

Report for the Royal Commission into
Institutional Responses to Child Sexual Abuse

MAY 2016

JURY REASONING IN JOINT AND SEPARATE TRIALS OF INSTITUTIONAL CHILD SEXUAL ABUSE: AN EMPIRICAL STUDY

Professor Jane Goodman-Delahunty
CHARLES STURT UNIVERSITY

Professor Annie Cossins
UNIVERSITY OF NEW SOUTH WALES

Natalie Martschuk
CHARLES STURT UNIVERSITY

The Royal Commission into Institutional Responses to Child Sexual Abuse commissioned and funded this research project. The authors carried it out with the assistance of the following researchers, people and organisations:

Project Manager: Natalie Martschuk

Research Assistance

Rowena Dench
Robyn Goodwin
Rhiannon Fogliati
Megan Fraser
Karen Gelb
Natalie Hodgson
Peta Kennedy
Alexandra Lonergan
Tanika Low
Melissa Martin
Jake Miyairi
William E Foote
Marco Olea
Rohan Lulham
Ida Nguyen
Stella Palmer

Christian Teixeira
Erin Warrick
Berenike Waubert de Puiseau
Nicholas Watt
Court Services, NSW Department of Justice
Catherine D'Elia
Michael Talbot
Anomaly and Stable Research

Actors

Nicole Abbey
Bernadette Cranage
Stephen Geronimos
Joel Hamilton-Foster
Peter Hayes
Judge Hock SC
Amanda Marsden

Daniel McBurnie
Kara Shead
Dean Michael Smith
John Van Putten
Richard Weinstein SC

Expert Advice

Judge Yehia SC
Nigel Balmer
Eunro Lee

Video Production

Rocket Productions Pty Ltd
Ajit Singh
Joel Hamilton-Foster
Paul Pallese

Disclaimer

The views and findings expressed in this report are those of the authors and do not necessarily reflect those of the Royal Commission.

Published

May 2016

Copyright information

Goodman-Delahunty, J., Cossins, A. & Martschuk, N. (2016). *Jury reasoning in joint and separate trials of institutional child sexual abuse: An empirical study*. Royal Commission into Institutional Responses to Child Sexual Abuse, Sydney.

ISBN 978-1-925289-62-6

© Commonwealth of Australia 2016

All material in this report is provided under a Creative Commons Attribution 4.0 Australia licence. Please see www.creativecommons.org/licenses for conditions and the full legal code relating to this licence.



The details of the relevant licence conditions are available on the Creative Commons website as is the full legal code for the CC BY 3.0 AU licence (<http://www.creativecommons.org/licenses>)

**Royal Commission into Institutional Responses to Child
Sexual Abuse**

**JURY REASONING IN JOINT AND
SEPARATE TRIALS OF INSTITUTIONAL
CHILD SEXUAL ABUSE:
AN EMPIRICAL STUDY**

May 2016



Professor Jane Goodman-Delahunty

Natalie Martschuk, Research Associate



Professor Annie Cossins

Preface

On Friday, 11 January 2013, the Governor-General appointed a six-member Royal Commission to inquire into how institutions with a responsibility for children have managed and responded to allegations and instances of child sexual abuse.

The Royal Commission is tasked with investigating where systems have failed to protect children, and making recommendations on how to improve laws, policies and practices to prevent and better respond to child sexual abuse in institutions.

The Royal Commission has developed a comprehensive research program to support its work and to inform its findings and recommendations. The program focuses on eight themes:

1. Why does child sexual abuse occur in institutions?
2. How can child sexual abuse in institutions be prevented?
3. How can child sexual abuse be better identified?
4. How should institutions respond when child sexual abuse has occurred?
5. How should government and statutory authorities respond?
6. What are the treatment and support needs of victims/survivors and their families?
7. What is the history of particular institutions of interest?
8. How do we ensure the Royal Commission has a positive impact?

This research report falls within theme five.

The research program means the Royal Commission can:

- obtain relevant background information
- fill key evidence gaps
- explore what is known and what works
- develop recommendations that are informed by evidence, can be implemented and respond to contemporary issues.

For more on this program visit www.childabuseroyalcommission.gov.au/research.

Table of contents

List of tables	9
List of figures	11
List of other materials	12
Glossary	13
Executive summary	22
Chapter 1: Introduction	38
1.1 Background	38
1.2 Aim of the research	39
1.3 Scope of the research.....	39
1.4 Key issues.....	39
1.5 Structure of the report.....	40
Chapter 2: Empirical overview	42
2.1 Judicial and law reform discussion of empirical studies.....	43
2.2 Empirical support for judicial assumptions of prejudice	45
2.2.1 Operationalising judicial hypotheses of prejudice.....	45
2.2.2 Dual process theory	48
2.3 Prior research on jury decision making in joint trials.....	50
2.3.1 Archival studies	50
2.3.2 Experimental trial simulations	54
2.3.3 Meta-analyses of trial simulation studies.....	61
2.4 Empirical support for legal safeguards against unfair prejudice	63
2.4.1 Jury instructions.....	63
2.4.2 Fact-based question trails	64

2.4.3	Group deliberation.....	66
2.5	Summary	69
Chapter 3: Research aims, method and design.....		71
3.1	Aims	71
3.2	Method.....	72
3.2.1	Online mock juror pilot study.....	72
3.3	Participants in the jury study	74
3.4	Research design.....	76
3.4.1	Trial simulation materials	77
3.4.2	Dependent measures	79
3.5	Research procedures.....	80
3.6	Data analysis	81
Chapter 4: Results		89
4.1	The influence of mock jurors' pre-trial expectations and attitudes	89
4.1.1	Research aim	89
4.1.2	Mock jurors' individual pre-trial biases and post-deliberation responses	89
4.2	The influence of the trial type on jury reasoning and verdicts	94
4.2.1	Research aim	95
4.2.2	Verdict by trial type.....	95
4.2.3	The influence of trial type on other dependent measures	98
4.3	Jury reasoning by type of trial.....	114
4.3.1	Research aim	115
4.3.2	The prevalence of impermissible reasoning in jury deliberations	115
4.3.3	Qualitative review of mock jurors' main reasons for their verdict	122

4.3.4	Qualitative review of deliberations in separate trials.....	133
4.3.5	Thematic analysis of deliberations about the focal complainant in joint trials	138
4.4	Influence of the number of counts and number of witnesses on jury reasoning and decision making.....	178
4.4.1	Research aim	178
4.4.2	What is the influence of the number of counts?	181
4.4.3	What is the influence of the number of witnesses on jury reasoning?.....	190
4.5	The influence of jury directions on jury reasoning and decision making.....	204
4.5.1	Research aim	205
4.5.2	What is the influence of context evidence directions?	206
4.5.3	What is the influence of tendency evidence directions in a separate trial?.....	212
4.5.4	What is the influence of directions on tendency evidence in a joint trial?	215
4.6	The influence of question trails on jury reasoning and decision making	222
4.6.1	Research aim	222
4.6.2	Overall effects of question trails.....	223
4.6.3	What is the influence of a question trail in a relationship evidence trial?	227
4.6.4	What is the influence of a question trail in a joint trial?.....	229
4.7	Self-reported cognitive effort by type of trial.....	231
4.7.1	Research aim	231
4.7.2	Self-reported cognitive effort	232
4.8	Fairness of the trial.....	238
4.8.1	Research aim	239
4.8.2	Fairness to defendant by trial type.....	239
4.8.3	Self-reported threshold for ‘beyond reasonable doubt’	248

Chapter 5: Discussion.....	249
5.1 Was there a joinder effect?	251
5.2 Did juries convict on the basis of impermissible reasoning?	253
5.3 Legal safeguards against prejudice.....	261
5.4 General conclusions about prejudice in joint trials	265
Chapter 6: Conclusion	266
6.1 Strengths of the study	267
6.2 Major outcomes.....	268
6.3 Limitations of the study.....	268
6.4 Implications for the criminal justice system.....	270
Appendices	272
Appendix A: Comparison of NSW juror and mock juror demographic characteristics (per cent)	273
Appendix B: Ethics approval from Charles Sturt University	275
Appendix C: Pre-trial questionnaire in the online juror study	277
Appendix D: Post-trial questionnaire in the online juror study	283
Appendix E: Online juror study	288
Appendix F: Descriptive statistics for pre-trial and post-trial measures by experimental group in the online juror study	322
Appendix G: Pre-trial juror questionnaire for the jury deliberation and reasoning study ..	324
Appendix H: Verdict form for separate trial with two counts.....	328
Appendix I: Verdict form for joint trial with six counts	329
Appendix J: Question trail for separate trial with relationship evidence.....	330
Appendix K: Question trail for joint trial.....	331
Appendix L: Post-trial juror questionnaire in the jury deliberation and reasoning study ..	335

Appendix M: Quantitative coding scheme for transcribed deliberations.....	343
Appendix N: Descriptive statistics for pre-trial and post-trial measures by experimental group in the jury deliberation and reasoning study	348
Appendix O: Jury reasoning in separate trials with appropriate jury directions	351
Appendix P: Jury directions.....	356

Please note that the online materials referred to in this report, including the trial scripts, can be found on the Royal Commission website www.childabuseroyalcommission.gov.au.

List of tables

<i>Table 1.</i> Participant demographic characteristics	75
<i>Table 2.</i> Trials by counts, Crown witnesses, jury directions, number of juries and mock jurors, and jurors' mean age and gender.....	77
<i>Table 3.</i> Inter-correlations of mock jurors' individual pre-trial biases and post-deliberation responses for the entire study sample.....	91
<i>Table 4.</i> Conviction rates and mean factual culpability of the defendant for counts of the focal complainant with moderately strong evidence, by trial group	96
<i>Table 5.</i> Factual culpability score of the defendant for each count, by complainant and type of trial	110
<i>Table 6.</i> Jury verdicts for counts with moderately strong evidence (Timothy) by type of trial for (a) non-penetrative and (b) penetrative offence (per cent)	111
<i>Table 7.</i> Proportion of factual errors in all jury deliberations, by trial type (per cent)	116
<i>Table 8.</i> Proportion of uncorrected factual errors in all jury deliberations, by trial type (per cent)	116
<i>Table 9.</i> Proportion of all jury deliberations with impermissible reasoning, by trial type (per cent/n).....	121
<i>Table 10.</i> Major categories of reasons for mock jurors' verdicts, by verdict type	123
<i>Table 11.</i> Reasons for mock jurors' decisions to convict by type of trial (per cent)	124
<i>Table 12.</i> Reasons for mock jurors' decisions to acquit by type of trial (per cent)	128
<i>Table 13.</i> Reasons for mock juror decisions on hung juries, by type of trial (per cent)	131
<i>Table 14.</i> Jury and mock juror verdicts for the non-penetrative and penetrative offences of the moderately strong complainant, by type of trial (per cent).....	184
<i>Table 15.</i> Jury and mock juror verdicts for each count, in joint trials with six or four Crown witnesses (per cent)	197
<i>Table 16.</i> Jury and mock juror verdicts for the moderate strength case in the relationship evidence trial, by type of jury directions (per cent).....	207

<i>Table 17.</i> Jury and mock juror verdicts for the moderate strength case with tendency evidence, by type of jury directions (per cent)	213
<i>Table 18.</i> Jury and mock juror verdicts for non-penetrative and penetrative offences in joint trials, with and without tendency evidence directions (per cent).....	218
<i>Table 19.</i> Mean accuracy of remembered case facts, perceived cognitive effort, deliberation time, and perceived factual culpability, by trial type and presence of a question trail	223
<i>Table 20.</i> Jury and mock juror verdicts in the relationship evidence trial, with and without a question trail (per cent).....	229
<i>Table 21.</i> Mock juror expectations of information they would receive at trial (per cent agreeing)	245

List of figures

<i>Figure 1. Mean factual accuracy, by trial type</i>	<i>100</i>
<i>Figure 2. Perceptions of the defendant’s sexual interest in boys, by trial type</i>	<i>102</i>
<i>Figure 3. Perceived criminal intent of the defendant, by trial type</i>	<i>103</i>
<i>Figure 4. Mean perceived credibility of the complainant with moderately strong evidence, by trial type</i>	<i>106</i>
<i>Figure 5. Key reasons for conviction, by trial type (per cent).....</i>	<i>127</i>
<i>Figure 6. Key reasons for acquittal, by trial type (per cent).....</i>	<i>129</i>
<i>Figure 7. Key reasons for verdicts by members of hung juries, by trial type (per cent)</i>	<i>132</i>
<i>Figure 8. Mean perceived factual culpability for each allegation, by case strength and trial type</i>	<i>182</i>
<i>Figure 9. Number of factual recall errors in deliberation, by total number of prosecution witnesses.....</i>	<i>191</i>
<i>Figure 10. Mean perceived factual culpability in the joint trial with four versus six witnesses.</i>	<i>194</i>
<i>Figure 11. Mock jurors’ perceptions of the judicial instructions, with and without context evidence directions</i>	<i>209</i>
<i>Figure 12. Mock jurors’ perceptions of the judge’s instructions with and without tendency evidence directions.....</i>	<i>214</i>
<i>Figure 13a and 13b. Reported cognitive effort in relationship evidence trials and joint trials, with and without a question trail (QT).....</i>	<i>225</i>
<i>Figure 14a and 14b. Self-reported cognitive effort, by type of trial.....</i>	<i>234</i>
<i>Figure 15. Self-reported cognitive effort by jury verdict and type of trial</i>	<i>236</i>
<i>Figure 16. Mock jurors’ perception of the convincingness of the defendant, criminal intent of the defendant, fairness of the judge’s instructions and overall fairness of the trial, by type of trial.....</i>	<i>239</i>
<i>Figure 17. Perceived fairness of the trial to the defendant by type of trial</i>	<i>244</i>

List of other materials

Mock trial case summary.....	84
Case Study 1: Decisions to convict and acquit for the weak claim	155
Case Study 2: Deliberations about tendency evidence	186
Case Study 3: Deliberations about a joint trial	199

Glossary

Accumulation prejudice: A type of impermissible reasoning that accords more weight to evidence than its true value, because multiple charges or multiple witnesses who give evidence against a defendant create the appearance of a stronger case against the defendant than exists in reality.

Basic separate trial: In this study the baseline trial involved one complainant, and a defendant who was charged with two counts of sexual abuse: one count of indecency (masturbation of the complainant), and one count of sexual intercourse (digital-anal penetration of the complainant). Three witnesses gave evidence: the complainant, the defendant and a former neighbour of the defendant (who gave evidence for the prosecution).

Character prejudice: A type of impermissible reasoning based on the unwarranted inference of criminality in a defendant who is thus considered to deserve punishment because he or she is a bad person.

Child sexual abuse: Child sexual abuse includes (i) non-contact sexual conduct by the defendant, such as exposure; (ii) non-penetrative sexual contact such as touching of the breast or masturbation; and (iii) penetration of a person's anus, vagina or mouth by a part of another person's body such as a finger, tongue or penis, or by an object. In this report, child sexual abuse refers to sexual conduct that matches the definitions of sexual assault offences classed as acts of indecency, indecent assault and penetrative sexual assault within the *Crimes Act 1900* (NSW).

Child sexual abuse knowledge: Numerous studies conducted in Australia¹ and internationally² have documented laypersons' common misconceptions regarding child sexual abuse. For example, many laypersons find it counterintuitive that a child who has been abused will continue to have contact with an abuser and show affection to that person. Common

¹ Anne Cossins, Jane Goodman-Delahunty and Kate O'Brien, 'Uncertainty and Misconceptions about Child Sexual Abuse: Implications for the Criminal Justice System' (2009) 16 *Psychiatry, Psychology and Law* 435; Jane Goodman-Delahunty, Natalie Martschuk and Anne Cossins, 'Programmatic Pretest-Posttest Research to Reduce Jury Bias in Child Sexual Abuse Cases' (2016) *Onati Socio-Legal Series*, (forthcoming); Bianca Klettke, David Hallford and David Mellor, 'Perceptions of Credibility of Sexual Abuse Victims Across Generations' (2015) 44 *International Journal of Law and Psychiatry* 91.

² Margaret Bull Kovera and Eugene Borgida, 'Expert Testimony in Child Sexual Abuse Trials: The Admissibility of Psychological Science' (1997) 11(7) *Applied Cognitive Psychology*; Jodi A Quas, William C Thompson and K Alison Clarke-Stewart, 'Do Jurors "Know" What Isn't So About Child Witnesses?' (2005) 29(4) *Law and Human Behavior* 425; Susan Morison and Edith Greene, 'Juror and Expert Knowledge of Child Sexual Abuse' (1992) 16(4) *Child Abuse and Neglect* 595.

misconceptions have been shown to influence mock jurors' views of evidence presented in child sexual abuse trials. People who hold strong misconceptions about child sexual abuse do not rate the evidence by child complainants as plausible or credible.³

Child Sexual Abuse Knowledge Questionnaire (CSA-KQ): A nine-item questionnaire developed from past research findings⁴; psychometrically validated⁵; and used to assess the extent of mock jurors' misconceptions, accurate knowledge and ignorance about child sexual abuse.⁶ A copy of the nine questions is included in Appendices C and G. The questionnaire includes two key factors. The first, 'The Impact of Sexual Abuse on Children', is based on five questions about how children who have been abused react, including the expectation that abused children will react in a similar way⁷, avoid the abuser⁸, try to escape⁹ and display strong emotions following the abuse¹⁰, and the likelihood that a medical examination will confirm the sexual abuse.¹¹ The second factor, 'Contextual Influences on Report', is based on four questions about the reliability of children's reports of experiences of sexual abuse. It measures the extent to which children are regarded as suggestible, specifically whether they make false claims of

³ Jane Goodman-Delahunty, Anne Cossins and Kate O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials: The Effect of Expert Evidence and Judicial Directions On' (2010) 28 *Behavioural Sciences and the Law* 769; Jane Goodman-Delahunty, Anne Cossins and Kate O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials' (2011) 44(2) *Australian and New Zealand Journal of Criminology* 196; Goodman-Delahunty, Martschuk and Cossins, above n 1.

⁴ Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

⁵ Jane Goodman-Delahunty, Anne Cossins and Natalie Martschuk, 'Validation of the Child Sexual Abuse Knowledge Questionnaire' [2016] *Psychology, Crime, & Law* (forthcoming).

⁶ Goodman-Delahunty, Martschuk and Cossins, above n 1.

⁷ Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3; Kovera and Borgida, above n 2; Morison and Greene, above n 2.

⁸ Kovera and Borgida above n 2; Quas, Thompson and Clarke-Stewart, above n 2.

⁹ Morison and Greene, above n 2.

¹⁰ Bradley D McAuliff and Margaret Bull Kovera, 'Do Jurors Get What They Expect? Traditional Versus Alternative Forms of Children's Testimony' (2012) 18(1) *Psychology, Crime & Law* 27; Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3; Klettke, Hallford and Mellor, above n 1; Daniel Bederian-Gardner and Deborah Goldfarb, 'Expectations of Emotions during Testimony: The Role of Communicator and Perceiver Characteristics' 32(6) *Behavioral Sciences & the Law* 829.

¹¹ Quas, Thompson and Clarke-Stewart, above n 2; Kovera and Borgida above n 2; Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

sexual abuse in response to repeated questions about what happened¹², coaching by adults¹³ and manipulation by adults¹⁴, or as a form of revenge against an adult.¹⁵

Cognitive load theory: A psychological theory predicting that the difficulty of learning new concepts or performing a task is associated with the volume and inherent difficulty of new information to be extracted from a source to make a decision, and the way information or tasks are presented. A high intrinsic cognitive load can be predicted when the evidence in a trial is lengthy, when legal directions are complex, or when multiple charges or counts must be decided.¹⁶ In this study, the intrinsic cognitive load of the jury task varied by the type of trial. Extraneous cognitive load is generated by presenting information in a format or manner that includes unnecessary information that unduly burdens the learner.

Coincidence evidence: Coincidence evidence is defined under the *Evidence Act 1995* (NSW) as evidence of the occurrence of two or more events to prove that a person did a particular act or had a particular state of mind on the basis that – having regard to any similarities in the events or the circumstances in which they occurred, or any similarities in both the events and the circumstances in which they occurred – it is improbable that the events occurred coincidentally.

Coincidence reasoning: An assessment of the improbability of two or more events occurring coincidentally, in order to prove that a person did a particular act or had a particular state of mind.

Context evidence: This evidence is adduced to explain the circumstances surrounding the charges; for example, by showing that the charged acts were part of an ongoing history between the defendant and the complainant.

¹² Quas, Thompson and Clarke-Stewart, above n 2; Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

¹³ Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

¹⁴ Kovera and Borgida, above n 2; Quas, Thompson and Clarke-Stewart, above n 2; Morison and Greene, above n 2; Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

¹⁵ Quas, Thompson and Clarke-Stewart, above n 2; Cossins, Goodman-Delahunty and O'Brien, 'Enhancing the Credibility of Complainants in Child Sexual Abuse Trials', above n 3; Cossins, Goodman-Delahunty and O'Brien, 'A Comparison of Expert Evidence and Judicial Directions to Counter Misconceptions in Child Sexual Abuse Trials', above n 3.

¹⁶ Tamsin Ede and Jane Goodman-Delahunty, 'Question trails in trials: Structured versus unstructured juror decision-making' (2013) 37(2) *Criminal Law Journal* 114, 118.

Convincingness: Plausibility, or the quality of causing someone to believe that something is true, certain or worthy of belief. In this study, mock jurors rated the extent to which each of the three complainants and the defendant were convincing on a seven-point scale (1 = strongly disagree; 7 = strongly agree).

Credibility: See *Observed Witness Efficacy Scale*.

Cross-admissible: When the evidence of a complainant is admissible evidence in relation to counts involving *another* complainant.

Dependent measure or variable: A variable measured in a study or experiment that is altered by the presence of one or more independent variables. In this study, the main dependent variables were conviction rates and the credibility of the focal complainant. These dependent variables were altered by the type of trial, which was the main independent variable.

Dual process theory: A psychological theory that refers to two ways in which a decision may be made: by the ‘central route’ or the ‘peripheral route’. The central route involves more effortful thinking, and is used when people have the cognitive capacity and time to think about an issue they regard as important. The peripheral route involves less thinking and is used more frequently by people who are distracted, have limited time to think about the message content or regard a topic as less significant.

Ecological validity: The extent to which the findings of a research study generalise according to real-life settings. In this study, the ecological validity depends on the extent to which the features of the simulated trials match those of real trials.

Effect size: A statistical term that refers to the magnitude or strength of the relationship between an independent variable and a dependent variable.

