected to find if the jury reasoning was motivated by character prejudice. Instead, the mean responses on the 7-point scale varied in accordance with the quantum of

relevant prejudicial incriminating evidence admitted in the respective four types of trials. This pattern of results was not consistent with the presence of unconscious character prejudice.

135. Further, the observed differences in the overall mean responses to the composite measure of perceived *Criminal Intent* of the defendant in the three trial types (Report, p. 103) in which evidence of other “bad acts” by the defendant was admitted (in addition to those charged) did not support an inference of verdicts driven by character prejudice. i.e., if jurors were reasoning on the basis of character prejudice we would expect to see ceiling effects or undifferentiated scores across on these measures across all trials. Significant differences emerged between perceived *Criminal Intent* in the joint trial versus separate and relationship evidence trials, and a marginally significant difference between the joint trial and tendency evidence trial. Moreover, individual jurors’ perceptions of the extent to which a defendant had *Innate Criminality* were unrelated to their ratings of *Criminal Intent* of the defendant in the simulated trial.
136. Finally, our analyses showed that mean scores on inferred *Criminal Intent* in each trial type predicted verdicts, not merely the presence of additional evidence of other bad acts by the defendant. As is shown in the figure below, jurors in the tendency evidence trial who acquitted inferred less *Criminal Intent* (on average a mean of 4.61), compared to their counterparts in the same trial who inferred more *Criminal Intent* (on average a mean of 6.23 out of 7). Higher rates of inferred criminal intent predicted convictions. If character prejudice motivated the reasoning, we would expect to find ceiling effects or statistically equivalent values across all trial types, or no differentiation between or within trial types, including tendency evidence versus joint trials. The observed differentiation substantiated the absence of reasoning based on character prejudice, and these findings converged with other measures of reasoning showing the absence of character prejudice.
137. The main distinguishing factor in the reasoning of jurors who voted to acquit and those who voted to convict was the degree of *Criminal Intent* inferred on the part of the defendant. The pattern of the findings in these two groups of jurors across all types of trials was parallel, showing that the *Criminal Intent* inferred was related to admitted evidence, irrespective of whether the ultimate verdict to by a juror was to acquit or to convict. This pattern of increases in inferred *Criminal Intent* in both groups of jurors demonstrated that more intent was inferred as more incriminating prejudicial evidence was admitted.



2. Extracts from the transcript of Case Study 46

Below are extracts from the transcript of the public hearing in Case Study 46 in which certain expert witnesses gave evidence in relation to concerns or queries about the Jury Reasoning Research.

2.1 Mr Tim Game SC

Mr Tim Game SC gave evidence concurrently with Mr Peter Morrissey SC on 29 November 2016.

The following exchange occurred between Mr Game and the Chair:

MR GAME: Can I ask a question? I'm asking you a question, Commissioner. How does one test, in an exercise like that, the dangers of impermissible reasoning without feeding something in to the study that involves impermissible reasoning?

THE CHAIR: What was done was that their reasoning was analysed, and it was done, as you know, very thoroughly, and it didn't show problems. And I don't know how you would do the other.

MR GAME: But as I understand it, the distinction was between introduction of tendency evidence and introduction of other counts. That can be material of the same kind, but how do you test feeding in material that involves impermissible - that is to say, wrong - reasoning and then test that?

THE CHAIR: You mean irrelevant material?

MR GAME: Material that involves an impermissible train of thought towards reasoning as to guilt, because you have to test the false reasoning.

THE CHAIR: I'm not sure I'm understanding you.

MR GAME: Well, the assumption behind the questioning is that the tendency evidence all involves permissible reasoning. I'm positing the position that one introduces something that involves impermissible reasoning, that is to say, that doesn't pass the test.³³

138. The suggestion in the above comment is to test susceptibility to impermissible reasoning using another research method, for example, by placing a research confederate on the jury who deliberately engages in impermissible reasoning, to test whether other jurors resist or follow suit, and whether this impacts the ultimate jury verdict.