External validity: The extent to which the findings of a research study generalise according to other people and other situations. In this study, the external validity depends on the extent to which research outcomes will be replicated in other trials with other complainants, with respect to other types of sexual abuse and in other jurisdictions.

Factual culpability: A dependent variable used in mock jury research to measure the likelihood that the defendant committed the acts with which they were charged. In this study, the defendant’s factual culpability was assessed by asking mock jurors to rate on a scale from 1 (very unlikely) to 7 (very likely) how likely it was that they engaged in the behaviours that were the basis for each count or charge against them in the simulated trial; for example, “How likely

is it that Mark Booth masturbated Timothy Lyons's penis between 1 December and 31 December, 1997?"

Focal complainant: Timothy Lyons, whose case was the moderately strong case. The evidence in Timothy Lyons's case was consistent across all trial types.

Heuristic: A mental shortcut derived from experience that eases the cognitive load of making a decision. Heuristics are a normal feature of cognitive adaptive functioning.

Impermissible reasoning: Reasoning that is logically unrelated to the evidence. In this study we assessed three types of impermissible reasoning: (a) inter-case conflation of the evidence; (b) accumulation prejudice; and (c) character prejudice.

Inculpatory: Indicating guilt.

Independent variable: A variable in a study that determines the value of a dependent variable. In this study the independent variables were the type of trial (separate or joint), the type of evidence presented (basic, relationship or tendency evidence), the number of witnesses, the number of counts, the jury directions (standard, context or tendency evidence), and the question trail.

Inter-case conflation of the evidence: A type of impermissible reasoning based on substitution of the facts in evidence about one complainant for facts in evidence about another complainant, in a joint trial involving two or more complainants.

Internal validity: The extent to which the effects detected in a study were caused by an independent variable in the study, rather than by biasing effects of unmeasured variables.

Inter-rater reliability: The extent to which different research coders or raters agree with one another.

Intra-class correlation (ICC): A descriptive statistic used to explain the consistency or conformity of measurements made by multiple observers in the same group on a quantitative measure. In multi-level analyses, intra-class correlation is the ratio of between-group variance to the total variance. In this study it describes the consistency of responses by mock jurors assigned to the same jury in response to post-trial questions about the simulated trial and deliberation they attended. Each jury is a cluster of jurors who watched the same video trial and deliberated as a group. The expectation is that because the mock jurors deliberated together about the evidence before they completed the post-trial questionnaires, their individual responses would *usually be* more similar to each other's responses than to those of mock jurors

who participated in a different jury or did not participate in group deliberations. The intra-class correlation can range between zero and one. The higher the ICC value, the more the differences in the observed values of a particular dependent measure can be attributed to the influence of unique jury groups. Even a small ICC value such as .05 or .10 reflects the fact that the mock jurors' responses were affected by the group process within the jury, and that they differed from responses that would be obtained from individual mock jurors in the absence of those influences. In social science research, the average ICC falls between .05 and .25.

Joinder effect: A statistically significant increase in the conviction rate for an offence when it is tried in a joint trial, compared to the conviction rate for the same offence when it is tried in a separate trial. A large-scale archival study of conviction rates in separate versus joint trials revealed an average increase of 9% in the conviction rate in joint trials.¹⁷

Joint criminal trial: When several counts in relation to two or more complainants or two or more defendants are united in a single criminal trial. A court may order a joint trial of all the counts against a single defendant if the evidence of each complainant is held to be cross-admissible. This decision is also subject to the number of counts that can be heard on the one indictment, which varies from jurisdiction to jurisdiction. The jury is required to return separate verdicts for each particular count. In this study, the joint trial involved three male complainants whose evidence was cross-admissible, with a total of six counts of sexual assault against the defendant.

Manipulation check: A measure to determine whether or not the modification of an independent variable had its intended effect on the study participants. In this study the manipulated variable was the strength of the allegations of the three male complainants, to ensure that mock jurors rated one claim significantly weaker than the others, thus testing the effect of a joint trial on the weak case.

Mock juror: An individual who participates in a trial simulation study in the role of a juror.

Observed Witness Efficacy Scale (OWES): An 18-item psychometrically validated verbal scale¹⁸ that measures the credibility of a witness's testimony in court. A copy of the 18 items, which are rated on a five-point scale (1 = not well; 5 = very well), is included in Appendix L, the post-trial questionnaire completed by mock jurors. The scale includes two factors. The first,

¹⁷ Andrew D Leipold and Hossein A Abbasi, 'The Impact of Joinder and Severance on Federal Criminal Cases: An Empirical Study' (2006) 59(2) *Vanderbilt Law Review* 101.

¹⁸ Robert J Cramer, Jamie DeCoster, Tess M S Neal and Stanley L Brodsky, 'The Observed Witness Efficacy Scale: A Measure of Effective Testimony Skills' (2013) 43(8) *Journal of Applied Social Psychology* 1696.

‘Poise’, predicts credibility and consists of 14 items that assess emotional control, anxiety management and confidence. The second factor, ‘Communication style’, consists of 10 items that measure verbal and nonverbal behaviours associated with witness trustworthiness, likeability and knowledge.

Permissible reasoning: Reasoning that is logically related to the evidence presented in a trial.

Prejudice: See *Unfair prejudice*.

Pre-trial attitudes/biases: Personal views and leanings that may influence a juror’s responses to and decisions about the evidence presented at trial, or in a simulated trial. The main attitudes of interest in this study were those favouring either the prosecution or the defence, as well as attitudes and knowledge about child sexual abuse.

Question trail: Question trails offer a structured approach to jury decision making. They highlight the important legal questions for the jury to consider by providing jurors with a set of logically ordered propositions relating to the issues in the case. Question trails are also referred to as decision trees, step directions and special verdicts.

Relationship evidence: See *Context evidence*.

Relationship evidence trial: In this study, the relationship evidence trial involved one complainant and two counts of sexual assault against the defendant. In addition to the evidence presented in the basic separate trial, the complainant gave additional evidence about grooming behaviours and uncharged acts of sexual abuse by the defendant.

Sensemaking: A psychological theory about group decision making, using a combination of cognitive and social mechanisms to reduce the ambiguity of a problem or task. The main steps in group sensemaking that apply to jury deliberations are (1) recognition of ambiguity and discrepant cues in the evidence; (2) drawing on past experiences to resolve these disparities; (3) generating plausible explanations of the competing claims; (4) evaluating those options through social engagement in deliberation with fellow jurors; and (5) returning a verdict within parameters prescribed by the rules of evidence and jury directions.

Separate trial: See *Basic separate trial*.

Sexual offences: The terminology and definitions of sexual offences vary between Australian states and territories. This report distinguishes non-contact offences such as indecent exposure; non-penetrative offences (indecent assault); and penetrative offences (sexual intercourse). The elements of each sexual offence prosecuted in our simulated trials were taken from the *Crimes*

Act 1900 (NSW). The elements of each sexual offence are set out in the judicial directions in Appendix P and the question trails in Appendices J and K.

Standard jury directions: In this study, the basic separate trial incorporated the following jury directions (a) to (h), referred to in this report as ‘standard jury directions’ and adapted from the New South Wales Judicial Bench Book¹⁹: (a) the presumption of innocence; (b) the onus on the Crown to prove the elements of each charge beyond reasonable doubt; (c) the requirement for a unanimous verdict; (d) elements of an act of indecency; (e) elements of sexual intercourse against a child; (f) separate charges and how to use the evidence; (g) delay in the complaint; and (h) practical problems for the defence caused by a long delay in reporting. A copy of these directions (1,448 words) is included in Appendix P. All experimental trials in this study included these directions.

Statistical power: Power in statistical terms is defined as the probability of detecting an effect given that the effect exists in the target population. The major factors that contribute to the power of an analysis are the sample size (N), the effect size, and the criterion or significance level ($\alpha = .05$ or smaller). To assess the effects of trial type on jury decisions in this study, the sample size was determined by a formal power analysis; the effect size was determined based on the magnitude of effects observed in past studies; and the significance level, as is customary in social scientific practice, was set at 95%.

Statistical significance: Statistical significance describes an outcome that is unlikely to have occurred by chance alone. A significant result does not necessarily imply a large or important practical difference. With a large sample size, even small differences can produce statistically significant results that in practical terms mean little or nothing. Conversely, the smaller the sample, the less likely it is that a test will render statistically significant results when an experimental effect is present. Factors that influence significance are similar to those contributing to power – that is, sample size; significance level applied, such as an alpha level smaller than .05; and effect size. Both statistical and practical significance (the implications of the results apart from statistical values) should be considered.

Source monitoring error: A type of memory error where a specific memory is incorrectly attributed to a specific recollected experience. This can occur for many reasons, when the

¹⁹ Judicial Commission of New South Wales, *Bench Books* (4 December 2009) <www.judcom.nsw.gov.au/bench-books>.

working memory is disrupted by stress or by the cognitive load, and is limited in its ability to properly encode the source of information.

Tendency evidence: Defined under the *Evidence Act 1995* (NSW) as evidence of the character, reputation or conduct of a person – or of a tendency that a person has or had – to prove that that person has or had a tendency to act in a particular way, or to have a particular state of mind.

Tendency evidence reasoning: The use of evidence to show that a person has or had a tendency to act or to think in a particular way, to infer a fact in issue in a criminal trial.

Tendency evidence trial: In this study, the tendency evidence trial involved one complainant and two counts of sexual assault against the defendant. In addition to the evidence presented in the basic separate trial, two additional witnesses gave evidence describing uncharged acts of sexual assault committed by the defendant against them that were similar in nature to the conduct alleged by the complainant.

Trial simulation: An experimental mock trial that employs community members or students as mock jurors. In this study, professional actors played the roles of the witnesses, and real barristers and a real judge played the roles of the legal professionals, while all performances were videotaped. Different versions of the videotaped trial were presented to jury-eligible members of the community.

Unfair prejudice: A real risk that the jury will misuse the evidence in a way that is unfair to the defendant, leading to a miscarriage of justice.

Variance: The distance from the mean or average score of each number in the data set.

Executive summary

Introduction

This report forms part of the Royal Commission into Institutional Responses to Child Sexual Abuse's research program in relation to criminal justice system's response to child sexual abuse in institutional contexts.

Child sex offenders are not a homogenous group and their offending behaviours vary widely. Offenders may offend against one victim or many victims, and they may engage in one incident of sexual abuse or multiple repeated incidents. The diversity and complexity of offending behaviours has a number of implications for the prosecution of child sex offenders.

The scope of this report

This study investigated the extent to which joint trials with cross-admissible tendency evidence infringed defendants' rights, and the extent to which joint trials posed a risk of unfair prejudice to the defendant. In particular, we investigated the reasoning processes of juries in a simulated joint trial of sex offences involving three complainants versus a separate trial involving a single complainant.

Our jury deliberation and reasoning study investigated these issues by presenting 10 different versions of a videotaped trial involving the same core evidence to a total of 1,029 jury-eligible mock jurors. The study tested the impact of evidence strength, the number of charges and the presence of specific judicial directions on jury decision-making in joint versus separate trials.

The five key aims of the project were to:

1. document juries' interpretation of cross-admissible evidence in a joint child sexual abuse trial, to determine the extent to which juries engage in impermissible reasoning regarding such evidence
2. compare the above decision-making processes with those of juries in a separate trial involving the same defendant
3. compare trial outcomes (acquittal, conviction or hung jury) in a joint versus separate trial involving the same defendant
4. examine the relationship between jurors' misconceptions about child sexual abuse, jury deliberations and decisions, and trial outcomes
5. determine the effect of question trail use on juries' reasoning and decisions.

Previous research

The research identified three types of potentially unfairly prejudicial reasoning in cases of joinder: (a) inter-case conflation of the evidence; (b) accumulation prejudice through accumulation of counts or witnesses; and (c) character prejudice. Past studies yielded a ‘joinder effect’ in the form of increased conviction rates when at least three similar crimes were joined in a single trial compared to when separate trials were held for these offences. No prior experimental studies of joinder examined jury decisions in cases of child sexual abuse. Past research focused almost exclusively on conviction rates and failed to distinguish logically related permissible reasoning on the one hand, from logically unrelated and impermissible uses of the evidence in joint trials on the other. Whether observed joinder effects were due to permissible or impermissible and unfairly prejudicial reasoning remains unknown.

The jury deliberation and reasoning study

To address methodological limitations of previous research, and provide empirical evidence on the issues raised by tendency evidence, we used an experimental jury simulation approach to examine the relationship between jury decision making and trial outcomes in a joint trial of alleged child sexual abuse. The trial type was one of four variations:

- (a) Separate trial with an adult male complainant with moderately strong evidence (a basic separate trial)
- (b) Separate trial with an adult male complainant with moderately strong evidence, in which relationship evidence about the defendant’s uncharged sexual acts and grooming behaviours was presented (a relationship evidence trial)
- (c) Separate trial with an adult male complainant with moderately strong evidence, in which tendency evidence from two prosecution witnesses was admitted (a tendency evidence trial)
- (d) A joint trial involving the same defendant and three adult male complainants, who gave weak, strong and moderately strong evidence, respectively (a joint trial).

Judicial directions were defined as one of five variations:

- (a) Standard jury directions
- (b) Standard jury directions plus a context evidence direction
- (c) Standard jury directions plus a context evidence direction with a question trail
- (d) Standard jury directions plus a tendency evidence direction
- (e) Standard jury directions plus a tendency evidence direction with a question trail.

A total of 1,029 mock jurors – 580 women and 449 men aged between 18 and 82 years – were randomly allocated to one of 90 juries to view an experimental trial and deliberate to a verdict with fellow jurors.

Key findings – Chapter 4

Part 4.1: The influence of mock jurors’ pre-trial expectations and attitudes

The first step in the analysis was to assess individual mock jurors’ pre-trial expectations and attitudes. The aim was to examine the contribution of their individual differences to jury reasoning and decision making, and ensure that observed differences in responses to the trials were the result of changes in the trial information and not due to pre-existing differences in the mock jurors assigned to any particular trial group. The results show that the more mock jurors knew about child sexual abuse, the less likely they were to endorse other types of pre-trial bias. Accurate recall of the case facts was higher among mock jurors with more accurate knowledge about child sexual abuse, and lower among those with high expectations of forensic evidence being presented at trial. Mock jurors with higher educational achievement were less likely to expect forensic evidence at trial. Mock jurors who were more knowledgeable about factors that influence a complainant’s reports of child sexual abuse – and who favoured the prosecution – rated the complainant as more credible.

Part 4.2: The influence of the trial type on jury reasoning and verdicts

The study compares jury reasoning and decisions across four different types of trials: a basic separate trial, a relationship evidence trial, a tendency evidence trial and a joint trial. These analyses are based on individual mock juror responses to a written questionnaire completed at the conclusion of their jury deliberations. The purpose of these quantitative analyses is to explore whether there was a ‘joinder’ effect and whether the verdicts were motivated by permissible or impermissible reasoning.

Was there a joinder effect?

As more inculpatory evidence against the defendant was added to the trials, conviction rates for both non-penetrative and penetrative offences against the focal complainant increased. Conviction rates in separate trials with relationship evidence and tendency evidence were significantly higher than those in the basic separate trial. However, there were no significant differences between conviction rates in the tendency evidence trial compared to the joint trial. Thus, we did not identify a joinder effect.

The findings demonstrated that increases in the culpability of the defendant and the credibility of the focal complainant were most prominent in response to sources of evidence that were independent of the focal complainant, and did not increase merely when more evidence was added or the claims were presented in a joint trial. In deliberations about the basic separate and relationship evidence trial, mock jurors were more likely to express the view that the evidence was unpersuasive because it was simply one person's word against that of another.

A major finding was that as more independent sources of evidence were introduced to support the focal complainant's account, the complainant's credibility increased, he was perceived as more convincing and his evidence was accorded more weight. Thus, there were no significant differences in the assessed credibility of the focal complainant in the basic separate trial versus the relationship evidence trial, although the addition of the relationship evidence increased the plausibility of the complainant's account and his evidence was rated as significantly more convincing. In line with this finding, mock jurors were more likely to blame the complainant in the basic separate trial than in any other type of trial. Similarly, ratings of the defendant's sexual interest in boys, inferences about his criminal intent and the factual culpability of the defendant were lowest in the basic separate trial and increased significantly in the tendency evidence and joint trials; that is, as more inculpatory evidence against the defendant was admitted.

Jury deliberations significantly increased ratings of the defendant's criminal intent and factual culpability in trials that involved relationship and tendency evidence. The inferred criminal intent of the defendant predicted the verdict at both juror and jury levels, irrespective of the type of offence. In the absence of tendency evidence, juries were more reluctant to convict for the more serious penetrative offences. Jury distinctions between penetrative and non-penetrative offences confirmed that they reasoned separately about the counts, making distinctions between the counts relating to the same complainant. The presence of tendency evidence increased convictions for both the non-penetrative and penetrative offences, in both separate and joint trials.

Were juries in joint trials more susceptible to inter-case conflation of the evidence?

We tested mock jurors' recall accuracy by asking multiple-choice questions about the case of the focal complainant. Results showed that trial complexity, not trial type, predicted the accuracy of factual recall. Accuracy on these questions was greatest in the less complex trials where only two witnesses appeared for the prosecution (an average of three errors), and decreased as more witnesses appeared for the prosecution, in both the tendency evidence and joint trials (an average of four errors). Mock jurors' formal education had no effect on their

factual recall accuracy. Across all juries, individual mock jurors who made fewer errors on the multiple-choice questions were more prone to acquit, and individual mock jurors who made more errors were more prone to convict, but this was unrelated to the type of trial.

Part 4.3: Jury reasoning by type of trial

To supplement the quantitative analyses reported in Part 2, we conducted a series of additional analyses using other sources of data to gain further insight into jury reasoning and decision making. Results reported in this section are drawn from quantitative and qualitative analyses of the content of the jury deliberations; open-ended responses by individual mock jurors about the main reasons for their verdicts; and a case study of jury reasoning in joint trials. These analyses focused on the prevalence of impermissible reasoning and jury susceptibility to unfair prejudice against the defendant. Contrary to expectations, juries in this study were not prone to impermissible reasoning and made very few factual errors. Most errors were corrected in jury deliberations.

The prevalence of impermissible reasoning in jury deliberations

A quantitative analysis of the content of jury deliberations in which all statements that might indicate unfair prejudice against the defendant were coded revealed that impermissible reasoning was rare, and when it might have occurred, it was more likely in the separate trials without tendency evidence than in the trials *with* tendency evidence.

We found a low rate of factual errors in jury deliberations; only 7.7 per cent of juries made more than two factual errors. Two or more factual errors were more likely to occur in jury deliberations about the joint trial; that is, the trial with the most complex evidence. When errors were made, the vast majority were corrected in the course of deliberations, demonstrating the ability of jury groups to self-correct. None of the 90 jury verdicts were based on inter-case conflation of the evidence.

We found no evidence of emotional or illogical reasoning by juries in any of the trials in which tendency evidence was admitted. We found only two jurors who appeared to use a lower standard of proof than the criminal standard, and only two jurors whose verdicts were driven by emotion. None of the juries featured a juror who reasoned illogically about the evidence.

A qualitative analysis of individual mock jurors' main reasons for their verdict revealed that 90 per cent of the decisions to convict were based on the consistency of evidence from multiple witnesses, the credibility of the witnesses and the pattern of grooming behaviour engaged in by

the defendant. Reasons that might indicate character prejudice as a reason for conviction were less than 3 per cent. These findings were supplemented by a qualitative thematic analysis of jury deliberations about the focal complainant in 33 joint trials, which did not uncover any conviction based on character prejudice.

Overall, our analyses of the reasons for decisions to convict provided negligible support for the notion that joint trials produce verdicts based on inter-case conflation of the evidence, character prejudice or accumulation prejudice. As instructed by the trial judge, mock jurors used their common knowledge and experience of the world in understanding the behaviours of the complainants and the defendant. Together, these findings provided no support for the hypothesis that joint trials lead to impermissible reasoning.

Part 4.4: Were juries in joint trials susceptible to accumulation prejudice?

This section tests the hypothesis that juries in a joint trial use the overall number of charges or witnesses to determine the guilt of the defendant. The results provide no support for the hypothesis that impermissible reasoning was triggered by accumulation of the counts or witnesses against the defendant. This conclusion was based on separate statistical analyses conducted on the accumulation of counts and accumulation of witnesses. Together, convergent results of quantitative and qualitative analyses on each issue confirmed that jurors and juries made logical and appropriate distinctions between the same types of offence alleged by different complainants, based on the strength of the evidence.

The findings demonstrate that the culpability of the defendant was predicted by mock jurors' assessments of the credibility of the complainants, not the overall number of counts and witnesses. We found no reliance on reasoning by accumulation in a joint trial, as there was no significant increase in conviction rates or in the defendant's factual culpability for allegations by the focal complainant in trials with six counts versus those with two counts.

Similarly, the addition of two prosecution witnesses in a joint trial did not increase conviction rates and, most notably, did not elevate the conviction rate for the complainant with the weak claim. In addition, mock juror ratings of victim blame did not vary in response to increases in the number of Crown witnesses, as might be expected if jurors were improperly accumulating the evidence. Rather, victim blame was predicted by individual mock jurors' misconceptions about child sexual abuse. Results of coding the content of the jury deliberations revealed no impermissible reasoning or reduction in the onus of proof in trials with more counts. As juries were exposed to more witnesses and their cognitive load increased, they made more factual

errors, but there were no observed differences in uncorrected or persistent errors across trials; that is, the results in separate and joint trials were similar. A case study of deliberations in a joint trial showed that juries in trials with six counts devoted most available deliberation time to the weak claim where the disparities in evidence were greatest, controverting the view that juries would gloss over these differences in a joint trial. A further case study of deliberations in a joint trial confirmed that no jury decision to convict or acquit was based on impermissible reasoning about the tendency evidence.

Part 4.5: The influence of jury directions on jury reasoning and decision making

In this section we examine whether there was support for judicial assumptions about the effectiveness of jury directions in reducing impermissible reasoning. As was noted above, we did not find that mock jury verdicts were based on impermissible reasoning. Nonetheless, we compared jury reasoning with and without specific jury directions provided to jurors in the relationship evidence trial regarding the uses of context evidence, and with and without specific jury directions on the uses of tendency evidence, provided to juries in the tendency evidence and joint trials. In addition, following the jury deliberations, we asked mock jurors a series of questions about how helpful the directions were.

The findings in this study are in line with a large body of empirical research demonstrating the ineffectiveness of most jury directions. Systematic statistical comparisons of jury reasoning and decisions in the relationship evidence trial, tendency evidence trial and joint trial accompanied by standard directions (on the one hand) and specific directions on the uses of relationship evidence and tendency evidence (on the other) yielded few differences. Overall, the relationship evidence direction was more effective than the tendency evidence direction, which produced no apparent benefits, irrespective of whether it was provided in a separate or a joint trial. Analyses of the content of jury deliberations revealed that error rates in using the context evidence and the tendency evidence were unaffected by the presence of these directions. More deliberation time was devoted to discussing ‘beyond reasonable doubt’ when standard jury directions were given than when juries received tendency directions.

The context evidence direction helped juries overcome their reluctance to convict for the penetrative offence, but the rate of conviction by juries for the non-penetrative offence was unaffected, although factual culpability ratings on both counts increased significantly in the presence of the context evidence direction. Consistent with the findings reported in Part 4.2, conviction rates were predicted by the convincingness of the complainant, irrespective of the

presence of the direction on context evidence. In addition, convictions for penetrative offences in trials with tendency evidence were predicted by higher child sexual abuse knowledge on the part of individual jurors, not the jury directions. In both separate and joint trials with tendency evidence, the judge's tendency evidence direction had no significant influence on the verdict, inferred criminal intent or the factual culpability of the defendant.

Self-report measures provided by mock jurors following their deliberations revealed that mock jurors who received context directions as opposed to the standard directions perceived the judge's instructions as more confusing; found it more difficult to assess witness credibility and apply the law; reported a higher cognitive load; and felt that the judicial instructions made it harder to understand the charges, recall the facts, weigh the evidence and assess the case for the prosecution. Similarly, compared to the standard directions, mock jurors rated tendency evidence directions as more difficult to understand, and perceived that these directions increased their cognitive load. However, mock jurors rated the charges as easier to understand when they were given tendency evidence directions in a joint trial than when they were not given these directions.

Part 4.6: The influence of question trails on jury reasoning and decision making

In this section, we examine the influence of a question trail on jury reasoning and decision making, to discern whether this assisted the juries in their deliberations. Overall, using a question trail appeared to increase the efficiency of jury decision making. The main finding was that using a question trail reduced the overall duration of deliberation in relationship evidence trials, where deliberations persisted far longer in the absence of a question trail. Mock jurors who used a question trail reported that they required significantly less cognitive effort to reach a unanimous verdict than was the case among those who deliberated without this aid.

The question trail had no influence on mock jurors' memory of the case facts. Separate analyses conducted on the relationship evidence trial showed that with the aid of a question trail, the defendant was rated significantly less factually culpable, and accordingly, the conviction rate for both the penetrative and the non-penetrative offences declined. Content analysis of deliberations in those trials revealed that with a question trail, a significantly greater proportion of deliberation time was devoted to discussing the counts and the judge's instructions. When given a question trail, the mock jurors perceived that they required less cognitive effort to

evaluate the defence case. This difference may account for the observed verdict shift from hung juries to acquittals.

Separate analyses conducted on the joint trial revealed that a question trail had no significant influence on the defendant's factual culpability or on conviction rates, regardless of the evidence strength or offence type. However, mock juries reported significantly more difficulty in understanding the charges in a joint trial when given a question trail than when deliberating without one.

Part 4.7: Self-reported cognitive effort by type of trial

In this project, the complexity of the four types of trials varied. Accordingly, the cognitive load imposed on the mock jurors varied by trial type, and was greater in the tendency evidence and joint trials than in the basic separate and relationship evidence trials. Following their deliberations, all mock jurors responded to a series of questions about the extent of effort they had expended in reasoning about the case and coming to a verdict. To gain further insight into mock jury reasoning and decision making, in this section we present the results of mock jurors' self-reports about the difficulty of their tasks, by trial type. As might be expected, mock jurors perceived that recalling the case facts was significantly more demanding in the tendency evidence trial than the basic separate trial, and that understanding jury instructions was more difficult in the joint trial than in the basic separate trial. Mock jurors reported that it required more effort to understand the charges as more inculpatory evidence was admitted.

Unexpectedly, mock jurors perceived that significantly more cognitive effort was required in the separate trial with relationship evidence and the tendency evidence trial than in a basic separate trial, while the joint and basic separate trials were perceived as requiring equivalent effort. The same pattern held for the tasks of assessing witness credibility, weighing the evidence, and evaluating the case for the prosecution and defence. Mock jurors rated reaching a unanimous verdict as significantly more difficult in the relationship evidence trial than in the basic separate trial. They rated deliberation in the relationship evidence and joint trials as more useful in understanding the case than in basic separate trials and in tendency evidence trials. Finally, mock jurors reported that deliberation significantly increased their confidence in the verdict reached.

Part 4.8: Fairness of the trial

A primary concern when considering the use of separate versus joint trials in child sexual abuse cases is the fairness of the trial to the defendant. In this section, we present a series of analyses

that assess the perceptions of juries of the fairness of the trial, by type of trial. These analyses draw on mock jurors' post-trial responses to a range of questions about the fairness of the trial, their expectation that they would be informed of any prior offending by the defendant, and the threshold they applied in interpreting the standard 'beyond reasonable doubt'. In addition, we draw on some coding of the content of jury deliberations.

The main outcome of these analyses is a series of convergent findings showing that mock jurors rated the joint trial as more fair to the defendant than the basic separate trial. As we expected, mock jurors inferred more criminal intent on the part of the defendant as more inculpatory evidence was admitted in the different types of trials, and intent was rated as equivalent in the tendency evidence and joint trials. These ratings show that mock jurors made a logical analysis of the inculpatory evidence presented in the different types of trials.

Other results are unexpected. First among them is the finding that mock jurors viewed the basic separate trial as significantly less fair to the defendant than trials that included more inculpatory evidence. Secondly, the defendant was rated as significantly less convincing in the separate trial with relationship evidence than in the joint trial with tendency evidence. Similarly, the mock jurors perceived the instructions from the judge as significantly less fair to the defendant in the basic separate trial than in the joint trial. A fourth measure that reflected an unexpected difference compared to what might be anticipated was the mock jurors' interpretation of the threshold 'beyond reasonable doubt' applied to convict the defendant. In the basic separate trial, the threshold applied was significantly lower (85.2 per cent) than that in the joint trial (92.1 per cent).

Finally, with respect to information about prior offending by the defendant, a substantial proportion (three-fifths) of the mock jurors expected that they would have been informed at trial of any prior child sexual abuse incidents, charges or convictions involving the defendant. Significantly more mock jurors who attended a separate trial believed that if other charges had been made against the defendant, they would have been informed. In the course of jury deliberation, our content analysis reveals that concern about prior allegations against the defendant were rarely expressed, and no significant relationship existed between trial type and comments made by mock jurors that they would or would not have been informed of prior allegations of sexual misconduct.

Discussion – Chapter 5

Part 5.1: Was there a joinder effect?

As expected, we found that conviction rates varied according to the strength of the inculpatory evidence presented at each type of trial. Conviction rates increased with the admission of more inculpatory tendency evidence. Since these increases in the conviction rate occurred in both the tendency evidence trial and the joint trial, these findings do not support the hypothesised joinder effect.

Although the conviction rates by juries and individual jurors in the joint trial were, on average, higher than those in the tendency evidence trial, these increases were not statistically significant, and were not due to the type of trial; that is, they were not due to the joinder of counts in the joint trial. In other words, we did not find a significant joinder effect.

Importantly, we did not find that the verdicts rendered were based on impermissible or prejudicial jury reasoning. Our analysis of credibility ratings confirmed that juries were sensitive to the source of additional prosecution evidence in assessing witness credibility. We can attribute increases in credibility ratings to systematic and permissible reasoning based on the probative value of the tendency evidence.

Multiple convergent findings showed that jury decision making in the tendency evidence trial was similar to that in a joint trial, indicating that the juries were not reasoning in an illogical and superficial manner in the joint trial when given cross-admissible tendency evidence, compared to the tendency evidence trial which involved one complainant and two witnesses who gave similar accounts of sexual abuse by the defendant. The admission of the tendency evidence, whether in the context of a separate or a joint trial, did not lead to impermissible reasoning.

Part 5.2: Were convictions in joint trials the result of impermissible reasoning?

Inter-case conflation of the evidence

To test the hypothesis that jurors would confuse or conflate the evidence tendered in support of different counts in joint trials, we compared the accuracy of jurors' factual recall and their factual culpability ratings of the defendant.

Accuracy of factual recall

The deliberations revealed that more jurors made factual errors in trials with tendency evidence than they did in trials without tendency evidence, driven in part by the higher number of witnesses in those trials. Jurors' mean recall accuracy scores decreased as trial complexity increased, indicating that the complexity of the trial evidence rather than joinder, *per se*, significantly predicted factual recall accuracy. Because juries promptly corrected their inaccuracies, we found no support for the hypothesis that persistent uncorrected errors were a feature of jury decision making in trials with tendency evidence, so no evidence emerged that errors of this nature had any causal effect on jury verdicts.

Compared to a real trial, where there is considerable repetition and more opportunity for juries to discuss the evidence and deliberate, our experimental simulations may have fostered the potential for more confusion than would arise in a real trial. This element of the trial simulation is likely to have contributed to the higher factual error rate observed in cases with more complex evidence. Nonetheless, in real child sexual abuse trials, whether separate or joint, a similar potential for confusion cannot be discounted as result of, for example, trial length, juror fatigue, juror disinterest, changing levels of concentration and trial complexity.

Factual culpability

The observed pattern of factual culpability ratings showed that juries relied on more systematic reasoning, rather than susceptibility to evidentiary conflation. This evidence of jury reasoning in response to additional evidence of the defendant's other criminal misconduct controverts the hypothesis that juries in joint trials or in trials with complex tendency evidence engaged in impermissible prejudicial reasoning because of inter-case conflation of the evidence.

Accumulation prejudice

To examine the extent to which elevated conviction rates in joint trials with tendency evidence were attributable to impermissible reasoning, we tested whether juries were prone to convict based on the overall number of charges against the defendant or the overall number of witnesses called by the prosecution.

Multiple counts

Courts have hypothesised that a defendant will be unfairly prejudiced in joint trials because juries are prone to reasoning that the defendant is guilty simply because of the number of charges brought by the prosecution.

To test whether juries were affected by the number of counts in the joint trial, we compared two trials in which the same evidence was presented by four witnesses called by the prosecution. The only salient difference between the trials was the number of charges against the defendant.

Factual culpability ratings differed by count and by complainant according to evidence strength, independently of the type of offence, so we did not find any accumulation prejudice as a result of multiple counts. As with verdicts, the factual culpability ratings reflected that juries were able to evaluate the culpability of the defendant for each separate count according to different evidentiary strength. Thus, juries displayed the ability to distinguish between the evidence of different complainants.

Multiple witnesses

Courts have hypothesised that juries are susceptible to the cumulative effects of multiple witnesses, which are expected to increase in joint trials. This version of the accumulation prejudice hypothesis holds that a defendant will be unfairly prejudiced because juries are prone to reasoning that the defendant is guilty simply because of the number of witnesses appearing for the prosecution.

To test jury susceptibility to the accumulative effect of multiple witnesses, we examined conviction rates when either four or six witnesses appeared for the prosecution in a joint trial. The addition of two prosecution witnesses did not significantly increase conviction rates, or ratings of the defendant's factual culpability, providing no support for the accumulation hypotheses. Most importantly, the presence of these witnesses did not increase the conviction rate for the count with the weakest evidence. These findings directly controverted the accumulation prejudice hypotheses in relation to multiple witnesses, by indicating that both jurors and juries evaluated the evidence of multiple witnesses based on its probative value, not simply the number of witnesses.

Character prejudice

Character prejudice arises when a juror uses the severity or number of allegations of criminal misconduct by the defendant to reason that the defendant is a person of bad character, and is therefore probably guilty of the current charges. Encompassed within this concept is the hypothesis that juries will be less concerned about convicting because the defendant deserves punishment for the prior misconduct, charged or uncharged.

The admission of inculpatory evidence about four other acts of sexual abuse from two additional independent witnesses, irrespective of whether they were witnesses or complainants, did not diminish the ratings of how convincing the defendant was, suggesting that jurors were not engaging in impermissible reasoning on the basis of character prejudice. If they had, these ratings would have differed significantly between the separate trial and the trials with tendency evidence, in which juries were exposed to evidence of the defendant's other acts of sexual abuse.

Thematic evaluation of the jury deliberations revealed that no juries in either the tendency evidence or joint trials impermissibly used the tendency evidence to conclude that the defendant was guilty because of the number of allegations of prior misconduct made. Furthermore, there was no evidence of verdicts motivated by emotional reactions to the severity of the allegations, such as a sense of horror regarding the allegations, or a desire to punish the defendant.

Ratings of the credibility and convincingness of the focal complainant showed that the complainant's credibility was enhanced by the evidence from independent witnesses or complainants who reported similar criminal conduct by the defendant, irrespective of whether the defendant was charged with counts pertaining to those individuals. Similarly, jurors' ratings of the convincingness of the focal complainant were significantly higher when tendency evidence was admitted, compared to the separate and relationship evidence trials that had no tendency evidence.

In sum, the low frequency and isolated examples of reasoning in deliberations involving inter-case conflation of the evidence, accumulation prejudice, or character prejudice suggests that the likelihood of impermissible reasoning, whether in joint or separate trials, is exceedingly low. This low probability suggests that there was negligible unfair prejudice to the defendant in joint trials or trials where tendency evidence was admitted.

Part 5.3: Legal safeguards against unfair prejudice

The law attempts to curtail the perceived unfairly prejudicial effect of joint trials and evidence of a defendant's other misconduct via judicial instructions or directions. We examine the extent to which judicial directions and/or fact-based question trails reduce juries' reliance on any impermissible reasoning, thereby mitigating unfair prejudice to the defendant.

Did judicial directions reduce any impermissible reasoning?

We found no differences in the ratings of the perceived criminal intent of the defendant, nor in ratings of his factual culpability for each of the counts when standard directions versus tendency directions were given in either the tendency evidence trials or joint trials. These results mirror findings from our deliberation analysis: that many juries appeared to either ignore or misunderstand the tendency direction and, consequently, failed to apply it or misapplied it.

Nonetheless, juries perceived that more cognitive effort was required when they were given tendency evidence directions versus standard judicial directions in the tendency evidence trial. Overall, it appears that the tendency evidence directions were not only difficult to understand, but were also difficult to apply. The directions, based on accepted legal practice, were not written in plain English; they were comprised of dense, legal language that, anecdotally, appears to pose comprehension problems for lawyers as well. The outcome may be an effect that favours the defence, rather than the prosecution.

Do question trails reduce impermissible reasoning?

The question trail helped juries in the relationship evidence trial reach a verdict more rapidly – on average 25 minutes faster than in the absence of a question trail. From a content analysis of jury deliberations, we found a second feature of the question trails: they increased the proportion of time that juries devoted to discussing the counts of child sexual abuse.

Part 5.4: General conclusions about unfair prejudice in joint trials

Although the expectation was that more complex trials with tendency evidence would result in more unfair prejudice to the defendant, we found more evidence of impermissible reasoning in the basic separate trial and in the relationship evidence trial than in the more complex trials. For

example, in the separate trials, juries were more likely to believe that there was an onus on the defendant to prove his innocence.

This finding is a crucial outcome of this study. Overall, the results show that it is unlikely that a defendant will be unfairly prejudiced in the form of impermissible reasoning as a consequence of joinder of counts or the admission of tendency evidence. Given the low probability, we found there is negligible risk to the defendant of a conviction based on reasoning logically unrelated to the evidence. We recommend further empirical research on jury directions and fact--based question trails.

Conclusion

Major outcomes

This project produced two particularly significant findings:

1. There was little indication that mock juries were susceptible to any joinder effect.
2. Even if there was a joinder effect, there was no evidence that jury conviction rates were the result of impermissible propensity reasoning resulting in unfair prejudice to the defendant.

We specifically looked for instances of verdicts driven by inter-case conflation of the evidence, reasoning by accumulation prejudice and character prejudice. Across four different types of trials, no convictions were made on those bases, and very few mock juror comments reflected emotionally motivated, superficial or impressionistic considerations. Overall, jury reasoning and verdicts were logically related to the probative nature of the admitted evidence.

Implications for the criminal justice system

While some individual mock jurors made errors, and others were susceptible to attitudinal biases and decision-making prejudice, these instances were infrequent. As a group, the juries monitored and corrected individual jurors' errors. The key to the mock juror and jury verdicts was their assessment of the credibility of the complainants, based on the source of the evidence in support of the charges.

In this study, we found 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.

Chapter 1: Introduction

Joint trials and the admission of tendency and coincidence evidence have been controversial issues in several Australian jurisdictions. As part of the Royal Commission into Institutional Responses to Child Sexual Abuse's examination of criminal justice issues, this study investigated the extent to which joint trials with cross-admissible tendency evidence infringed defendants' rights, and whether jury reasoning and decisions in joint trials resulted in unfair prejudice to the defendant.

1.1 Background

The Royal Commission into Institutional Responses to Child Sexual Abuse commissioned this report as part of its examination of the criminal justice system in relation to child sexual abuse in institutional contexts. Child sex offenders are not homogenous group and their offending behaviours vary widely. Offenders may offend against one victim or many victims, and they may engage in one incident of sexual abuse or multiple repeated incidents. The diversity and complexity of offending behaviours has a number of implications for the prosecution of child sex offenders.²⁰

Allegations by two or more complainants may be prosecuted in a joint trial if the complainants' evidence is cross-admissible, in which case, the jury hears the evidence of all complainants. Alternatively, when the complainants' evidence is not cross-admissible, separate trials are usually conducted. Each jury in each separate trial will only hear the evidence of one complainant about the charges involving that particular complainant. Because different Australian jurisdictions have developed different principles for deciding when a joint trial will be held, this study was devised to address the specific judicial concerns about holding joint trials of multiple allegations of child sexual abuse.

The cross-admissibility of complainants' evidence in joint trials raises concerns about the fairness of the trial to the defendant. Of particular concern is the possibility that juries will use the evidence of multiple allegations to engage in impermissible propensity reasoning; that is, to conclude that because there are so many allegations against the defendant, the defendant must be guilty as charged.

²⁰ Annie Cossins, 'The Behaviour of Serial Child Sex Offenders: Implications for the Prosecution of Child Sex Offences in Joint Trials' (2011) 35(3) *Melbourne University Law Review* 821.

1.2 Aim of the research

The Commission's identification of the issue of multiple allegations of child sexual abuse gives rise to the main topic of this jury research project. Because joint trials and the admission of tendency and coincidence evidence have been controversial, and there have been a number of law reforms on the topic, this study investigated the extent to which joint trials with cross-admissible tendency evidence infringed a defendant's rights, and whether jury reasoning and decisions in joint trials resulted in unfair prejudice to the defendant. In particular, we investigated the reasoning processes of juries in a simulated joint trial of child sex offences involving three complainants, versus a separate trial involving a single complainant. The study examined the relationship between jury decision making and trial outcomes. It aimed to gather empirical research to facilitate the development of evidence-based rather than anecdotally driven legal policies about the prosecution of multiple allegations of child sexual abuse.

1.3 Scope of the research

Our jury deliberation and reasoning study investigated these issues by using 10 different video-recorded trials involving the same core evidence. A total of 1,029 jury-eligible mock jurors were randomly assigned to one of the 10 trials, which varied from 45 to 110 minutes in length. Altogether, we conducted 90 simulated jury trials (see Chapter 3 for details). After watching a realistic video simulation of a child sexual assault trial in which the defendant faced either two or six charges, juries deliberated to reach a verdict. Specifically, the study tested the impact on jury decision-making of evidence strength, the number of charges and the presence of specific judicial directions in joint versus separate trials.

1.4 Key issues

We identified a number of key issues in consultation with the Royal Commission, so we could investigate how different trial types impact fairness to the defendant. The five key aims of the project were to:

1. document juries' interpretation of cross-admissible evidence in a joint child sexual abuse trial, to determine the extent to which juries engage in impermissible reasoning in relation to such evidence
2. compare the above decision-making processes with those of juries in a separate trial involving the same defendant

3. compare trial outcomes (acquittal, conviction or hung jury) in a joint versus separate trial involving the same defendant
4. examine the relationship between jurors' misconceptions about child sexual abuse, jury deliberations and decisions, and trial outcomes
5. determine the effect of question trails on juries' reasoning and decisions.

The central research challenge was to distinguish between the legitimate inculpatory effects of additional evidence against the defendant, and impermissible forms of reasoning that undermine the presumption of innocence. To this end, we included some less realistic trials in the simulation experiment so we could isolate potential causes of observed differences in jury responses to variations in the evidence and type of trial. Although evidence and jury directions usually work in tandem, to differentiate between the effects of additional evidence and the effects of relevant jury instructions, in certain experimental trials we varied the evidence only, while in others we only varied the number of witnesses, the number of charges or the type of judicial directions. Use of this programmatic methodology enabled a systematic analysis of the factors that influenced jury reasoning or created a risk of prejudice to the defendant. Because precise measures of reasoning by groups are difficult to ascertain, we assessed jury group deliberations and individual juror decisions, using multiple convergent measures of inferences drawn from the evidence to compile a composite picture of factors that influence jury reasoning and decision making in joint versus separate trials.

1.5 Structure of the report

In Chapter 2, we offer some critique of the psychological research cited by courts and law reform bodies to support policies and practices antithetical to joint trials. We review the three major sources of potential unfair prejudice identified by courts and law reform agencies, and explain how these might manifest in jury decision making. Next, we examine the extent to which broader empirical literature on jury reasoning and decision making support judicial assumptions about the prejudicial effect of joint trials. We distinguish findings derived from archival analyses, field studies and experimental trial simulations, and briefly review the strengths and weaknesses of each approach. We evaluate the empirical support for procedural measures designed to safeguard against unfair prejudice to defendants in joint trials. Despite the absence of any prior experimental research on joinder in child sexual abuse cases, we draw lessons from past findings for the present project, and outline the premises for a jury

deliberation and reasoning study that is high in internal and external validity, and has sound ecological validity.

Chapter 3 describes the research aims and method of the jury deliberation and reasoning study, and includes a summary of the approach taken in analysing the jury deliberations, by combining quantitative and qualitative methods.

Chapter 4 presents the findings of the jury deliberation and reasoning study. First, we report the overall results of mock juror attitudes, expectations and preconceptions that may influence their reasoning and decisions. Following this we provide quantitative and qualitative analyses of juror reasoning and jury deliberations, illustrated by excerpts from deliberations and case studies. After this, we assess the influence of the number of counts of child sexual abuse, the number of witnesses, jury directions and question trails on jury reasoning. Finally, we consider the effect of the type of trial on jurors' self-reported cognitive effort, and the fairness of trials.