139. We agree that there are many different methodological approaches that can be used to assess the presence of or resistance to impermissible reasoning, each of which has its own advantages and disadvantages. Perhaps future researchers may adopt the type of approach proposed in the above comment. The approach we opted for was to provide juries with information about multiple allegations against the defendant, and observe how they responded, without contriving to mislead them; in other words, to assess jury decision-making in trials with different types of evidence, rather than placing jurors with different social persuasion strategies in the jury room, which is a factor outside the control of the court. In some juries in the JRR, the events suggested by this comment arose spontaneously. In those deliberations, when one juror made a comment that reflected prejudicial reasoning, other jurors corrected this.

2.2 Mr Peter Morrissey SC – 29 November 2016

Mr Peter Morrissey SC gave evidence concurrently with Mr Tim Game SC on 29 November 2016.

³³ Transcript of T Game SC, Case Study 46, 29 November 2016, T24009:16-46.

Mr Morrissey said:

I wanted to add - if I may take 30 seconds on this? The efficiency issues are one thing. The prejudice is another. What you can't test for in a mock trial where they know that they're not actually dealing with a real, damaged individual or a potentially dangerous accused is that you'll never get that emotional hijacking, which is what we're concerned about, because they will just simply know: I know I'm doing an intellectual exercise and that's what I'll do. Their heart will never be troubled by the realities of a courtroom, which can be harrowing, and, if not properly managed, hijacking.³⁴

140. We agree, as noted in our limitations section (Report, p. 268) that the emotional engagement in a simulation may not be equivalent to that in a real trial. However, see above (paras. 81-88) on the emotional investment of the mock jurors in the simulated trials, to the extent that some jurors found it a harrowing experience. We also incorporate by reference our comments above about the fact that intense emotions tend to dissipate over time in real trials. Thus, it cannot be assumed that there is a causal link between a harrowing experience in the courtroom and so-called “hijacking” in the jury room.
141. Thus, while we readily concede that a trial simulation lacks the real-world consequences of a decision to convict or acquit in a real trial, the responses of the mock jurors in this study nonetheless showed that they were very engaged and did not regard this as a mere intellectual exercise. Their interactions with other members of the jury revealed considerable emotional intensity.

2.3 Mr Stephen Odgers SC – 2 December 2016

Mr Stephen Odgers SC gave evidence concurrently with Mr Arthur Moses SC on 2 December 2016.

Mr Odgers said:

I would make the observation that it seems to me that one has to approach that research with some caution. For example, as I understand it, the research that was done with mock juries involved multiple allegations of child sexual abuse. I'm not aware that in any of the scenarios mock juries were actually informed that, for example, the accused had been convicted of an earlier offence or admitted that he had committed or that it was not in dispute.

So that highlights one point, which is that in those research scenarios, one of the greatest concerns about tendency evidence - that is, that a jury will be informed that the accused has, in fact, done that kind of act before - was not present in these scenarios, so that one risk of prejudice was not present. It was much more likely that

³⁴ Transcript of P Morrissey SC, Case Study 46, 29 November 2016, T24010:23-33.

the jury would be engaging in coincidence reasoning rather than in what I will call tendency reasoning. So that is an important qualification to the conclusions that have been drawn from that research report.

142. We agree that the JRR, like any research, should be approached with caution. As was noted above, we distinguished the aims of the JRR from studies examining the impact of prior convictions on jury reasoning. Other researchers have examined the impact of disclosing prior convictions in other types of criminal cases,³⁵ and also the impact of disclosing prior convictions for child sexual abuse.³⁶ See our responses above at paras. 7-8 and 23 about prejudice from prior convictions or admissions posing a different type of prejudice to that investigated in this study of joinder.
143. We respectfully disagree with the characterisation of the JRR as having failed to present the risk of prejudice in the form of evidence that the accused has previously committed similar acts to those alleged in the present trial. See our response at paras. 36-39 above. The facts given to mock jurors in the JRR were those most likely to arise in institutional cases of alleged child sexual abuse where multiple complainants make allegations against the same perpetrator who has not previously been convicted of a child sex offence.