Chapter 5 summarises the extent to which mock juries engaged in legitimate or impermissible forms of reasoning, and how jury reasoning in a joint trial with cross-admissible evidence differed from jury reasoning in a separate trial without that evidence. In addition, the strengths and weaknesses of the research design and outcomes are evaluated.

Finally, Chapter 6 summarises the implications of our empirical study for the development of evidence-based legal policies regarding the prosecution of multiple child sexual abuse allegations.

Chapter 2: Empirical overview

A review of past research showed little support for prejudicial effects when similar offences were joined at trial. In addition to other methodological limitations, the prior studies focused primarily on conviction rates and failed to distinguish logically related permissible reasoning from logically unrelated and impermissible uses of the evidence.

Given that the law regulating the ordering of joint trials and admission of evidence of a defendant's other criminal misconduct is premised on assumptions about the reasoning abilities and behaviour of juries, it is imperative that we critically assess these assumptions against empirical research.²¹ Indeed, Eames J has pointed out that “in the absence of such research, it is a field in which anecdote, self-assurance and self-delusion abound within the ranks of the legal profession and the judiciary”.²² Both the Australian Law Reform Commission (ALRC) and the High Court have explicitly cited empirical research to justify the onerous barriers to admitting evidence of a defendant's other criminal misconduct into evidence.²³

The focus of this chapter is threefold. First, we offer some critique of the psychological research courts and law reform bodies cite in support of the prejudicial consequences of joint trials. Second, we examine the extent to which the broader empirical literature on jury reasoning and decision making supports judicial assumptions about the prejudicial effect of joint trials. Third, we evaluate the empirical support for procedural measures designed to safeguard against unfair prejudice to defendants in joint trials.

²¹ Australian Law Reform Commission, New South Wales Law Reform Commission and Victorian Law Reform Commission, *Uniform Evidence Law*, ALRC Report No 102 (2005), 592 [18.8]; see also Dorne Boniface, ‘The Common Sense of Jurors v the Wisdom of the Law: Judicial Directions and Warnings in Sexual Assault Trials’ (2005) 28(1) *University of New South Wales Law Journal* 261; Mark Nolan, ‘More Creativity, Less Criticism: An ‘Evidence Based’ Approach to Jury Reform’ (2003) 15(3) *Legaldate* 5.

²² Geoffrey Eames, ‘Towards a Better Direction—Better Communication with Jurors’ (2003) 24 *Australian Bar Review* 36, 39.

²³ *BRS v The Queen* (1997) 191 CLR 275, 322; *Uniform Evidence Law Report*, above n 21, 592 [18.8]; Australian Law Reform Commission, *Evidence*, Interim Report No 26 (1985), 795–809.

2.1 Judicial and law reform discussion of empirical studies

In *BRS v The Queen*²⁴, Kirby J stated that “the need for care in the admission of evidence about events prejudicial to an accused has been further reinforced by empirical studies”. His Honour cited psychological studies demonstrating that assumptions about the ‘cross-situational consistency’ of human behaviour were frequently unwarranted, and interpreted findings to imply that ‘bad character evidence’ is often unduly prejudicial to the defendant.²⁵ The cited body of psychological research established that human behaviour is frequently dependant on situational factors, more than stable personality traits.²⁶ As Kirby J noted, people readily attribute the behaviour of others to stable personality traits, and underestimate the influence of situational factors, a phenomenon known as the ‘fundamental attribution error’.²⁷ In his Honour’s opinion, these findings indicated that “[o]nce lay decision-makers know facts about the background and character of the defendant, the risk is acute that the focus on the particular offences charged will be lost”.²⁸ Hence, Kirby J suggested that the “defendant may be judged by reasoning that anyone shown to have acted in the criminal or discreditable way proved *must* be guilty of the offences charged, so long as they bear some similarity to the facts established by the evidence”.²⁹ In his Honour’s view, this risk was especially salient in light of research challenging the judicial presumption that jurors are willing and able to use a defendant’s other criminal misconduct in the limited way judges instruct them to do so.³⁰

Previously, the ALRC’s 1984 interim report on the law of evidence, chaired by Kirby J, had cited empirical research to confirm the need for stringent controls on character, tendency and coincidence evidence, and by extension the holding of joint trials. In fact, the ALRC believed that the existing psychological research demonstrated such a ‘real danger’ that jurors would give character and tendency evidence disproportionate weight that it recommended such evidence be admissible only in ‘exceptional circumstances’.³¹ In support of their recommendations, the ALRC referred to ‘the halo effect’, a finding that research participants

²⁴ (1997) 191 CLR 275, 322, citing F M Neasey, ‘Similar Fact Evidence and Propensity Reasoning’ (1985) 9 *Criminal Law Journal* 232, 240; Donald K Piragoff, *Similar Fact Evidence: Probative Value and Prejudice* (Carswell, 1981) 292.

²⁵ *BRS v The Queen* (1997) 191 CLR 275, 322.

²⁶ *Ibid.*

²⁷ *Ibid.*

²⁸ *Ibid.*

²⁹ *Ibid* 322–3.

³⁰ *BRS v The Queen* (1997) 191 CLR 275, 322, citing Evelyn Goldstein Shaefer and Kristie L Hansen, ‘Similar Fact Evidence and Limited Use Instructions: An Empirical Investigation’ (1990) 14 *Criminal Law Journal* 157, 179.

³¹ *Evidence Interim Report*, above n 23, 795–809.

who were unfamiliar with a particular person tended to generalise from one prominent ‘good’ or ‘bad’ characteristic to an overall impression of the person as a whole.³² Moreover, the Commission concluded that evidence of a defendant’s other criminal misconduct would be particularly prejudicial based on research showing that unfavourable information had a greater impact than favourable information in colouring participants’ subsequent judgments of a person.³³ Finally, to further justify its view of the prejudicial nature of evidence of a defendant’s other criminal misconduct, the Commission erroneously relied on purported research on the ‘regret matrix’³⁴, a utility-based theory, ostensibly predicting that juries were less concerned about the possibility of mistakenly convicting the defendant if they were aware of the defendant’s previous misconduct. Examination of the sources cited revealed that they consisted of academic and doctrinal speculations, and that no such jury research existed.

Returning to the topic in 2006, the ALRC reiterated its view that the laws limiting the potential for juries to misuse or overestimate the value of tendency, coincidence, credibility and character evidence were supported “to a considerable extent by a substantial body of psychological research”.³⁵ The Commission noted that the reports cited in its previous reports and by the High Court continued to be supported by empirical studies.³⁶ However, the research findings upon which it relied examined general psychological research on reasoning processes; they were not studies of jury reasoning. Systematic jury research has been conducted since the 1950s.³⁷ Although empirical legal research on jury behaviour has long been established as a field in its own right, judicial assumptions about the prejudicial effects of joint trials were not explicitly underpinned by any supporting empirical evidence, and did not take into account any studies of these issues conducted specifically in the jury context.

The present study is driven by policy; that is, the aim of the study is to examine the extent to which there is empirical support for judicial hypotheses regarding the prejudicial effect of cross-admissible evidence in a joint trial of child sex offences. As was noted above, these hypotheses

³² *Evidence Interim Report*, above n 23, 795–809; Harold H Kelley, ‘The Warm-Cold Variable in the First Impressions of Persons’ (1950) 18 *Journal of Personality* 431; David J Schneider, ‘Implicit Psychological Theory: A Review’ (1973) 73 *Psychological Bulletin* 294, 299.

³³ *Evidence Interim Report*, above n 23, 800, citing Solomon E Asch, ‘Forming Impressions of Personality’ (1946) 41 *Journal of Abnormal and Social Psychology* 258.

³⁴ *Evidence Interim Report*, above n 23, 800–3, citing Richard Eggleston, *Evidence, Proof and Probability* (Weidenfield and Nicolson, 2nd ed, 1983); Richard O Lempert and Stephen A Saltzburg, *Modern Approach to Evidence* (West Publishing, 1977) 151–3.

³⁵ *Uniform Evidence Law Report*, above n 21, 80.

³⁶ *Uniform Evidence Law Report*, above n 21, 80–85.

³⁷ Edith Greene, Sonia R Chopra, Margaret Bull Kovera, Steve D Penrod, V Gordon Rose, Regina Schuller and Christina A Studebaker, ‘Jurors and Juries: A Review of the Field’ in James R P Ogloff (ed), *Taking Psychology and Law into the Twenty-First Century* (Kluwer Academic Press, 2004) 225.

have limited theoretical and empirical foundations. We must delineate and define them with specificity before testing them. Next, we must examine how judicial assumptions of prejudice – those developed by reference to the psychological research as well as those that have evolved independently – have been applied to develop researchable questions that have been empirically tested in studies of jury reasoning and decision making.

2.2 Empirical support for judicial assumptions of prejudice

The distinction drawn in the case law between ‘moral’ and ‘reasoning’ prejudice is a helpful starting point in conceptualising the prejudice that a defendant may face in a joint trial. However, the categories overlap substantially in certain instances. For example, ‘reasoning prejudice’ encompasses “erroneous generalisation and oversimplification” and “the formation of social attitudes before or despite objective evidence”.³⁸ Yet these types of reasoning biases may arise from social perceptions or attitudes that derive from normative judgments – namely ‘moral prejudice’. A further problem with this nomenclature is that decisions influenced by social perceptions or attitudes deriving from normative values are, by definition, classified as ‘morally prejudicial’. One of the fundamental justifications for trial by jury is that it brings the conscience and values of the community to bear on issues in a way that a judge cannot.³⁹ Accordingly, most jury decisions will be underpinned to some extent by social or normative values and cannot be categorised as prejudicial for this reason alone. Judicial concerns over ‘moral’ and ‘reasoning’ prejudice must be clarified with more specificity to be empirically assessed.

2.2.1 Operationalising judicial hypotheses of prejudice

Despite an extensive body of legal doctrinal commentary on jury prejudice against the defendant in joint trials, few academics and judges have summarised the risks as succinctly as the US Fourth Circuit Court of Appeal⁴⁰, which in 1976, in the case of a defendant who was charged with two bank robberies involving similar facts, distinguished the risk of three types of prejudice to a defendant posed by the joinder of multiple offences as:

- (a) inter-case conflation of the evidence – confusion of the evidence across charges
- (b) accumulation prejudice - accumulation of the evidence across multiple charges, giving it undue weight

³⁸ Key Sun, ‘Two Types of Prejudice and their Causes’ (1993) 48(11) *American Psychologist* 1152, 1152–3.

³⁹ Xavier Connor, ‘Trial by Jury—Can it Survive?’ (1987) 61 *Law Institute Journal* 818.

⁴⁰ *U.S. v Foutz*, 540 F.2d 733 (4th Cir. 1976).

- (c) character prejudice – the use of evidence from one crime to infer criminality on the part of the defendant.

Research on joint versus separate trials has typically examined assumptions about prejudice by testing these three central hypotheses to determine the risk of prejudice to the defendant.

Inter-case conflation of the evidence

A common hypothesis is that jurors will confuse or conflate the evidence tendered in support of different charges in joint trials, which has been referred to as ‘inter-case evidentiary conflation’.⁴¹ Because courts will not order joint trials unless the charges are similar in nature, concern arises that juries may be particularly susceptible to inter-case conflation of the evidence within joint trials.⁴² The Victorian Law Reform Commission has argued that:

Where a person is charged with separate sexual offences against several complainants there is a risk that, if the same jury hears all of the counts, it might use evidence relating to an offence charged in one count to decide that the person has also committed a different offence, even though there may be insufficient evidence to support a conviction for the second offence.⁴³

Of course, confusion of the evidence may not necessarily be prejudicial to a defendant. Put simply, the research question is whether juries are capable of separating the counts against the defendant in reaching their verdicts in a joint trial.

Accumulation prejudice

The second concern or hypothesis when trials are joined is that a defendant will be prejudiced by juries ‘accumulating’ evidence across different charges, such that any given item of evidence would be perceived as stronger when presented with the evidence of other complainants (‘accumulation of evidence’).⁴⁴ As Leipold and Abbasi submitted, intuitively at least, the more counts the defendant is charged with, the more likely it is that a jury will conclude that the

⁴¹ Sarah 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, 320; See also Kenneth S Bordens and Irwin A Horowitz, ‘Information Processing in Joined and Severed Trials’ (1983) 13(4) *Journal of Applied Social Psychology* 351; Sarah Tanford and Steven Penrod, ‘Biases in Trials Involving Defendants Charged with Multiple Offenses’ (1982) 12(6) *Journal of Applied Social Psychology* 453; Leipold and Abbasi, ‘The Impact of Joinder and Severance on Federal Criminal Cases: An Empirical Study’ above n 17.

⁴² *Drew v United States*, 118 U.S.App.D.C. 11, 331 F.2d 85, 88 (1964).

⁴³ Victorian Law Reform Commission, *Sexual Offences: Discussion Paper*, (2001) 133 [8.51].

⁴⁴ Dennis J Devine, *Jury Decision Making: The State of the Science* (New York University Press, 2012) 59.

defendant is guilty of *something*.⁴⁵ The hypothesis is that juries are prone to conclude that because there are multiple charges or multiple witnesses supporting the prosecution's case, the defendant is guilty. The impermissible reasoning would be to place greater weight on a particular item of evidence merely because it is presented along with other evidence that is not probative of the defendant's guilt in its own right.⁴⁶

However, in some joint trials, evidence relevant to one offence may appropriately reinforce the evidence relevant to another offence. Accumulation of evidence would not be unfairly prejudicial if, for example, together with probative evidence of the defendant's tendency from other complainants, the evidence of one complainant was strengthened within the overall factual matrix. The critical research question is whether, faced with elevated jury conviction rates in a joint trial compared to the same charges in a separate trial, convictions based on permissible reasoning can be distinguished from those based on impermissible reasoning.

Character prejudice

The third common formulation of unfair prejudice is that jurors will use evidence from one crime to infer criminality on the part of the defendant; that is, apply weak circumstantial reasoning to reach their verdict by reasoning of 'he did it once, he will do it again' or 'he's the type of person who would do such a thing', rather than systematically considering the probative value of the evidence pertaining to each different count.⁴⁷ Although similar to 'moral' prejudice, this conception of prejudice avoids the implication that decisions deriving from normative values are automatically prejudicial. A variation of character prejudice raised by the ALRC is whether jurors' anger or outrage over the defendant's previous criminal misconduct will, consciously or unconsciously, lead them to punish the defendant for those prior actions rather than the charges they have been tasked to evaluate.⁴⁸ Some commentators have suggested that jurors who are susceptible to this form of improper reasoning may be satisfied with less rigorous evidence and may reduce the standard of proof because of the other evidence against the

⁴⁵ Leipold and Abbasi, above n 17. See also Tamara Rice Lave and Aviva A Orenstein, 'Empirical Fallacies of Evidence Law: A Critical Look at the Admission of Prior Sex Crimes' (2013) 81(3) *University of Cincinnati Law Review* 795, 798.

⁴⁶ Tanford, Penrod and Collins, above n 41, 320; See also Bordens and Horowitz, above n 41; Leipold and Abbasi, above n 17.

⁴⁷ See also Lave and Orenstein, above n 45, 798–9.

⁴⁸ *Evidence Interim Report*, above n 23, 800–3, citing Eggleston, above n 34; Lempert and Saltzburg, above n 34, 151–3.

defendant.⁴⁹ Diluted slightly, the hypothesis is that jurors will be less cautious about convicting a defendant with prior criminal allegations since “their consciences are eased because they know that the defendant is not a blameless character”.⁵⁰

Whatever the precise mechanism, the research question is whether jurors in joint trials are more prone than those in separate trials to convict on the basis that the defendant has a ‘criminal disposition’⁵¹, rather than based on the evidence tendered for each of the crimes charged.

2.2.2 Dual process theory

Dual process models of decision making⁵² provide a useful framework for evaluating the extent to which any of three hypothesised impermissible reasoning processes in joint trials might be prejudicial to a defendant. Dual process models of decision making – such as the Elaboration Likelihood Model⁵³ and the Heuristic-Systematic Model⁵⁴ – have been applied to jury decisions⁵⁵ as they are well suited to examining the influence of persuasive communications and information on jury decisions. During the course of a trial, the jury is exposed to a series of persuasive messages from legal counsel, witnesses and the trial judge, and from other jurors during their deliberations. Dual process models distinguish two ways in which decisions are made: (a) the central route, which involves more effortful thinking and is used when juries have the cognitive capacity and time to think about an issue that they regard as important; and (b) the peripheral route, which involves less thinking and is used more frequently when juries are distracted, or have limited time to think about the message content or a topic they view as less

⁴⁹ Richard O Lempert, Samuel R Gross and James B Liebman, *A Modern Approach to Evidence* (Gale Cengage Publishing, 3rd ed, 2000); John Kaplan, ‘Decision Theory and the Fact-Finding Process’ (1968) 20(6) *Stanford Law Review* 1065.

⁵⁰ Lave and Orenstein, above n 45, 799.

⁵¹ Irwin A Horowitz and Kenneth S Bordens, ‘Prejudicial Joinder of Multiple Offenses: Relative Effects of Cognitive Processing and Criminal Schema’ (1986) 7(4) *Basic and Applied Social Psychology* 243; Tanford, Penrod and Collins, above n 41, 320; See also, Bordens and Horowitz, above n 41.

⁵² Shelley Chaiken and Yaacov Trope, *Dual Process Theories in Social Psychology* (Guilford Press, 1999); Eliot R Smith and Jamie DeCoster, ‘Dual Process Models in Social and Cognitive Psychology: Conceptual Integration and Underlying Memory Systems’ (2000) 4(2) *Personality and Social Psychology Review* 108.

⁵³ Nancy Pennington and Reid Hastie, ‘Evidence Evaluation in Complex Decision Making’ (1986) 51(2) *Journal of Personality and Social Psychology* 242.

⁵⁴ Shelley Chaiken, Akiva Liberman, and Alice H Eagly, ‘Heuristic and Systematic Information within and beyond the Persuasion Context’ in Jim S Uleman and John A Bargh (eds), *Unintended Thought* (Guilford Press, 1989) 212; Richard E Petty and John T Cacioppo, ‘The Elaboration Likelihood Model of Persuasion’ in Leonard Berkowitz (ed), *Advances in Experimental Social Psychology* (vol 19) (Academic Press, 1986).

⁵⁵ Brian H Bornstein and Edie Greene, ‘Jury Decision Making: Implications for and from Psychology’ (2011) 20(1) *Current Directions in Psychological Science* 65.

significant.⁵⁶ Joint trials typically involve a greater amount of evidence than separate trials, so juries may be more likely to rely on peripheral route or heuristic reasoning in those cases.⁵⁷

Conflation of evidence, accumulation prejudice and character prejudice are examples of more peripheral than central reasoning. As information becomes more complex and the cognitive load on working memory is increased, it becomes difficult to evaluate the information systematically, and people rely increasingly on strategies or ‘heuristics’ to assess information.⁵⁸ Heuristics are mental shortcuts derived from experience that ease the cognitive load of making a decision. The more jury reasoning is based on heuristic reasoning strategies – and the less it is based on a systematic consideration of the evidence – the greater the potential unfair prejudice to the defendant. For example, a jury reasoning that because the defendant is facing multiple charges ‘he or she must be guilty of something’ would be improper, because the ultimate verdict is unduly influenced by weak, ‘extra legal’, circumstantial reasoning, and non-systematic consideration of the evidence. Thus, one useful approach in assessing the prejudice of joint trials is to compare the extent to which jurors engage in more effortful and systematic reasoning – as opposed to less cognitively demanding heuristic reasoning – compared to jurors in separate trials.

However, it does not follow that all heuristic reasoning amounts to unfair prejudice against a defendant. Heuristics are a normal feature of cognitive adaptive functioning⁵⁹, and increased reliance on heuristics is a typical adaptation to increased factual complexity.⁶⁰ For example, numerous studies have demonstrated that legally trained professionals such as judges and magistrates appropriately rely on heuristics to make sentencing decisions when pressured by time constraints and a high volume of cases.⁶¹ Programmatic research has challenged the view that heuristic decisions imply more errors than logical or ‘rational’ decisions, by showing that ignoring some of the information, rather than weighing all the options, can lead to more accurate

⁵⁶ Ian Freckelton, Jane Goodman-Delahunty, Jacqueline Horan and Blake McKimmie, *Expert Evidence and Criminal Jury Trials* (2016, Oxford University Press).

⁵⁷ Irwin A Horowitz and Kenneth S Bordens, ‘The Consolidation of Plaintiffs: The Effects of Number of Plaintiffs on Jurors’ Liability Decisions, Damage Awards, and Cognitive Processing of Evidence’ (2000) 85(6) *Journal of Applied Psychology* 909.

⁵⁸ Richard E Petty and John T Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (Springer-Verlag, 1985).

⁵⁹ Gerd Gigerenzer, ‘Why Heuristics Work’ (2008) 3 *Perspectives on Psychological Science* 20.

⁶⁰ Gerd Gigerenzer, P M Todd & The ABC Research Group, *Simple Heuristics that Make us Smart* (Oxford University Press, 1999).

⁶¹ Bettina von Helversen and Jorg Rieskamp, ‘Predicting Sentencing for Low Level Crimes: Comparing Models of Human Judgment’ (2009) 15 *Journal of Experimental Psychology: Applied* 375; Mandeep K Dhali, ‘Psychological Models of Professional Decision Making’ (2003) 14 *Psychological Science* 175.

decisions, including decisions in legal contexts where samples are small and there is low predictability.⁶²

In sum, the mere fact that juries use heuristic strategies in joint trials and the mere fact that conviction rates are higher in joint trials do not, on their own, determine unfair prejudice against a defendant. Heuristic strategies may also be adaptive and appropriate, and higher conviction rates in joint trials may also result from permissible reasoning based on the probative value of other complainants' evidence regarding the defendant's tendency and character.

2.3 Prior research on jury decision making in joint trials

Researchers have used three main methodologies to study the potentially prejudicial effects of joint trials: archival studies, experimental trial simulation studies and meta-analyses. Each of these approaches has its strengths and weaknesses, which vary according to the method and scientific rigour applied. In particular, some paradigms of research typically provide greater insight into the extent of prejudice, if any, engendered by heuristic reasoning in joint trials. We will discuss the results of studies within each of these paradigms to examine the extent to which they support any of the three types of impermissible reasoning and unfair prejudice.

2.3.1 Archival studies

Archival studies entail the systematic analysis of actual court verdicts and records⁶³, typically with large samples. The advantage of such studies is that the observed relationships can be generalised across all jury trials with greater confidence than, for example, relationships between variables observed in a single trial or simulation. No two real trials will be exactly the same, so a finding that is robust across many trials is more likely to be broadly applicable to all relevant jury trials.