Another point to be made is - and I think Mr Morrissey made this on Tuesday - that the mock juries would have known that these were not real people; that, in fact, when they were being told about allegations it was a situation where it was unlikely that it would generate a kind of emotional response from awareness that a real person in front of you was in fact somebody who had engaged in child sexual abuse undoubtedly in the past.

So the concerns about emotional reactions, about undercutting the standard of proof as a result of awareness of somebody's previous significant misconduct, concerns about tendency to overweigh or give too much weight to such material - I have great concerns that the research would not, in the way it was conducted, have thoroughly elucidated those issues and that great caution should be taken in relying on the conclusions from that.³⁷

144. Please refer to our response above at paras. 17-25 regarding tests of unconscious bias in the JRR, and our explanation of how this was accomplished. As was noted above, the results showed that no support emerged for the hypotheses that the standard or proof was undercut in the presence of tendency evidence. In fact,

³⁵ Devine and Caughlin, above n 10; Eisenberg and Hans, above n 28.

³⁶ Michelle Cowley and Juliette B Colyer, 'Asymmetries in Prior Conviction Reasoning: Truth Suppression Effects in Child Protection Contexts' (2010) 16(3) *Psychology, Crime & Law* 211.

³⁷ Transcript of S Odgers SC, Case Study 46, 2 December 2016, T24353:37-T24354:27.

statistically significant findings were in the opposite direction (Report, pp. 324-342).

145. We agree that determinations about the appropriate weight to accord to relevant prejudicial evidence are difficult to make, as are assessments that “too much weight” is accorded to such evidence, but those are the issues which all judges must consider when admitting or excluding evidence. Our study assessed the presence of three forms of unfairly prejudicial reasoning and whether unwarranted convictions were returned by juries using one or more of those impermissible forms of reasoning.

Mr Odgers also said:

I am sorry, I accept that nothing was disclosed in the way they reasoned to show those kinds of prejudice. But what I'm saying to you is the information they were given was of a certain kind which, necessarily, in my view, meant that you wouldn't expect certain kinds of prejudice, because, for example, they were not told that the accused had, in fact, engaged in child sexual abuse on other occasions, which is one of the greatest concerns in this area; secondly, they weren't confronted by real world, as I've already pointed out and Mr Morrissey pointed out, so, therefore, you wouldn't expect an emotional response generated by such information; thirdly, just because juries don't, when they reason, appear to be engaged in prejudicial thinking or giving too much weight to material - one should be careful about this. One of the concerns is subconscious responses to information and that a juror might, with the best will in the world, be affected in a way which is prejudiced by information but then, in order to justify their conclusion that the person should be convicted, will advance reasons explaining it which seem, on the face of it, entirely appropriate, and may not even be aware of the extent to which they've been prejudiced. I don't think studies of this kind will necessarily reveal those kinds of concerns.³⁸

146. We respectfully disagree. As was noted above at paras. 36-39, the relationship evidence trial and the tendency evidence trials (as well as the joint evidence trial) all included evidence about the defendant engaging in prior acts of child sexual abuse.
147. As noted above, the purpose of the research was not to investigate effects of a prior conviction.
148. See responses above about the degree of emotional engagement observed, and the research design features to assess unconscious prejudice and bias. As we pointed out above, the examination of explicit statements by juries and jurors about their reasons for their verdicts was ancillary to and complementary to the core quantitative experimental outcomes. The experimental procedures included

³⁸ Transcript of S Odgers SC, Case Study 46, 2 December 2016, T24357:13-36.

random assignment of mock jurors to trial conditions, 'blind' observations of jury decision making and interactions, 'blind' coding by research assistants of their deliberations and responses to written questions to thoroughly test the data for the presence of subconscious responses to the evidence presented in the different trials.