One strength of archival studies is that they evaluate the verdicts of real-life juries, which have greater gravity due to their binding consequences. This is a feature that experimental trial simulations are less able to emulate.

⁶² Gerd Gigerenzer and Wolfgang Gaissmaier, 'Heuristic Decision Making' (2011) *62 Annual Review of Psychology* 451.

⁶³ Norbert L Kerr and Robert M Bray, 'Simulation, Realism and the Study of the Jury' in Neil Brewer and Kipling D Williams (eds), *Psychology and Law: An Empirical Perspective* (Guilford Press, 2005) 322, 327.

Conviction rates

Archival studies have been used extensively to examine the relationship between adducing evidence of a defendant's prior convictions and jury verdicts. Almost unanimously, they have demonstrated elevated conviction rates where a defendant's prior conviction is adduced. This finding has implications for joint trials involving multiple complainants, as the law assumes that evidence of a defendant's prior convictions has the same prejudicial effect as evidence from multiple complainants.⁶⁴ One seminal archival study⁶⁵, compared more than 3,576 criminal jury verdicts in the US to the presiding judges' own opinions about the appropriate case verdict.⁶⁶ The comparison revealed that conviction rates were 27 per cent higher in cases where the defendant's criminal record was known to the jury. Subsequent studies revealed that these differences were greater when the other evidence in the case did not support a conviction, thus giving some support to the accumulation prejudice hypothesis.⁶⁷ However, the conviction rates for non-testifying defendants with unknown priors – compared to defendants with known priors – were very similar.⁶⁸ Furthermore, juries were more likely to convict if the defendant had numerous prior convictions, even if these past convictions were not revealed to the jury.⁶⁹

Subsequent studies suggest that the prejudice to the defendant of a prior criminal history can be greater if their criminal record is *not* admitted in evidence. An archival study across several US states discovered that juries acquitted in 42.6 per cent of cases where evidence of the defendant's criminal record (if any) was not admitted, whereas judges acquitted in 27.6 per cent of these cases.⁷⁰ By comparison, judges and juries acquitted at a similar rate if the defendant's prior criminal record was tendered. Findings such as these led Lord Atkinson to state in his review of the criminal procedure of England and Wales that:⁷¹

⁶⁴ Christopher M Shanahan, *Bias Arising from Mock juror Decision Making in Joint Criminal Trials*, (Master of Psychology (Forensic) Thesis, Charles Sturt University, 2000).

⁶⁵ Harry Kalven Jr and Hans Zeisel, *The American Jury* (University of Chicago Press, 1966) 143–6. See also Valerie P Hans and Neil Vidmar, 'The American Jury at Twenty-Five Years' (1991) 16(2) *Law and Social Inquiry* 323.

⁶⁶ Leipold and Abbasi, above n 17.

⁶⁷ Theodore Eisenberg and Valerie Hans, 'Taking a Stand on Taking the Stand: the Effect of a Prior Criminal Record on the Decision to Testify and on Trial Outcomes' (2009) 94 *Cornell Law Review* 1353.

⁶⁸ *Ibid.*

⁶⁹ Martha A Myers, 'Rule Departures and Making Law: Juries and Their Verdicts' 13 *Law and Society Review* 781, 792–3.

⁷⁰ Daniel Givelber and Amy Farrell, 'Judges and Juries: The Defence Case and Differences in Acquittal Rates' (2008) 33(1) *Law and Social Inquiry* 31.

⁷¹ Robert Auld, *Review of the Criminal Courts of England and Wales* (Stationary Office, 2001) 564.

... the reality of the present law is that it mostly does not conceal from the tribunal of fact that a defendant has some, though not precisely what, criminal record. In the resultant scope for speculation, it is thus capable of engendering as much or more prejudice against him [as the admission of priors would]. And it is not an honest system in that it does not do what it is claimed to do.

Archival analyses of joint trials and separate trials have also shown that juries were more likely to convict when multiple counts were joined.⁷² An examination of 19,057 US trials revealed that defendants in joint trials were 9 per cent more likely to be convicted of the most serious charge than a defendant tried on a single count. The conviction rate rose further if a second count was charged, and again if a third count was tried, but plateaued after that point for additional charges.⁷³ This finding was replicated in an analysis of 4,310 United Kingdom (UK) rape trials, where the probability of conviction rose significantly as the number of charges increased, thus giving some credence to the accumulation prejudice hypothesis.⁷⁴

Cases with one count versus five counts of rape had a 40 per cent and 80 per cent conviction rate respectively, although this effect plateaued in cases with more than five counts.⁷⁵ In a study conducted by the Judicial Commission of NSW, the proportion of guilty and not guilty verdicts for child sexual abuse cases in joint trials was quite close, while the vast majority of separate trials resulted in verdicts of not guilty, indicating a significant correlation between joinder of charges and higher conviction rates.⁷⁶

Limitations of archival and field studies

From the results of archival studies, we can only draw very limited conclusions about the prejudice engendered by evidence of a defendant's other criminal misconduct generally – and in joint trials specifically. The mere fact that conviction rates are generally higher in joint trials tells us very little about their assumed prejudicial effect. Judges presume that the prejudicial effect of joint trials derives from an increase in impermissible jury reasoning; that is, inter-case conflation of the evidence, accumulation prejudice and/or character prejudice. But these studies

⁷² Leipold and Abbasi, above n 17.

⁷³ Ibid.

⁷⁴ Cheryl Thomas, *Are Juries Fair*, (Research Report No 1/10, Ministry of Justice, February 2010).

⁷⁵ Ibid.

⁷⁶ Patricia Gallagher, Jennifer Hickey and David Ash, *Child Sexual Assault: An Analysis of Matters Determined in the District Court of New South Wales during 1994 (1997)* Research Monograph 15, Judicial Commission of New South Wales. The study examined 201 child sex offences with multiple victims, of which 158 offences were heard in joint trials and 43 in separate trials. Differences in the type of cases approved for joint or separate trials – and those tried by judge as opposed to jury – were not specified.

do not reveal the extent to which the observed increases in conviction rates in joint trials can be attributed to any of the three hypothesised sources of unfair prejudice.

The core of the problem is that a comparison of verdicts in joint trials versus separate trials – or cases in which the defendant’s criminal record was or was not adduced – cannot reveal a causal relationship between joinder (or the admissibility of prior criminal misconduct) and conviction rates. Real-life trials involve unique and highly complex variables. No archival study can exclude the possibility that differences in verdicts were influenced by numerous other confounding variables.⁷⁷

Likewise, establishing a causal relationship between joinder of multiple counts and higher conviction rates – or evidence of a defendant’s other criminal misconduct and higher conviction rates from archival studies – is challenging. Many potentially confounding variables cannot be controlled, manipulated or eliminated in archival studies. This is further problematic because cases in which the defendant does or does not have a criminal record are likely to differ in a number of ways other than the evidence presented of a defendant’s criminal history. For instance, a person’s criminal record is likely to be a factor in deciding who is investigated by police, is targeted for prosecution or receives a plea bargain. Indeed, the finding that juries were more likely to convict if the defendant had numerous prior convictions, regardless of whether these convictions were disclosed to the juries, suggested that the prosecution’s case tends to be stronger in cases involving defendants with a prior record even without the record itself being tendered as evidence. Likewise, archival studies comparing joint and separate trials cannot reveal whether joinder of charges is unfairly prejudicial to the defendant, as they cannot eliminate the possibility that some other factor – for example, the fact that prosecution evidence was generally stronger for each charge in a case involving multiple counts – influenced the outcome of cases.⁷⁸

Archival studies are further limited by the fact that jury deliberations are conducted in secret. Attempts to observe the reasoning process of actual juries other than based on results alone face

⁷⁷ For instance, a field study comparing black and white defendants showed that black defendants received longer sentences (Henry A Bullock, ‘Significance of the Racial Factor in the Length of Prison Sentences’ (1961) 52 *Journal of Criminal Law and Criminology* 411). However, subsequent studies suggested that this difference was largely attributable to the defendants’ prior criminal history and the severity of the charge alleged (Edward Green, ‘Inter- and Intra-Racial Crime Relative to Sentencing’ (1964) 55(3) *Journal of Criminal Law, Criminology and Police Science* 348).

⁷⁸ A practical tool to assess the internal validity and methodological strength of prior studies is the Maryland Scale of Scientific Methods developed by Lawrence W Sherman, Denise C Gottfredson, Doris L MacKenzie, John Eck, Peter Reuter, and Shawn D Bushway, ‘Preventing Crime: What Works, What Doesn’t, What’s Promising,’ (1998) *National Institute of Justice Research in Brief* 4.

serious legal and practical barriers. Thus, archival studies are unable to observe the actual reasoning and deliberative process of juries – arguably one of the most fruitful avenues for observing the extent to which jurors engage in unfairly prejudicial reasoning in joint trials.

2.3.2 Experimental trial simulations

Experimental trial simulations are one of the most prominent research methods in research on juries. A key criterion for increasing the value of experimental trial simulations for the purpose of law reform is to focus less on what have been identified as ‘estimator’ variables and more on what are called ‘system variables.’⁷⁹ Estimator variables are features of trials that may influence conviction rates, but which courts in the real world cannot control, such as the characteristics of the defendant, witnesses or evidence in a case. Research focused on ‘system’ variables that are under the control of the court or the litigants – such as the type of trial, the nature of jury directions, and the use of jury aids such as question trails – provides a more fruitful avenue for effecting policy changes.

This research strategy has several significant advantages over archival studies. Crucially, causal conclusions can be more readily drawn from trial simulations because researchers control and construct the elements of a trial that they are interested in studying.⁸⁰ Inferences about the causal relationships between variables of interest – for example, the influence of joinder of counts on conviction rates – can be determined with greater confidence than in archival studies because researchers are able to reduce the extraneous ‘noise’ present in real trials. Because the only differences across experimental conditions are manipulated by the researcher/s prior to observing the behaviour in interest, researchers can isolate whether these differences caused any observed differences in jury reasoning and case outcomes.

Another advantage is that the identical research problem can be replicated multiple times in an experimental simulation, whereas archival studies will always be beset by the possibility that the trials differed based on some confounding variable (such as the number of witnesses, juror demographics or personalities, or the strength of the prosecution case) as well as the variable of interest (joinder of charges). By contrast, a videotaped trial simulation can be shown to many

⁷⁹ Margaret Bull Kovera, ‘Evaluating Jury Research’ (Speech delivered at The Third International Conference on Empirical Studies of Judicial Systems: Citizen Participation around the World, Taipei, Taiwan, 5–6 September 2014) 14.

⁸⁰ Reid Hastie, Steven Penrod, Nancy Pennington, *Inside the Jury* (Harvard University Press, 1983) 327.

mock jurors or mock juries. Doing so reduces the possibility that an extraneous variable influenced the outcomes.⁸¹

Trial simulations also have the methodological advantage of facilitating direct observations of the process of jury decision making, as well as the outcome of the trial.⁸² As was noted above, in Australia it is not legally feasible to observe the deliberation or decision making processes of actual juries. Experimental simulations offer a unique opportunity to systematically observe the factors underpinning jury decision making and to assess their relative influence on the ultimate outcome of the case.

The extent to which the results of trial simulations apply to real-world trials depends on two main factors. The first is internal validity or the extent to which the behaviour observed in simulations can be attributed to experimental manipulations of the relevant variables. A trial simulation would have high internal validity if, for example, an increase in guilty verdicts could be attributed to changes in the order of charge presentation, the strength of prosecution evidence, and so on, but not some other confounding variable. The second factor is external validity – that is, the extent to which the findings in a simulation apply beyond the context of that particular experiment. Internal validity is necessary, but not sufficient, to establish external validity: “if random or systematic error makes it impossible for the experimenter to draw any conclusions from the experiment, the question of the generality of these conclusions never arises”.⁸³

To date, nine published studies have used the trial simulation paradigm to examine the reasoning process of jurors in joint trials compared to separate trials. All but one of these studies was conducted in the 1980s.⁸⁴ Each study applied roughly the same paradigm. First, case summaries of charges involving the same defendant were pre-tested to determine the perceived evidentiary strength of the prosecution case for each charge⁸⁵, although not all studies included this step. Second, mock jurors were presented with a simulated trial, either in the form of a joint

⁸¹ Ibid 327.

⁸² Ibid 328.

⁸³ Elliot Aronson, Timothy D Wilson and Marilyn B Brewer, ‘Experimentation in Social Psychology’ in Daniel T Gilbert, Susan T Fiske and Gardner Lindzey, *The Handbook of Social Psychology* (McGraw Hill, 4th ed, 1998) 85.

⁸⁴ Norbert L Kerr and Sawyers, ‘Independence of Multiple Verdicts Within a Trial by Mock Jurors’ (1979) 10 *Representative Research in Social Psychology* 16 is the sole exception; Sarah Tanford and Stephen Penrod, ‘Social Interference Processes in Juror Judgments of Multiple-Offense Trials’ (1984) 47 *Journal of Personality and Social Psychology* 749.

⁸⁵ James Farrin, ‘Rethinking Criminal Joinder: An Analysis of the Empirical Research and its Implications for Justice’ (1989) 52(4) *Law and Contemporary Problems* 325.

trial, or a separate trial. In other words, a single charge against a defendant was presented to some participants (the control group), while the same charge (the focal charge) was presented to other participants along with one or more additional charges (the joint trial condition/s).⁸⁶ In some studies, the joint charges were for similar offences (such as multiple burglaries) and in other studies, for dissimilar offences (such as rape and homicide). Finally, participants were asked to deliver a verdict, generally 'guilty' or 'not guilty', although some mock jurors rated the defendant's guilt on a linear scale. Some researchers also asked jurors further questions that targeted specific sources of prejudice. These tasks often included memory recall or recognition tests for the evidence presented in support of each charge (targeting the 'inter-case conflation of the evidence' hypothesis), or assessments of the defendant's personal qualities (targeting the 'character prejudice' hypothesis). The studies then compared the verdict and other outcomes obtained in the joint trial condition with the results obtained for the trials in which charges were prosecuted separately. In one study, although the mock jurors participated in a group deliberation, they returned individual verdicts. Thus, in all of the prior simulation experiments on joint trials, the unit of analysis was the decision of each individual mock juror; none analysed the jury's decision-making process.

Conviction rates

With few exceptions, results of these studies demonstrated 'a joinder effect', in that joining multiple charges in a single trial increased the rate of conviction for the focal charge compared to trials where the charges were tried separately.⁸⁷ Joinder of charges increased conviction rates in studies that used different types of stimulus materials (written, audio or video summaries), and participant mock jurors (undergraduates or jury-eligible citizens).

Researchers have typically taken these findings on conviction rates to imply that joint trials are unfairly prejudicial to defendants, as the same evidence was presented for the focal charge in both the joint and separate trial simulations. According to these studies, the rate of conviction

⁸⁶ Ibid.

⁸⁷ See, eg, Kerr and Sawyers, above n 84; Norbert L Kerr, Douglas L Harmon and James K Graves, 'Independence of Multiple Verdicts by Jurors and Juries' (1982) 12(1) *Journal of Applied Social Psychology* 12; Irwin A Horowitz, Kenneth S Bordens and Marc S Feldman, 'A Comparison of Verdicts Obtained in Severed and Joined Criminal Trials' (1980) 10 *Journal of Applied Social Psychology* 444; Horowitz and Bordens, above n 51; Bordens and Horowitz, above n X41; Edith Greene and Elizabeth F Loftus, 'When Crimes are Joined at Trial' (1985) 9(2) *Law and Human Behaviour* 193; Tanford and Penrod, above n 41; Tanford and Penrod, above n 81; Tanford, Penrod and Collins, above n 41. Irwin A Horowitz and Kenneth S Bordens also considered the influence of the number of plaintiffs in a civil case on juror verdicts and awards: 'The Consolidation of Plaintiffs: The Effects of Number of Plaintiffs on Jurors' Liability Decisions, Damage Awards, and Cognitive Processing of Evidence' (2000) 85(6) *Journal of Applied Psychology* 909.

should not have varied for the focal charge if the jury in the joint trial condition was not reasoning in a manner unfairly prejudicial to the defendant. However, this interpretation vastly oversimplifies legal concerns about prejudice. As discussed above, the unfairly prejudicial effect to the defendant is not based on increased conviction rates in joint trials alone, but on specific forms of impermissible prejudicial reasoning that lead to increased convictions rates. Moreover, binary verdict measures are a blunt instrument, and are uninformative about any reasoning processes, unfairly prejudicial or otherwise. The real concern with joint trials is that jurors are more likely to engage in one or more of the three identified types of unfairly prejudicial reasoning. As such, these trial simulation studies are only capable of informing the question of unfair prejudice to the extent that they evaluated the prevalence of unfairly prejudicial reasoning in joint trials compared to separate trials. Next, we examined the outcome variables other than verdict used in prior experimental simulation studies designed to target specific sources of unfair prejudicial reasoning.

Prejudicial reasoning

Inter-case conflation of the evidence

Prior research has tested the inter-case conflation of the evidence hypothesis by assessing mock jurors' factual recall of the evidence presented for each charge, although not all studies included these measures. Some studies that included these measures found no evidence of confusion.⁸⁸ Although several studies revealed that joinder of charges was associated with elevated levels of errors in recognition or free recall tasks, only one study associated these errors with increased perceptions of the liability of the defendant (in a civil trial in which the joined charges were similar to the focal charge) but these errors had no effect on mock jurors' verdicts.⁸⁹ In fact, one study found that evidentiary conflation occasionally weighed in favour of the defendant.⁹⁰

The absence of findings of inter-case conflation may be due to methodological attributes in the prior studies. The complexity of the information in the joined trials varied widely. In some studies, only one additional charge was joined, while others included several additional charges of a similar and dissimilar nature. In some studies the written case descriptions were very brief, reducing the possibility of evidentiary conflation.⁹¹ Others used a free recall task, which did not specifically target the recollection of evidence that influenced jurors' ultimate verdicts.⁹²

⁸⁸ Greene and Loftus, above n 87.

⁸⁹ Tanford, Penrod and Collins, above n 41.

⁹⁰ Greene and Loftus, above n 87.

⁹¹ Ibid.

⁹² Id.

Overall, the existing research – and in particular, the studies in which similar as opposed to dissimilar charges were joined – did not support the hypothesis of unfair prejudice to the defendant by jurors’ and juries’ inter-case conflation of the evidence.

Accumulation prejudice

There is mixed empirical support for the proposition that joint trials are unfairly prejudicial to defendants because jurors will tend to accumulate evidence impermissibly across each of the charges. Most studies revealed no accumulation effect in the form of elevated conviction rates⁹³, although one study demonstrated that convictions increased for the weaker but not the stronger charges in the joint versus the separate trials.⁹⁴

Another study demonstrated that joinder of charges had a small influence on mock jurors’ perception of the strength of both prosecution and defence evidence, which appeared to reflect their logical inferences in response to the additional presentation of evidence by both parties in the joint trial.⁹⁵ However, a study that examined reasoning processes by asking mock jurors for their ‘thoughts’ regarding evidence presented in the simulation, discovered no significantly more favourable evaluations of either prosecution or defence evidence in joint trials compared to separate trials.⁹⁶

A further apparently logical inference reported in some studies, depending on the similarity of the joined charges, was that mock jurors rated prosecution evidence stronger in joint trials, compared to the same evidence presented in separate trials,⁹⁷ and the concomitant perception of defence evidence as weaker in joint trials. These findings indicated that mock jurors were more likely to accumulate inculpatory than exculpatory evidence in joint trials, as might be expected.

One topic related to the ‘accumulation prejudice’ hypothesis that has not yet been thoroughly empirically tested is whether juries ‘accumulate’ the number of witnesses rather than the content of the witness statements. In an experimental simulation in which the joinder of multiple plaintiffs was manipulated in a civil trial, the liability of the defendant increased as every plaintiff was added, up to a maximum of four.⁹⁸ In this case, however, the number of claims

⁹³ Tanford and Penrod, above n 81; Bordens and Horowitz, above n 41; Tanford, Penrod and Collins, above n 41.

⁹⁴ Kerr and Sawyers, above n 84.

⁹⁵ Bordens and Horowitz, above n 41.

⁹⁶ Tanford and Penrod, above n 81.

⁹⁷ Ibid.

⁹⁸ Horowitz and Bordens, above n 57.

and number of witnesses was confounded. If juries were using this kind of heuristic processing in their decision making, the number of witnesses rather than the content of the witness statements would be more influential on their verdicts. On the other hand, juries that exhibited systematic processing would use the content of cumulative witness evidence, not the number of witnesses alone, in their decision making.

In sum, there is limited support for the hypothesis that joint trials pose a risk of unfair prejudice to the defendant because of improper accumulation of charges or witnesses.

Character prejudice

Studies that demonstrated that complexity of evidence leads to reliance on heuristics and stereotypes in decision making suggested that because jurors in joint trials are evaluating more evidence than those in separate trials, they may favour the explanation that the defendant has a criminal disposition because it is the simplest explanation available.⁹⁹ Prior simulation studies typically evaluated this hypothesis by comparing mock juror verdicts with their ratings of the defendant's character. Most findings disclosed that mock jurors in joint trials were more likely to give unfavourable ratings of the defendant's criminal character.¹⁰⁰ One study asked participants to rate the defendant based on three dimensions: dangerousness, likeability and believability.¹⁰¹ Defendants who were tried in a joint trial were rated less favourably on all dimensions¹⁰², compared to those tried in separate trials.

Some studies have used path analyses to evaluate how mock jurors' perceptions of the defendant's criminal character influence jury verdicts. In one study, correlation of jurors' verdicts – and their subsequent ratings of the defendant's character on an 11-point scale for factors including honesty, dangerousness and likelihood to commit future crime – revealed that participants in joint trials were more likely to attribute criminal characteristics to the defendant, and more likely to convict, compared to separate trials. As a result, the researchers posited that participants in joint trials were more likely to form a view of the defendant having a criminal disposition relatively early in the trial process, and to evaluate subsequent evidence in light of this 'criminal schema'.¹⁰³ However, path analyses demonstrated only a *correlation* between the

⁹⁹ Shanahan, above n 64, 15.

¹⁰⁰ Horowitz and Bordens, above n 41; Roselle J Wissler and Michael J Saks, 'On the Inefficacy of Limiting Instructions: When Jurors Use Prior Conviction Evidence to Decide on Guilt' (1985) 9(1) *Law and Human Behaviour* 37.

¹⁰¹ Greene and Loftus, 'When Crimes are Joined at Trial' above n 87.

¹⁰² *Ibid.*

¹⁰³ Horowitz and Bordens, above n 51.

verdicts and subsequent ratings of the defendant's character, not a causal relationship between these factors. Indeed, it can be just as credibly argued that participants systematically reasoned that the defendant had a criminal disposition *after* making a judgment about his guilt.¹⁰⁴ Thus, there is limited empirical evidence supporting the prejudicial effects on verdicts as a result of heuristic processing of the defendant's criminal disposition.

In a trial simulation study in which participants evaluated the probability of the defendant's guilt based solely on the number and nature of charges alleged, mock jurors in the joint trial conditions were not significantly more likely to perceive the defendant as guilty. This finding was in contrast to what would be expected if jurors' decisions were guided by character prejudice because of the number of charges, rather than a systematic evaluation of the evidence itself. On the other hand, verdicts and defendant character evaluations were affected to a greater extent by the number of charges the jury actually evaluated, rather than the number of charges the jury was aware of.¹⁰⁵ This indicated that mere knowledge of the charges was insufficient to affect verdicts and that jurors were more likely to develop a 'criminal schema' when actually evaluating the charges during the deliberation phase.¹⁰⁶

Research has suggested that defendants in child sexual abuse cases may be particularly susceptible to prejudice based on their criminal disposition. For example, one trial simulation found that defendants with prior child sexual abuse convictions were viewed by mock jurors as more likely than defendants with other types of prior convictions to commit the crime alleged, to have escaped punishment in the past, to lie in court, to be untrustworthy and to be more deserving of punishment.¹⁰⁷

In summary, even though the existing empirical research has generally established a correlation between participants' adverse inferences about a defendant's criminal disposition and guilty verdicts, it has not established a causal relationship between these factors.¹⁰⁸

¹⁰⁴ Irwin A Horowitz and Kenneth S Bordens, 'Joinder of Criminal Offences: A Review of the Legal and Psychological Literature' (1985) 9 *Law and Human Behaviour* 349.

¹⁰⁵ Horowitz and Bordens, above n 51.

¹⁰⁶ *Ibid.*

¹⁰⁷ Sally Lloyd-Bostock, 'The Effects on Juries of Hearing about the Defendant's Previous Criminal Record: A Simulation Study' [2000] *Criminal Law Review* 734, 741–54.

¹⁰⁸ Greene and Loftus, above n 87, 204–5. It is not difficult to construct an experiment to test for this causal connection. For example, the study could incorporate ratings of the defendant at various points throughout a joined trial: Bordens and Horowitz, above n 41, 349.

Limitations of prior trial simulation studies of joinder

As the preceding discussion indicated, few conclusions can be drawn from the existing research about the influence (if any) of unfairly prejudicial reasoning processes underlying jury decision making on verdicts in joint trials compared to separate trials. Interestingly, although many of these studies were examining what legal psychologists refer to as a ‘system’ variable – namely the very nature of the trial procedure itself – they tested few related system variables in conjunction with this manipulation. Only four of the studies provided standard appropriate jury directions to test the effectiveness of limiting jury directions when charges were joined. None of the prior studies used question trails. Only one study included group deliberations in the trial simulation, where adults in the community served as mock jurors. That study made no effort to record and analyse the content of the juries’ discussions.

Finally, many of the prior studies had weak ecological validity due to a number of methodological limitations, such as the use of brief written trial summaries or audiotaped materials, and the reliance on undergraduate students as mock jurors. Only one of the prior studies used realistic video trial simulations lasting 30 to 45 minutes in the separate trial versions, and 90 to 120 minutes in the joint trial versions. These features of the past studies further limit the extent to which they can inform policy changes.

2.3.3 Meta-analyses of trial simulation studies

Meta-analyses have been used to estimate the robustness of the joinder effect across trial simulation studies. Meta-analyses quantitatively collate results of existing empirical studies. In the joint trial simulation context, this essentially involves mathematically averaging results from existing trial simulations to estimate the relationship between the joinder of charges and trial verdicts. The idea is that this relationship can then be more accurately generalised across all jury trials, rather than one particular simulation, because errors with individual studies will typically be balanced out.¹⁰⁹ For example, a recent meta-analysis of 272 jury simulation studies revealed the surprisingly small but nonetheless significant influence of prior criminal records on jury verdicts, showing that defendants whose prior criminal record was known to juries were

¹⁰⁹ Dennis J Devine, *Jury Decision Making: The State of the Science*, above n 44, 13.

somewhat more likely to be convicted.¹¹⁰ The researchers noted that the effect was likely to be moderated by other factors, such as the similarity of the prior offence to the current charge.

We were only able to locate two meta-analyses on joinder. In the first, the researchers calculated the mean correlation between joinder of charges and conviction on the focal charge, to exhibit a moderately strong effect.¹¹¹ The second aggregated results of 10 jury simulation studies conducted on joinder between 1980 and 1985.¹¹² Most notably, significant joinder effects emerged only when three identical or similar offences were joined at trial, and not when the charges were dissimilar.

The importance of this finding is that the conviction rates increased only when there was a logical connection between the types of charges presented in a joint trial, and not when they were not logically related. For instance, the joinder effect occurred in a joint trial of a defendant for three service station burglaries, or three burglaries at different locations. The observation that joinder of charges increased conviction rates (the ‘joinder effect’) was robust, despite the fact that many studies used differing methodologies. These meta-analyses provided a moderately strong indication that conviction rates for the same charge would increase if presented in a joint rather than separate trial. The analyses did not address whether some of these outcomes were the product of appropriate inferences about the inculpatory force of the logically related additional evidence available in the joint trials, as they provided no indication of the reasoning processes that underlie these observed differences in outcomes in joint versus separate trials. As such, the meta-analyses do not inform the question of whether unfair prejudicial reasoning was engendered by the joinder of charges any more than the prior simulations themselves.

And yet, jury researchers have typically taken these findings to imply that joint trials are unfairly prejudicial to defendants, since the same evidence was presented for the focal charge in both the joint and separate trial simulations. This interpretation of the findings is premised on the view that the rate of conviction should not have varied for the focal charge unless the jury in the joint trial condition was reasoning in a manner unfairly prejudicial to the defendant. However, this interpretation vastly oversimplifies legal concerns about prejudice. As discussed

¹¹⁰ Dennis J Devine and D E Caughlin, ‘Do they Matter? A Meta-Analytic Investigation of Individual Characteristics and Guilty Judgments’ (2014) 20 *Psychology Public Policy and Law* 109.

¹¹¹ 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) 231.

¹¹² Tanford, Penrod and Collins, above n 41.

above, the unfairly prejudicial effect to the defendant is not based merely on a finding of increased conviction rates in joint trials; it involves specific forms of inappropriate or impermissible prejudicial reasoning that lead to increased convictions rates. This is because increased conviction rates in joint trials may be a result of more prosecutorial evidence being presented in a joint trial that is probative of the defendant's guilt compared to the evidence presented in a separate trial, a fact that is likely to occur in real-life trials where the prosecution's case can be either weak or strong, depending on the available evidence.

The real concern with joint trials is the risk of jurors being more likely to engage in one or more of the three identified categories of unfairly prejudicial reasoning. As such, the trial simulations in these nine studies are only capable of informing the question of prejudice to the extent they evaluated the prevalence of heuristic or systematic reasoning in joint compared to separate trials.

2.4 Empirical support for legal safeguards against unfair prejudice

The law attempts to curtail the perceived risk of an unfairly prejudicial effect of joint trials – and evidence of a defendant's other misconduct more generally – via one procedural measure: jury instructions, also known as judicial instructions or judicial directions. In addition, the normal process of jury deliberation will usually counteract or correct particular types of unfairly prejudicial reasoning. Next, we discuss the extent to which empirical evidence supports the corrective effect of each of these measures.

2.4.1 Jury instructions

A handful of studies have explored the effect of cautionary or limiting jury instructions on appropriate use of the evidence as a legal remedy for unfairly prejudicial joinder. In three studies, the limiting jury instructions had no effect¹¹³; only one study showed an effect, and that was obtained in the context of a more realistic trial simulation with undergraduate students serving as mock jurors, tested only at the level of individual mock jurors, and not among juries.¹¹⁴ Thus, no research to date has explored the effects on jury reasoning of cautionary jury instructions to guide juries on the use of evidence admitted in joint trials.

¹¹³ Greene and Loftus, above n 87; Tanford and Penrod, 'Social Interference Processes in Juror Judgments of Multiple-Offense Trials', above n 84; Horowitz and Bordens, 'Joinder of Criminal Offences', above n 104.

¹¹⁴ Tanford, Penrod and Collins, above n 41.

A considerable body of empirical research has cast some doubt on how effectively jury instructions aid jurors' understanding of the permissible use of different items of evidence, and by extension, the extent to which these instructions safeguard against unfairly prejudicial reasoning.¹¹⁵ For example, a trial simulation study that presented issues somewhat analogous to those in joint trials showed that jurors were more likely to convict a defendant when given evidence of a prior conviction rather than evidence of a prior acquittal or no prior record. This effect was mediated by mock jurors' attributions about the criminal propensity of the defendant, which replicates the findings regarding the influence of character evidence in joint trials. The judge's limiting instructions – which aimed to guide jurors in the appropriate use of prior record evidence – were ineffective. However, those materials were administered only to individual mock jurors; there was no test of how juries used the instructions in group deliberations.¹¹⁶

2.4.2 Fact-based question trails

One recurring problem with limiting judicial instructions, as with many other types of jury directions, is that they can be difficult to understand and apply. In fact, one very early study found that only one of 18 jurors was able to recall a judge's oral instructions with sufficient detail to follow them correctly.¹¹⁷ Mock jurors' comprehension levels generally ranged between 50 and 70 per cent¹¹⁸, and comprehension of legal rules did not significantly differ between jurors who were given typical oral instructions and those who received no instructions.¹¹⁹

¹¹⁵ E Allan Lind, 'The Psychology of Courtroom Procedure', in Norbert L Kerr and Robert M Bray (eds) *The Psychology of the Courtroom* (Academic Press, 1982) 13; Lora M Levett, Erin M Danielson, Margaret Bull Kovera and Brian L Cutler, 'The Psychology of Juror and Jury Decision Making' in Neil Brewer and Kipling D Williams (eds), *Psychology and Law: An Empirical Perspective* (Guildford Press, 2005).

¹¹⁶ Mary Dodge and Edith Greene 'The Influence of Prior Record Evidence on Juror Decision Making' (1995) 19(1) *Law and Human Behavior* 67–78.

¹¹⁷ Harold M Hoffman and Joseph Brodley, 'Jurors on Trial' (1952) 17 *Modern Law* 235, 245.

¹¹⁸ R P Charrow and V R Charrow, 'Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions' (1979) 79(7) *Columbia Law Review* 1306; Phoebe C Ellsworth, 'Are Twelve Heads Better than One?' (1989) 52 *Law and Contemporary Problems* 205; Edith Greene, 'Judge's Instruction on Eyewitness Testimony: Evaluation and Revision' (1988) 18 *Journal of Applied Social Psychology* 252; Laurence J Severance and Elizabeth F Loftus, 'Improving the Ability of Jurors to Comprehend and Apply Criminal Jury Instructions' (1982) 17 *Law and Society Review* 153; Joel D Lieberman, 'The Psychology of the Jury Instruction Process' in Joel D Lieberman and Daniel A Krauss (eds), *Jury Psychology: Social Aspects of Trial Processes: Psychology in the Court Room, Vol. 1* (Ashgate Publishing, 2009) 129.

¹¹⁹ Geoffrey P Kramer and Doreen M Koenig, 'Do Jurors Understand Criminal Jury Instructions? Analysing the Results of the Michigan Juror Comprehension Project' (1990) 23(3) *University of Michigan Journal of Law Reform* 401.

Comparably poor levels of comprehension have also been observed in Canada and New Zealand.¹²⁰

A distinction must be drawn between understanding a rule, applying a rule, and the capacity to recall and articulate a rule. One study found juries unequivocally recalled responses correctly more than half of the time, but were clearly incorrect 20 per cent of the time.¹²¹ Another study showed that juries' recognition of judicial instructions was better than their recall, and that juries performed better at these tasks than when they were required to apply the directions to a new set of facts.¹²² This is likely due to the fact that reproduction or replication is a high-level task, especially when the rule is expressed in language unfamiliar to jurors. As such, researchers must be cautious about drawing conclusions about comprehension or application from the failure to recall an instruction.

The form and nature of the instructions may have some influence on their efficacy. In other words, low comprehension may be due to linguistic rather than conceptual complexity.¹²³ Plain language judicial instructions have improved comprehension of both procedural and substantive instructions.¹²⁴

The well-documented difficulties that juries experience when trying to understand and apply standard judicial directions have generated interest in fact-based question trails. An emerging body of jury research has examined the effectiveness of fact-based question trails as an aid to jury decision making. So far, promising but limited evidence suggests that jurors apply judicial instructions more accurately when using question trails than without.¹²⁵ One recent Australian study revealed that jurors who were given question trails exhibited significantly better comprehension of complex substantive issues compared to those who received standard jury instructions or no instructions at all.¹²⁶ In a Victorian experimental trial simulation study, 1,007 jury-eligible citizens viewed a videotaped sexual assault trial with two charges against the

¹²⁰ V Gordon Rose and James R P Ogloff, 'Evaluating the Comprehensibility of Jury Instruction: A Method and an Example' (2001) 25 *Law and Human Behaviour* 409; Eames, above n 240; Warren Young, Neil Cameron and Yvette Tinsley, '*Juries in Criminal Justice Trials*' (Research Report No 29, New Zealand Law Commission, 2001).

¹²¹ Ellsworth, above n 118.

¹²² Carolyn Semmler and Neil Brewer, 'Using a Flowchart to Improve Comprehension of Jury Instructions' (2002) 9(2) *Psychiatry, Psychology and the Law* 262.

¹²³ R P Charrow and V R Charrow, 'Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions' (1979) 79(7) *Columbia Law Review* 1306.

¹²⁴ *Ibid.*

¹²⁵ Catriona McKay, Mark Nolan and Michael Smithson, 'Effectiveness of Question Trails and Jury Decision Aids: the Jury's Still Out' (2014) 21(4) *Psychiatry, Psychology and Law* 492.

¹²⁶ Ede and Goodman-Delahunty, above n 16.

defendant (indecent assault and rape).¹²⁷ All mock jurors participated in group deliberations in juries of 10 to 12 persons before returning a verdict. Although that study did not analyse the content of jury deliberations, a comparison of pre-deliberation and post-deliberation measures of comprehension and application of the jury directions showed significant increases in comprehension following deliberation when juries were given a fact-based question trail compared to standard jury directions. Use of the fact-based question trail in deliberation also increased the unanimity of verdicts compared to verdicts based on standard jury directions, although this difference was not statistically significant.

The use of question trails remains untested in the joint trial context, but the existing findings suggest that question trails could potentially reduce the risk of unfairly prejudicial reasoning that may arise from conflation of the evidence, and reliance on heuristics triggered by the increased cognitive load in joint trials. Since few prior experimental simulation studies have incorporated question trails into their design, their potential value as a safeguard against impermissible prejudicial reasoning in joint trials is a matter ripe for investigation.

2.4.3 Group deliberation

Most prior-trial simulation studies did not include group deliberation despite this being a key component of the jury decision-making process.¹²⁸ Where group deliberation was included, only the verdict and not the process of deliberation itself was evaluated. As such, only limited insight was gained from past trial simulations regarding the extent to which deliberation safeguarded against unfairly prejudicial reasoning in joint trials.

Trial simulation studies comparing individual jurors' verdicts and jury verdicts in North America¹²⁹, South Korea¹³⁰ and Taiwan¹³¹ demonstrated a leniency effect associated with jury deliberation. This may be because deliberation caused individual jurors to become more

¹²⁷ B L Spivak, *Asking the Right Questions: Examining the Utility of Fact Based Jury Directions Using a Simulated Trial Paradigm* (PhD thesis, Monash University, 2016).

¹²⁸ Devine, *Jury Decision Making*; above n 44; Mark Nolan and Jane Goodman-Delahunty, *Legal Psychology in Australia* (Thomson-Reuters, 2015) 291.

¹²⁹ Devine, above n 44, M152–180.

¹³⁰ Kwangbai Park, Sangjoon Kim, Eunro Lee and Hyesun Seo, 'Social Conformity and Cognitive Conversion During Jury Deliberations: A Content Analysis of Deliberation Arguments in the First Officially Simulated Jury Trial in Korea' (2005) 19(3) *The Korean Journal of Social and Personality Psychology* 1–21.

¹³¹ Kuo-Chang Huang and Chang-Ching Lee, 'Mock Jury Trials in Taiwan – Paving the Ground for Introducing Lay Participation' (2014) 38(4) *Law and Human Behaviour* 367; Chang-Ching Lin and Kuo-Chang Huang, 'Differential Effects Between Legal and Quantified Definitions of Standard of Proof' (Paper presented at The Third International Conference on Empirical Studies of Judicial Systems: Citizen Participation around the World, Taipei, Taiwan, 5–6 September 2014) 14.

sensitive to the error of wrongful conviction, resulting in more acquittals when comparing verdicts before and after group deliberation and varying the experimental manipulations of deliberation. However, in one Australian simulation where a control group tested the effects of deliberation, there was no leniency effect.¹³² Thus, observed increases in acquittals in trial simulations that involved deliberation may be attributable to other factors. One hypothesis – confirmed by some follow-up research – was that jurors favouring acquittal in deliberation had the most influence with the jury, especially when the jury was required to reach a unanimous (rather than majority) decision.¹³³ Group polarisation effects may account for some of these findings.¹³⁴ Other research demonstrated that deliberation reduced biases present in individual juror verdicts.¹³⁵ In another study, mock jurors also exhibited greater understanding of the trial evidence following group deliberations.¹³⁶ In one study of jury deliberation in a simulated child sexual abuse trial about expert evidence, the researchers coded deliberations to record instances of references to the defendant's moral character, but did not include these results in their published report.¹³⁷

Conversely, research into the impact of introducing evidence of a defendant's prior convictions revealed that conviction rates by mock juries after group deliberation were significantly higher than those by mock jurors who returned individual verdicts without deliberation.¹³⁸ Additionally, a series of joint trial simulations using the same trial simulation materials disclosed a stronger joinder effect when mock jurors engaged in group deliberation.¹³⁹ Consequently, group deliberation appeared to have *some* potentially unfairly prejudicial influence on jury decision making in joint trials. The absence of deliberation in all but one of the previous empirical joint trial stimulation studies is a methodological shortcoming that

¹³² Jane Goodman-Delahunty, Natalie Martschuk and Anne Cossins, 'Programmatic Pretest-Posttest Research to Reduce Jury Bias in Child Sexual Abuse Cases' above n 1.

¹³³ Robert J MacCoun and Norbert L Kerr, 'Asymmetric Influence in Mock Jury Deliberations: Jurors' Bias for Leniency' (1988) 54 *Journal of Personality and Social Psychology* 21; Norbert L Kerr and Robert J MacCoun, 'Is the Leniency Asymmetry Really Dead? Misinterpreting Asymmetry Effects in Criminal Jury Deliberation' (2012) 15 *Group Processes and Intergroup Relations* 585.

¹³⁴ Goodman-Delahunty, Martschuk and Cossins, above n 1. See also, Dan Simon, *In Doubt: The Psychology of the Criminal Justice System* (Harvard University Press, 2012) 201, 367.

¹³⁵ M F Kaplan and L E Miller, 'Reducing the Effects of Juror Bias' (1978) 36 *Journal of Personality and Social Psychology* 1443.

¹³⁶ Ellsworth, above n 118.

¹³⁷ Margaret Bull Kovera, April W Gresham, Eugene Borgida, Ellen Gray and Pamela C Regan, 'Does Expert Psychological Testimony Inform or Influence Juror Decision Making? A Social Cognitive Analysis' (1997) 82(1) *Journal of Applied Psychology* 178–191, 191.

¹³⁸ Valerie P Hans and Anthony N Doob, 'Section 12 of the Canada Evidence Act and the Deliberations of Simulated Juries' (Research Paper No 337, Cornell Law Faculty Publications, 3 January 1976).

¹³⁹ Tanford, Penrod and Collins, above n 41; Tanford and Penrod, above n 84.

inhibits the generalisation of these findings to all jury decisions. Moreover, the previous trial simulations that included deliberation did not analyse the content of the jury deliberations themselves, so we can only draw limited conclusions about the extent to which deliberation mitigated or increased the incidence of prejudicial reasoning in joint trials, and whether any observed joinder effects were in fact based on permissible jury reasoning.

To date, few studies have examined jury deliberation in simulated child sexual abuse trials. One recent exception is an Australian study that incorporated online deliberation¹⁴⁰, which has obvious limitations for comparison with in-person group deliberation. As deliberation is one of the legal procedures expected to reduce jury errors, it is essential to add this legal process to the jury simulation paradigm, to assess its impact on factors that may influence conviction rates.

A social process known as ‘sensemaking’ offers a useful theoretical framework through which to examine the reasoning and deliberation process of mock juries. Broadly speaking, sensemaking is the process by which someone explains previous events, typically under uncertain or ambiguous circumstances.¹⁴¹ Sensemaking suggests that decision makers use a combination of cognitive and social mechanisms to reduce the ambiguity of their situation. The research applying this approach identifies six main types of sensemaking:¹⁴²

- (1) the recognition of a discrepant set of cues in the ongoing flow of events;
- (2) the retrospective consideration of experiences;
- (3) the generation of plausible explanatory speculations;
- (4) enactment through written and oral communication;
- (5) social contact with other individuals and their ideas; and
- (6) the involvement of issues about identity and reputation.

The applicability of these processes in assessing jury deliberations is as follows:

- (1) Different witnesses present discrepant cues in the course of an adversarial trial.
- (2) Juries draw on their common experience to assess and resolve these disparities.
- (3) Jury decision making includes testing the plausibility of explanations for the evidence in issue, often in competing narratives suggested by the prosecution and defence.
- (4) Juries engage in oral communication as a group to assess these explanations.

¹⁴⁰ Samantha J Tabak and Bianca Klettke, ‘Mock Jury Attitudes towards Credibility, Age, and Guilt in a Fictional Child Sexual Assault Scenario’ (2014) 66 *Australian Journal of Psychology* 47.

¹⁴¹ Gary Klein, Sterling Wiggins and Cynthia O Dominguez, ‘Team Sensemaking’ (2010) 11(4) *Theoretical Issues in Ergonomics Science* 304.

¹⁴² Ibid; Ian D Colville, Brown AD and Annie J Pye, ‘Simplexity: Sensemaking, Organizing and Storytelling’ (2012) 65(1) *Human Relations* 5.

- (5) The participation of 12 jurors in the decision-making process requires social contact and engagement with the ideas of fellow jurors to reach consensus.
- (6) The verdicts are made within parameters prescribed by rules of evidence and jury directions, to guide assessment of credibility and plausibility.

This theory suggests that the discursive nature of jury deliberation is likely to have an influential role on jury reasoning and decision making. Two jury researchers have noted that:¹⁴³

[juries] perform their tasks as a group, which means that they can depend on each other to recollect different facts, challenge each other's view and misconceptions, and reach a shared judgement, which requires all of the jurors, who come from different backgrounds and walks of life, to agree that the evidence is strong enough to justify a unanimous verdict.

2.5 Summary

The foregoing review of the literature on joinder identified three central hypotheses associated with potential unfair prejudice in cases that join multiple charges against a defendant in a single trial. From these hypotheses, we derived a series of key research questions. We then discussed the extent to which prior archival studies and jury-trial simulation experiments have addressed those questions.

Although past empirical studies that compared verdicts in separate and joint trials with multiple charges against the same defendant in civil and in criminal cases yielded a moderately strong joinder effect, in general, the studies had two major weaknesses. First, the emphasis on conviction rates – using verdicts as a primary indicator – did not adequately distinguish influences of joinder that were logically related to the issue of guilt from those that were logically unrelated. In other words, verdict alone is too blunt and insensitive an instrument to assess prejudicial reasoning, as it allows for no assessment of legally appropriate prejudice based on additional inculpatory information provided to juries in the joint trials. Second, only four studies included standard jury instructions provided in a joint trial to guide juries on the appropriate uses of the evidence – and where those were provided, the studies did not include group jury deliberation about the evidence, but were conducted exclusively with individual mock jurors. The sole prior study to include deliberation did not examine the content of jury deliberations, so there was no assessment of the juries' reasoning in distinguishing verdicts

¹⁴³ Nancy S Marder and Valerie P Hans, 'Introduction to Juries and Lay Participation: American Perspectives and Global Trends' (2015) 90 *Chicago-Kent Law Review* 789, 799 citing *Ballew v Georgia*, 435 U.S. 223, 232–38 (1978).

based on influences of joinder that were logically related to the issue of guilt from those that were logically unrelated. For these reasons, there are questions around the interpretation of those findings as evidence of unfair prejudice to the accused.

In the present study, we used an experimental jury-trial simulation to further examine the risk of unfair prejudice to a defendant in the context of a child sexual abuse trial. A major advantage of the trial simulation method is that it permits insight into jury reasoning during the group deliberation process. In addition, we examined the effectiveness of jury directions and a question trail to help juries reach a verdict based on evidence that is logically related to the issue of guilt.

Chapter 3: Research aims, method and design

The trial simulation with 90 mock juries aimed to determine whether a joinder effect emerges in a child sexual abuse trial with multiple complainants – and if so, whether it is the result of impermissible prejudicial reasoning. The study tested the influence of evidence strength; the number of charges and witnesses; and whether specific judicial directions and question trails reduced the risk of unfair prejudice to a defendant in a joint trial.

3.1 Aims

This research project aimed to address two questions that were not adequately answered by previous studies:

1. Are joinder effects the result of jury reasoning logically unrelated to the evidence?
2. If so, do jury directions mitigate risks of unfair prejudice in a joint trial?

We sought to answer these questions in an experimental jury-simulation study of a joint trial for alleged child sexual abuse offences with three complainants, versus a separate trial with a single complainant.

The study aimed to:

- determine whether juries engaged in impermissible reasoning in relation to cross-admissible tendency evidence in a joint trial with three complainants, focusing in particular on the incidence of errors based on inter-case conflation of the evidence, accumulation prejudice and character prejudice
- compare the above decision-making processes with those of juries in a separate trial of the same defendant
- compare the outcomes (acquittal, conviction or hung jury) of a joint versus a separate trial
- examine the relationship between jury reasoning, decision making and trial outcomes to determine the effect of jury directions and question trails on jury reasoning and trial outcomes.

Three broad hypotheses guided the research design:

1. Conviction rates in joint trials will be higher than in separate trials, but this increase will not be due to impermissible forms of jury reasoning.

2. Judicial directions will help juries distinguish between evidence in relation to two or more counts of child sexual abuse, and will reduce the risk of convictions based on impermissible reasoning.
3. Question trails will help juries distinguish between evidence in relation to two or more counts of child sexual abuse, and will reduce the risk of convictions based on impermissible reasoning.

3.2 Method

3.2.1 Online mock juror pilot study

Before conducting the trial simulation study, it was necessary to pre-test the mock trial materials to ensure that the strength of the evidence in the cases of three separate complainants was significantly different so that we could assess jury responses to differences in the allegations of each complainant in the context of a joint trial. In addition, the pilot study tested whether the study instruments were sensitive to differences in juror responses to the evidence in the absence of group deliberations.

Based on facts derived from a review of child sexual abuse trials, six mock trial scripts were prepared about the trial of *R v Booth*. This trial contained allegations of sexual abuse by a defendant, Mark Booth, who, as a soccer coach, had a relationship of care and control in an institutional setting with each of three complainants in the 1990s. In all trials, the defendant was charged with indecency in the form of masturbation of the complainant. This was the focal offence for the manipulation check. The six trial summaries presented the following evidence:

1. Separate trial with a single complainant (Simon) who gave weak evidence (779 words).
2. Separate trial with a single complainant (Justin) who gave strong evidence (1,203 words).
3. Separate trial with a single complainant (Timothy) who gave moderately strong evidence (1,217 words).
4. Separate trial with a single complainant (Timothy) who gave moderately strong evidence, plus supporting relationship evidence about the prior grooming acts of the accused (1,285 words).

5. Separate trial with a single complainant (Timothy) who gave moderately strong evidence, with supporting tendency evidence from two other witnesses (Simon and Justin) (2,129 words).
6. Joint trial with three complainants (Simon, Justin and Timothy) who gave weak, strong and moderately strong evidence respectively (2,365 words).

Using a between-subjects experimental design, we recruited a total of 300 jury-eligible citizens as online mock jurors and randomly allocated them to one of six trial groups (weak case, moderate case, strong case, relationship evidence trial, tendency evidence trial or joint trial). Each mock juror read one trial summary. Copies of the six trial summaries are in the online materials on the Royal Commission website and a full description of the method and results of the pilot study with individual online mock jurors is in Appendix E.

A major aim of the manipulation check was to ensure that the conviction rate for the claims of the moderately strong complainant were approximately 50 per cent in a separate trial, to test with confidence the differences in verdict and other measures between the basic separate trial and other trial types. The results of the pilot study indicated that the evidence for this complainant was evenly balanced, yielding a conviction rate from individual mock jurors between 50 and 60 per cent. Since conviction rates by juries as a group are typically lower than those for individual mock jurors, conviction rates in this range were satisfactory, and established a good premise on which to assess the between-subjects effects on the dependent measures of adding more inculpatory evidence in the relationship evidence trial, the tendency evidence trial and the joint trial.

Where a complainant alleged multiple counts against the defendant, the conviction rate for that complainant was averaged. Using those averages, chi-square analysis of mock juror verdicts showed a significant relationship between the conviction rate and evidential strength.¹⁴⁴ The stronger the evidence against the defendant, the more likely jurors were to convict (weak claims: 24.0 per cent, moderately strong claims: 58.7 per cent, strong claims: 73.6 per cent). In the separate trials, the conviction rate for indecency in the form of masturbation by the defendant was 24.0 per cent for the complainant with the weak evidence, 54.3 per cent for the complainant with the moderately strong evidence, and 67.9 per cent for the complainant with the strong evidence. The conviction rate for the third count by the complainant with the strong evidence, indecency in the form of masturbation of the defendant, was 71.7 per cent. Conviction

¹⁴⁴ $\chi^2 = 26.43, p < .001, \text{Phi} = .421$.

rates for the penetrative offences were lower than for indecency: 52.2 per cent and 58.5 per cent respectively for the complainant with the moderately strong evidence and the strong evidence. Since a joint trial was expected to produce an increase, it was important to ensure that the evidence in the strongest case did not yield a ‘ceiling’ effect, allowing no room for an increase in convictions. Conviction rates in the observed ranges in the separate trials were satisfactory to test the effect of joinder.

Supplementary analyses using other measures confirmed that as the evidence strength increased, the online jurors were more likely to agree that the defendant had a sexual interest in boys and to perceive the defendant as factually culpable for the alleged offences. However, as with the verdict, other differences between the moderately strong and the strong case were not as pronounced. For instance, differences in the perceived criminal intent of the defendant and blame attributed to the complainant differed significantly between the weak and strong cases, and showed trends in the expected direction when comparing the moderately strong case with either the weak or the strong case. Given that the major question in a joint trial is the impact of joinder on the claim with the weakest evidence – and conviction rates for the weakest claim were significantly lower than those of the moderately strong and strong claims – overall, we deemed the manipulation of the evidence strength for the three complainants to be successful.

3.3 Participants in the jury study

A total of 1,029 jury-eligible citizens (hereafter referred to as mock jurors) comprising 580 women and 449 men aged between 18 and 82 years ($M = 43.5$, $SD = 15.43$) were randomly allocated to one of 90 mock juries.¹⁴⁵ Table 1 shows demographic characteristics of the participants. Most participants had completed a tertiary degree (47.1 per cent) or were currently undertaking tertiary education (13.8 per cent). Twenty-four per cent had completed a trade certificate, diploma or equivalent, and the remainder had completed high school or less (15.3 per cent). The majority of the participants were employed (62.7 per cent) and were parents (54.0 per cent).

¹⁴⁵ Appendix A compares the demographic characteristics of the mock jurors in this study with those of NSW empanelled jurors and jurors who participated in studies conducted in the past 10 years.

Table 1. Participant demographic characteristics

	Per cent	Number
Gender		
Male	43.6	449
Female	56.4	580
Age group (in years) (unknown: $n = 3$)		
18–29	25.3	260
30–39	17.1	175
40–49	16.1	165
50–59	23.0	236
60+	18.5	190
Highest education qualification (unknown: $n = 19$)		
Less than high school	1.6	16
High school	13.6	137
Some university courses	13.8	139
Trade certificate or diploma	24.1	244
Bachelor's degree	30.9	312
Post-graduate diploma or master's degree	13.0	131
Doctoral degree	3.2	32
Occupation (unknown: $n = 19$)		
Employed for wages (full time or part time)	62.7	633
Student	11.2	113
Retiree or pensioner	8.5	86
Self-employment	7.4	75
Unemployed	6.5	65
Home duties	3.8	38
Parental status (unknown: $n = 19$)		
Parent	55.0	556
Non-parent	45.0	454

3.4 Research design

The jury study was designed to test the impact of evidence strength, the number of charges and the presence of specific judicial directions on jury decision making in joint versus separate trials.

The evidence strength in the trials was varied in terms of the type of alleged sexual abuse (indecenty versus penetration), the type of witness (complainant or supporting witness) and the number of Crown witnesses (one, three or five). The trials were varied in terms of the number of complainants (one versus three), the number of charges against the defendant (two versus six) and the type of judicial directions (standard, specific, and specific with question trails).

To incorporate these variables, the trial type was one of four variations:

- (a) Separate trial of an adult male complainant with moderately strong evidence (Timothy) (Trial 1)
- (b) Separate trial of an adult male complainant with moderately strong evidence (Timothy), in which relationship evidence about the defendant's uncharged sexual acts and grooming behaviours was presented (Trials 2, 3 and 4)
- (c) Separate trial of an adult male complainant with moderately strong evidence (Timothy), in which tendency evidence from two prosecution witnesses (Simon and Justin) was admitted (Trials 5 and 6)
- (d) Joint trial involving the same defendant and three adult male complainants (Simon, Justin and Timothy), who gave weak, strong and moderately strong evidence respectively (Trials 7–10).

Judicial directions involved one of five variations:

- (a) Standard judicial directions (Trials 1, 2, 5 and 10)
- (b) Standard judicial directions plus a context evidence direction informing the jury that they could use the relationship evidence for context to help them better understand the charges, but could not use the relationship evidence as evidence of the defendant's tendency to have a sexual interest in and engage in sexual activities with young boys under the age of 12 (Trial 3)
- (c) Standard judicial directions plus a context evidence direction and a question trail (Trial 4 – a copy of the question trail is attached in Appendix J)
- (d) Standard judicial directions plus a tendency evidence direction informing the jury that they could only use the evidence of other sexual acts by the defendant once they had

made findings beyond reasonable doubt that those acts had occurred (Trials 6, 7 and 9)

- (e) Standard judicial directions plus a tendency evidence direction, plus a question trail (Trial 8 – a copy of the question trail is attached in Appendix K).

Table 2 provides an overview of the research design and the 10 trials.¹⁴⁶

Table 2. Trials by counts, Crown witnesses, jury directions, number of juries and mock jurors, and jurors’ mean age and gender

Exp. Group	Counts	Crown witnesses (number and evidence type)	Jury directions	Juries (number)	Mock Jurors (number)	Mean age	Male jurors (%)
Separate trial with one complainant							
1	2	2	Standard	9	105	41.5	55.2
2	2	2 RE	Standard	12	135	40.5	45.2
3	2	2 RE	Standard, RE	9	103	42.3	44.7
4	2	2 RE	Standard, RE, QT	10	107	44.2	38.3
5	2	4 TE	Standard	8	85	42.2	45.9
6	2	4 TE	Standard, TE	9	112	48.1	41.4
Joint trial with three complainants							
7	6	6 TE	Standard, TE	8	93	44.8	35.5
8	6	6 TE	Standard, TE, QT	8	100	43.8	46.0
9	6	4 TE	Standard, TE	9	108	44.9	39.4
10	6	6 TE	Standard	8	83	42.8	44.6

Note: QT = question trail; RE = relationship evidence; TE = tendency evidence.

3.4.1 Trial simulation materials

In the joint trial, the defendant was charged with six counts of child sexual assault against three unrelated complainants with whom the defendant had a relationship of care and control in an institutional setting as a soccer coach. Three counts of sexual assault involved the complainant with strong evidence (Justin), two counts involved the complainant with moderately strong

¹⁴⁶ A power analysis indicated that 10 people per jury group and nine juries per experimental trial condition were sufficient to conduct multilevel analyses, and to conduct targeted comparisons between paired subsets of the trials.

evidence (Timothy) and one count involved the complainant with weak evidence (Simon). In the separate trial, the defendant was charged with two counts of child sexual assault involving the complainant with moderately strong evidence (Timothy).

A series of scripts were written; the joint trial formed the basis for all other trials. The facts in the joint trial were drafted according to the type of allegations that would be prosecuted in NSW as a joint trial, with a mixture of similar and dissimilar sexual behaviours by the defendant, and similar and dissimilar circumstances in which the alleged sexual assaults occurred.¹⁴⁷ Copies of the full scripts for all trials can be found in the online materials on the Royal Commission website. A summary of the joint trial is at the end of this chapter.

All trials included the opening address of the trial judge; opening addresses by the prosecution and defence; evidence-in-chief of each complainant and, where relevant, supporting witnesses; cross-examination of each complainant and, where relevant, supporting witnesses; evidence-in-chief of the defendant; cross-examination of the defendant; closing addresses by the prosecution and the defence; and the judge's summing-up. The joint and three separate trials were videotaped to increase the ecological validity of the study materials, producing simulated trials ranging from 45 to 110 minutes in length. Professional actors were employed to play the roles of all complainants and witnesses, while actual barristers and a New South Wales District Court trial judge performed the roles of the trial judge, defence counsel and prosecutor. All trials included relevant standard judicial directions and warnings from the Criminal Trial Courts Bench Book and the Sexual Assault Trials Handbook (standard judicial directions).¹⁴⁸

To test the influence of the judicial directions on jury reasoning, the content of the judge's summing-up was varied in the joint trial and in the separate trial with relationship evidence. Different versions of the separate trial with relationship evidence included one without a context evidence direction (Trial 2) and one with it (Trial 3).

Different versions of the joint trial included one with a tendency evidence direction (Trial 7) and one without (Trial 10).

To test the impact of question trails on jury reasoning, some of the juries exposed to relationship evidence and some of the juries exposed to joint trials received a question trail after the judge's summing-up. The question trails were drafted with the assistance of a New South Wales District Court trial judge based on similar question trails given to juries in cases before her.

¹⁴⁷ All trials were checked for accuracy by a NSW District Court judge and NSW criminal law practitioner.

¹⁴⁸ *Bench Books*, above n 19.

To test the impact of the number of witnesses on jury decisions, the joint trial transcript was varied to create a version without the evidence of the two supporting witnesses for the complainant with the strong evidence (Trial 9). Aside from these two witnesses, all other features of this trial were the same as those in Trial 7.

3.4.2 Dependent measures

Mock jurors completed an online pre-trial written questionnaire when they registered to attend the experiment (Appendix G). They completed an onsite post-trial written questionnaire at the conclusion of their jury deliberations (Appendix L).

Pre-trial questionnaire

The pre-trial questionnaire assessed mock juror attitudes and biases via:

- (a) the Pre-trial Juror Attitude Questionnaire¹⁴⁹ (PJAQ), which looked at six factors (*Conviction Proneness, System Confidence, Cynicism to the Defence, Social Justice, Racial Bias* and *Innate Criminality*)¹⁵⁰
- (b) the Forensic Evidence Evaluation Bias Scale¹⁵¹ (FEEBS), which looked at two factors (*Pro-Prosecution Bias* and *Pro-Defence Bias*)
- (c) a nine-item *Child Sexual Abuse Knowledge* Questionnaire¹⁵² (CSA-KQ), which looked at two factors (*Impact of Sexual Abuse on Children* and *Contextual Influences on Report*).

A copy of the pre-trial juror questionnaire is attached in Appendix G.

Post-trial questionnaire

We assessed differences in jury reasoning processes and verdicts by comparing jury and juror ratings of factual accuracy; the perceived criminal intent of the defendant; the credibility and convincingness of the complainant(s); the blameworthiness of the complainant(s); distinctions

¹⁴⁹ Len Lecci and Bryan Myers, 'Individual Differences in Attitudes Relevant to Juror Decision Making: Development and Validation of the Pre-Trial Juror Attitude Questionnaire (PJAQ)' (2008) 38(8) *Journal of Applied Social Psychology* 2010.

¹⁵⁰ Factors of interest in the present study were *Conviction Proneness* and *Innate Criminality*.

¹⁵¹ Lisa L Smith and Ray Bull, 'Identifying and Measuring Juror Pre-Trial Bias for Forensic Evidence: Development and Validation of the Forensic Evidence Evaluation Bias Scale' (2012) 18(9) *Psychology, Crime, and Law* 797.

¹⁵² Goodman-Delahunty, Cossins and Martschuk, above n 5.

between charges of different severity; the defendant's sexual interest in boys; the factual culpability of the defendant; the verdict; and the consistency of verdicts across charges.

The Observed Witness Efficacy Scale¹⁵³ (OWES) – via two factors, *Poise* and *Communication Style* – addressed perceptions of the credibility of the complainant with moderately strong evidence (Timothy), whose evidence was common to all 10 trials.

We also gathered jury perceptions of the fairness of the trial to the complainant(s) and the defendant; the perceived fairness of the trial; expectations of related evidence; and self-reported cognitive effort to understand the facts and the law, and to reach a verdict. Finally, jurors completed the Positive and Negative Affect Scale¹⁵⁴ (PANAS) to assess their current *Positive Affect* and *Negative Affect* (attentive or distressed, for example). A copy of the post-trial juror questionnaire is included in Appendix L.

Verdict form

Copies of the verdict forms for separate and joint trials are attached as Appendix H and I respectively.

3.5 Research procedures

Charles Sturt University Human Research Ethics Committee granted approval to conduct this research (Appendix B).

Mock jurors were recruited by a market research company and by placing notices about the study in the Jury Assembly Room at the Downing Centre in Sydney, which serves the NSW District and Supreme Courts.¹⁵⁵ Participants were offered a \$100 payment for their time. When mock jurors registered online for the study, they completed the necessary consent forms – including permission to audiotape and videotape their deliberations – and completed the pre-trial questionnaire.

Six to 12 simulated trials were conducted over eight consecutive days, at venues in the Sydney CBD and the Parramatta Justice Precinct. The order of the trials was varied to control any order effects based on the time of day or day of the week. Upon arriving at the trial venue, participants

¹⁵³ Cramer, DeCoster, Neal and Brodsky, above n 18.

¹⁵⁴ John R Crawford and Julie D Henry, 'The Positive and Negative Affect Schedule (PANAS): Construct Validity, Measurement Properties and Normative Data in a Large Non-Clinical Sample' (2004) 43(3) *British Journal of Clinical Psychology* 245.

¹⁵⁵ The NSW Department of Justice granted approval to place invitations in the Downing Centre jury assembly room.

were randomly assigned to one of the trials and attended a video-recorded trial lasting between 45 and 110 minutes. Each mock juror received a written copy of the judicial directions was handed to each mock juror when the judge commenced her oral summing-up.

Mock juries had approximately 90 minutes to reach a unanimous decision on all counts. Their deliberations were audiotaped and videotaped. A randomly assigned mock juror in each jury was handed the verdict form, then asked to record the group decision for each count and return it to the researchers. After all verdicts were recorded, each mock juror completed the post-trial questionnaire. All study participants received a written debriefing statement before being released.

3.6 Data analysis

Because it is difficult to ascertain precise measures of group reasoning, we assessed both individual juror decisions and jury group decisions and deliberations, using multiple methods and convergent measures of inferences drawn from the evidence, to compile a composite picture of factors that influenced jury reasoning and decision making. The two major sources of data were the information gathered from individual mock jurors on their written pre-trial and post-trial questionnaires, and the group deliberations of the 90 juries.

We quantitatively analysed individual mock jurors' responses to the pre-trial and post-trial questionnaires using SPSS and MPlus. Details of the statistical tests performed are presented in footnotes that include the specific *p* values indicating the significance level, and also the effect sizes, to allow assessment of the magnitude of the observed effects.

All differences reported in the results section are statistically significant. The sole exception is the reported values of the verdicts, which are descriptive unless noted otherwise. The effect size reported for verdict – a binary dependent variable – is the Odds Ratio, reported only for analyses conducted using individual mock jurors' verdict responses, not those of jury groups.

Where multi-level analyses were conducted to take into account the allocation of mock jurors to groups and the resulting non-independence of their decisions, results are presented first at the level of the juror, based on individual mock jurors' responses, and second at the level of the jury, using the jury groups to report effects within juries. The intra-class correlation (ICC), which can range between .0 and 1.0, is the degree of dependence of the observation, defined as

the ratio of between-group variance to the total variance.¹⁵⁶ In this report, the ICC values provide a measure of the variance on any particular measure attributable to a jury group, such as the effects of jury deliberation on the responses of the juries within one or more particular trials or conditions. The higher the ICC value, the more the differences in the observed values of a particular outcome variable can be attributed to the unique jury groups. Even a small ICC value such as .05 or .10 indicates that the responses are affected by the group process within juries, and that they differ from responses that would be obtained from individual mock jurors in the absence of those influences. In social science research, the average ICC falls between .05 and .25.

Where regression analyses were conducted, the regression coefficient β reflects the strength of the impact of an independent or predictor variable on an outcome or dependent variable.¹⁵⁷ The statistical significance test for regression analyses is the Z-test. A Z value of 1.96 is equivalent to $p = .05$.

The recorded jury deliberations were transcribed for quantitative and qualitative analyses. Trained research assistants who were blind to the experimental group conducted quantitative coding of deliberations and assessed (a) mock juries' understanding of the evidence; (b) juries' understanding of the judicial directions; and in particular (c) the presence of factual errors, unfair prejudice against the defendant, and any verdicts motivated by inter-case conflation of the evidence, accumulation prejudice, and/or character prejudice. The coding scheme used for this purpose is in Appendix M.

Two research assistants independently coded 10 per cent of the deliberation transcripts to ensure a high degree of inter-coder agreement and establish inter-rater reliability. Similarly, inter-rater reliability was calculated for the coding conducted during deliberations. Disparities were resolved by simplifying the coding scheme to achieve consensus. The interrater reliability was assessed using intra-class correlation for continuous coding and was .98 for the live coding and .99 for the transcript coding scheme.¹⁵⁸

¹⁵⁶ Barbara M Byrne, *Structural Equation Modeling with Mplus* (Routledge Taylor & Francis Group, 1st ed, 2012); Christian Geiser, *Data Analysis with Mplus* (Guilford Press, 2013).

¹⁵⁷ Ibid.

¹⁵⁸ Klaus Krippendorff, 'Reliability in Content Analysis: Some Common Misconceptions and Recommendations' (2004) 30(3) *Human Communication Research* 411.

Results of the foregoing quantitative analyses were illustrated by excerpts from the jury deliberations, and by three case studies comparing the decision making of different groups of juries.

We undertook qualitative analyses of (a) short post-trial written responses from individual jurors to an open-ended question about the major reason for their verdict; and (b) transcripts of group deliberations in the 33 joint trials.

We used NVivo to analyse individual jurors' stated reasons for their verdicts, to identify the main themes in their reasoning, and to assess the prevalence of any unfairly prejudicial reasoning in the form of accumulation prejudice, and character prejudice. We thematically analysed group deliberations using the sensemaking approach described in Chapter 2, to track the collective decision process, and to discern whether the jury decision making relied on permissible or impermissible forms of reasoning.

Chapter 4 presents the findings of the jury reasoning and deliberation study. The results of the study are presented in eight sections:

Part 1: The influence of mock jurors' pre-trial expectations and attitudes

Part 2: The influence of the trial type on jury reasoning and verdicts

Part 3: Jury reasoning by type of trial

Part 4: The influence on jury verdicts of the number of counts and number of witnesses

Part 5: The influence of jury directions on jury reasoning and decisions

Part 6: The influence of question trails on jury reasoning in separate and joint trials

Part 7: Self-reported cognitive effort by type of trial

Part 8: Fairness of the trial.

Chapter 4 presents the quantitative and qualitative results in conjunction with the issues that they addressed. Excerpts from jury deliberations are provided throughout Chapter 4.

R v Booth: Summary of Joint Trial

The accused Mr Mark Booth is charged with six counts of sexual assault against **three** complainants:

- 1: one** count of an act of indecency against Simon Rutter
- 2 and 3: two** counts of acts of indecency against Justin McCutcheon
- 4: one** count of sexual intercourse against Justin McCutcheon
- 5: one** count of an act of indecency against Timothy Lyons
- 6: one** count of sexual intercourse against Timothy Lyons

Mr Booth has pleaded not guilty to all charges.

In court, Simon Rutter testified that:

He was a member of the under-12 boys Kogarah soccer team in July 1993, when Mr Booth was the coach. When Simon joined up and tried on his uniform, the accused grabbed his crotch as he complained his shorts were too tight. The accused offered to look after Simon on Saturdays until Simon's mother finished work. Simon visited Mr Booth's house a number of times. In August 1993, Simon was pushed into the accused's swimming pool after a thunderstorm. Afterwards, Mr Booth told Simon to take a shower. As he helped Simon dry himself, he reached over Simon's shoulder and stroked Simon's penis while pressing himself against Simon. Simon froze but when the accused touched his bottom, he ran out of Mr Booth's house to his mother's waiting car and never returned to the accused's house.

On cross-examination, Simon Rutter:

- Stated that although Simon told the police these events happened in March 1993, in court he said they happened in August 1993.
- Stated he was unaware that no rain fell in August 1993 and that neither Mr Booth nor any of his neighbours owned a swimming pool in 1993.
- Denied fabricating the claims to get victim's compensation.

Justin McCutcheon testified in court that:

He was a member of the under-12 boys Kogarah soccer team in 1995 when Mr Booth was the coach. Mr Booth became a close friend of Justin's father and was often invited to the McCutcheon household for drinks and dinner. Because Justin's father travelled a lot for work,

Mr Booth became a surrogate father to Justin. He took him to swimming lessons, the cricket and the movies.

In September or October 1995, Justin stayed at Mr Booth's house after going to a movie with him. Because there was no spare bed, Justin was told he had to share Mr Booth's bed. As he was getting undressed, Mr Booth told him to take off all his clothes. Mr Booth tickled Justin and played with his genitals. He took Justin's hand and placed it on his own penis, forcing him to rub Mr Booth's penis. Mr Booth masturbated Justin's penis and his own for at least half an hour, maybe longer. About a week later when travelling home from a swimming lesson in Mr Booth's car, Mr Booth stopped in a park, pulled Justin's head into his crotch, and forced his penis into Justin's mouth.

On cross-examination, Justin McCutcheon:

- Stated he did not escape from Mr Booth's bed because he froze out of fright and did not know what to do.
- Stated he didn't tell either of his parents because Mr Booth was his father's best friend, but he did tell his best friend at school.
- Stated that he refused to go on outings with Mr Booth anymore which caused arguments with his parents.
- Stated that he was suspended from school for forging his parents' signatures on letters sent to the school. That was after he had been abused when his behaviour changed.
- Denies making up his allegations because he was jealous of his father's friendship with Mr Booth.

Timothy Lyons testified in court that:

He was a member of the under-12 boys Kogarah soccer team in May 1997 when Mr Booth was the coach. He was one of a group of Mr Booth's favourite boys. Every Saturday, Mr Booth took them to McDonald's after training. Like one of the boys, Mr Booth mucked around and joked, and was great fun to be with. He was also very touchy-feely and used to ruffle the boys' hair or put an arm around their shoulders.

Timothy loved the attention because his dad didn't live with him and his mother. Mr Booth used to drive Timothy home after McDonald's and would sometimes stay for a cup of tea with his mother. In August 1997, on the back porch of the Lyons' house, Mr Booth confided that he thought Timothy's mother was sexy and pointed to his crotch, saying that that will

happen to him when he started looking at girls. When Timothy said it already did, Mr Booth asked him what he had ‘down there’.

When Mrs Lyons had an operation in hospital in December 1997, Timothy stayed at Mr Booth’s house. They ate pizza and watched a video on a TV on top of a chest of drawers in Mr Booth’s bedroom. At first, Timothy thought the video was *Hercules* then said it was the *Babe* film, *Pig in the City*.

He had to share Mr Booth’s bed. While in bed, Mr Booth stroked Timothy’s arms and leg to stop him worrying about his mother. He then touched Timothy’s penis on the outside of his underpants. When Mr Booth placed Timothy’s hand on his own penis, Timothy pulled his hand away. Mr Booth then took off Timothy’s underpants and rubbed Timothy’s penis with an oily substance. Mr Booth rolled Timothy onto his stomach and inserted his finger into Timothy’s anus until Timothy cried out and told him to stop. Timothy ran away to his own house the next day and refused to let Mr Booth in.

On cross-examination, Timothy Lyons:

- Denied that he made up his allegations because he was jealous of his mother’s renewed friendship with Mr Booth. He also denied that he knew about the current relationship between them because he had not spoken with his mother for a couple of years.
- Stated he was unaware that *Pig in the City* was not released until December 1998.
- Disagreed that Mr Booth did not have a TV or chest of drawers in his bedroom in 1997 because Mr Booth had two houses. In December 1997, they had gone to his second house. He didn’t realise the information about the second house was important and that he ought to have mentioned it before.

Ellen Samuels testified in court that:

Mr Booth did the gardening at her townhouse and collected her mail when she was overseas. She gave Mr Booth a set of keys for her house when she was away. In December 1997, when she travelled overseas for her 50th birthday, Mr Booth looked after her townhouse. There was a TV in her bedroom on top of a chest of drawers.

Mark Booth testified that:

He was the coach of the Kogarah under 12 boys soccer team from June 1993 to December 1997.

When new boys tried on uniforms, the boys' parents were always present. He befriended Simon Rutter because Simon was always alone after soccer training, waiting for his mother. He felt sorry for him, and offered him lunch which became a regular thing. It assisted Simon's mother, who never knew what time she would be finishing work. He did not push Simon into a swimming pool in 1993. He did not have a swimming pool in 1993, nor did any of the neighbours in his street. Simon did not shower at his house and he did not masturbate Simon's penis. As a boy, Simon was a bit of handful. He didn't like being told what to do, and was sometimes a bit loose with the truth. Simon, who did not have a father, had to be brought into line sometimes, something he did not like.

He knew Justin's father and he and Mr McCutcheon became good mates. He was like an uncle to Justin, and helped out the McCutcheon family because Mrs McCutcheon needed assistance with her three children while her husband was away. When Justin stayed overnight at his house, he slept on the couch. He did not share his bed with Justin, touch him, or masturbate Justin's penis, or try to get Justin to touch his penis. He did not force his penis into Justin's mouth. The McCutcheons were his friends. They invited him into their home, and he did not betray their trust. He did the family a favour by taking Justin on outings when his father was away.

He took a group of boys to McDonald's after soccer training in 1997 as a reward for playing well. Timothy Lyons was one of these boys. When he had tea on the Lyons' back porch, he and Timothy were in full view of Mrs Lyons while she was in the kitchen. He did not make suggestive comments about Mrs Lyons or ask Timothy what he had 'down there'. Mr Booth helped Mrs Lyons when she was in hospital because someone had to look after Timothy.

He did not watch a video on a TV in his bedroom with Timothy and he did not take Timothy to any other residence. He did not have a key to Mrs Samuels' house. He did not share his bed with Timothy nor did he rub Timothy's penis or insert his finger into Timothy's anus. Mr Booth took Timothy to the hospital to see his mother because Timothy was worried about her.

On cross-examination, Mr Booth:

Denied grooming Mrs Rutter to gain access to her son, or grooming Simon so he could sexually abuse him. He denied grooming the McCutcheon family to gain access to Justin so

he could sexually abuse him. He also denied grooming Mrs Lyons to gain access to her son or grooming Timothy to make him feel special.

He stated that contact with Mrs Rutter and Simon ceased because he thinks Simon had lied to his mother. He did not phone Mrs Rutter to find out why there was not further contact. With 11 boys on the team, he could not have been expected to follow up on each one. He had tried to help Mrs Rutter but it didn't work out, so he moved on.

Justin stayed on the soccer team until he was 11 or 12 years old. Contact with the McCutcheon family only ended after Justin left the team. It was normal to lose contact with families after their son left the team.

Timothy did not run away from his house. He decided to remain at the hospital with his mother. He did not contact Mrs Lyons after she came out of hospital and couldn't remember why.

Chapter 4: Results

4.1 The influence of mock jurors' pre-trial expectations and attitudes

This section tested the extent to which individual mock jurors' pre-trial expectations and attitudes differed. The aim is to assess how their individual differences contributed to jury reasoning and decision making, and to ensure that observed differences in responses to the trials were caused by changes in the trial information and not by pre-existing differences in the jurors assigned to any particular trial group. The results showed that the more mock jurors knew about child sexual abuse, the less likely they were to endorse other types of pre-trial bias. Accurate recall of the case facts was higher among mock jurors who had more accurate knowledge of child sexual abuse and lower among those with high expectations that forensic evidence would be presented at trial. Mock jurors with higher educational achievement were less likely to expect forensic evidence at trial. Mock jurors who were more knowledgeable about factors that influence reports of child sexual abuse and who favoured the prosecution rated the complainant as more credible.

4.1.1 Research aim

Preliminary analyses of individual mock jurors' expectations and attitudes aimed to explore how jurors' knowledge of child sexual abuse, their expectations of forensic evidence and their general attitudes towards the criminal justice system contributed to jury reasoning and decision making. In addition, these analyses were conducted to ensure that observed differences in responses to the trials were a result of changes in the trial information and not due to pre-existing differences in the knowledge, attitudes and expectations of mock jurors assigned to any particular trial group.

4.1.2 Mock jurors' individual pre-trial biases and post-deliberation responses

To explore the influence of the expectations and predispositions mock jurors bring to the trial, we conducted a series of correlations for measures of the mock jurors' pre-trial attitudes and post-trial responses. The results are set out in Table 3.

Results of analyses of mock jurors' pre-trial and post-trial attitudes yielded a negative correlation between mock jurors' *Child Sexual Abuse Knowledge* (that is, *Impact of SA on*

Children and Contextual Influences on Report), and their pre-trial expectations about forensic evidence (FEEBS *Pro-Prosecution Bias*: a specific bias related to evidence favouring the prosecution, and conclusions in cases where forensic evidence is lacking). Similarly, mock jurors' *Child Sexual Abuse Knowledge* was negatively correlated with juror pre-trial attitudes measured with the PJAQ (*Confidence in the Justice System, Conviction Proneness, Cynicism about the Defence, Social Justice, Racial Bias and Innate Criminality of Defendants*). The more mock jurors knew about child sexual abuse, the less likely they were to endorse other types of pre-trial bias.

Post-deliberation responses revealed a positive correlation between mock juror *Child Sexual Abuse Knowledge* about the *Impact of Child Sexual Abuse*, and memory for the case facts. Mock jurors with more accurate *Child Sexual Abuse Knowledge* were more likely to recall the case facts accurately. Conversely, mock jurors' pre-trial *Pro-Prosecution Biases* about forensic evidence were negatively correlated with accurate factual recall. Mock jurors with strong expectations that forensic evidence would have been presented by the prosecution were less accurate in recalling the case facts.

Mock jurors' *Child Sexual Abuse Knowledge* about *Contextual Influences on Report* was positively correlated with the perceived credibility of the complainant. Similarly, their *Pro-Prosecution Bias* was positively correlated with perceived complainant credibility. Mock jurors with greater *Child Sexual Abuse Knowledge* about *Contextual Influences on Report* and mock jurors with greater *Pro-Prosecution Bias* were more likely to perceive the complainant with the moderately strong evidence as credible.

Mock juror education was positively correlated with some pre-trial dispositions. A small but significant effect emerged, showing that mock jurors with higher educational achievement were less likely to expect forensic evidence at trial. Furthermore, on the PJAQ, mock jurors were less likely to report *System Confidence*, to show *Conviction Proneness* or to endorse the measures of *Racial Bias*. Mock jurors with higher formal educational achievement were also more likely to report *Positive Affect* at the conclusion of the trial simulation. The more mock jurors knew about the *Impact of Child Sexual Abuse* the less likely they were to report *Positive Affect* at the conclusion of the simulated trial.

Where individual juror attitudes and expectations exerted an influence on the jury reasoning or decisions under consideration, the results are reported in that context throughout Chapter 4

Table 3. Inter-correlations of mock jurors' individual pre-trial biases and post-deliberation responses for the entire study sample

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. CSA-KQ <i>Impact of Sexual Abuse on Children</i>	-												
2. CSA-KQ <i>Contextual Infl.</i>	.393**	-											
3. CSA-KQ <i>sum</i>	.872**	.792**	-										
4. FEEBS <i>Pro-Prosecution</i>	-.452**	-.181**	-.396**	-									
5. FEEBS <i>Pro-Defence</i>	-.445**	-.269**	-.438**	.582**	-								
6. PJAQ <i>System Confidence</i>	-.406**	-.233**	-.393**	.489**	.290**	-							
7. PJAQ <i>Conviction Proneness</i>	-.425**	-.246**	-.413**	.467**	.345**	.621**	-						
8. PJAQ <i>Cynicism in Defence</i>	-.364**	-.338**	-.421**	.358**	.354**	.516**	.526**	-					
9. PJAQ <i>Racial Bias</i>	-.386**	-.242**	-.385**	.426**	.338**	.509**	.470**	.425**	-				
10. PJAQ <i>Social Justice</i>	-.161**	-.233**	-.231**	.218**	.233**	.268**	.251**	.457**	.241**	-			
11. PJAQ <i>Innate Criminality</i>	-.395**	-.287**	-.414**	.474**	.430**	.642**	.550**	.560**	.545**	.371**	-		
12. Factual memory	.102**	.060	.099**	-.119**	-.078*	-.061	-.082**	-.052	-.103**	.010	-.063*	-	
13. Cred M <i>Poise</i>	-.055	.063*	-.003	.089**	.018	.059	.042	.025	.033	.025	.081**	-.088**	-
14. Cred M <i>Communic. style</i>	-.035	.057	.007	.071*	.001	.035	.026	.018	.011	.003	.050	-.061*	.924**
15. Cred M <i>overall</i>	-.042	.067*	.008	.079*	.011	.047	.028	.021	.024	.023	.069*	-.075*	.988**
16. Positive affect	-.131**	-.029	-.102**	.075*	.027	.102**	.062	.038	.081**	-.019	.052	.030	.064*
17. Negative affect	-.195**	-.129**	-.198**	.181**	.240**	.108**	.126**	.146**	.158**	.078*	.186**	-.103**	.001
18. Cognitive effort overall	-.015	.013	-.003	.021	-.004	.050	.046	.000	-.010	.006	.004	-.045	.155**
19. Criminal intent	-.017	.117**	.051	.046	.009	.040	.042	.053	-.024	.059	.032	-.103**	.365**
20. Trial fairness to defendant	.065*	.050	.070*	-.058	-.087**	.013	-.024	-.026	-.049	-.048	-.026	-.066*	.093**
21. Convincingness of M	.056	.125**	.104**	.042	-.032	.008	-.042	-.025	-.079*	.052	-.010	-.048	.389**
22. Complainant blame of M	-.197**	-.237**	-.256**	.101**	.199**	.086**	.089**	.098**	.199**	.022	.151**	-.064*	-.194**
23. Factual culp M nonpen	.058	.157**	.122**	.002	-.076*	.016	-.004	.009	-.077*	.039	.021	-.090**	.381**
24. Factual culp M pen	.063*	.153**	.124**	.013	-.068*	.027	.013	.016	-.077*	.043	.027	-.085**	.352**
25. Trial fairness to M	-.021	.001	-.014	.028	.027	.037	-.028	.016	.043	-.033	-.004	-.129**	.170**
26. Convincingness of W	.028	.149**	.099*	.052	.000	-.012	-.050	-.091*	-.039	-.034	.006	-.070	.181**
27. Complainant blame of W	-.216**	-.254**	-.280**	.053	.175**	.072	.112**	.124**	.216**	.033	.159**	-.123**	-.155**
28. Factual culp W nonpen	.009	.162**	.092*	.053	-.002	.024	-.034	-.027	-.099*	-.001	.022	.023	.244**
29. Trial fairness to W	-.086*	.009	-.052	.018	.020	.090*	.009	-.003	.052	-.069	.017	-.034	.097*
30. Convincingness of S	.090*	.197**	.165**	.072	-.032	-.027	-.063	-.082	-.137**	-.050	-.034	.080	.229**
31. Complainant blame of S	-.197**	-.250**	-.265**	.062	.159**	.072	.090*	.094*	.169**	.026	.139**	-.142**	-.169**
32. Factual culp S nonpen 1	.067	.214**	.159**	.040	-.083	.059	-.004	-.031	-.112**	.000	.022	.095*	.301**
33. Factual culp S nonpen 2	.053	.197**	.141**	.040	-.087*	.073	-.008	-.034	-.116**	-.001	.026	.102*	.305**
34. Factual culp S pen	.083*	.217**	.172**	.035	-.090*	.060	.006	-.016	-.076	.004	.031	.087*	.275**

	14	15	16	17	18	19	20	21	22	23	24	25	26
35. Trial fairness to S													
	.959**	-											
15 Cred M overall													
16. Positive affect	.067*	.061	-										
17. Negative affect	-.012	-.003	.112**	-									
18. Cognitive effort overall	.156**	.163**	.088**	.034	-								
19. Criminal intent	.351**	.361**	-.022	.079*	.152**	-							
20. Trial fairness to defendant	.096**	.092**	.043	-.093**	.074*	.186**	-						
21. Convincingness of M	.420**	.405**	.019	-.033	.151**	.516**	.147**	-					
22. Complainant blame of M	-.212**	-.201**	.118**	.106**	.101**	-.399**	-.168**	-.299**	-				
23. Factual culp M nonpen	.370**	.384**	-.025	.090**	.205**	.700**	.195**	.555**	-.424**	-			
24. Factual culp M pen	.350**	.357**	-.013	.066*	.188**	.669**	.173**	.537**	-.436**	.901**	-		
25. Trial fairness to M	.161**	.170**	.087**	-.067*	.122**	.238**	.475**	.195**	-.102**	.229**	.202**	-	
26. Convincingness of W	.186**	.178**	.016	.004	.127**	.349**	.170**	.476**	-.155**	.345**	.354**	.161**	-
27. Complainant blame of W	-.184**	-.165**	.058	.144**	-.101*	-.254**	-.180**	-.159**	.739**	-.263**	-.305**	-.078	-.290**
28. Factual culp W nonpen	.245*	.247**	-.007	.056	.162**	.423**	.154**	.344**	-.280**	.675**	.612**	.149**	.499**
29. Trial fairness to W	.102*	.097*	.099*	-.074	.046	.152**	.401**	.093*	-0.053	.135**	.115**	.742**	.158**
30. Convincingness of S	.266**	.249**	.061	-.091*	.134**	.430**	.121**	.718**	-.227**	.490**	.452**	.203**	.473**
31. Complainant blame of S	-.201**	-.177**	.096*	.154**	-.060	-.321**	-.170**	-.218**	.801**	-.281**	-.309**	-0.059	-.132**
32. Factual culp S nonpen 1	.322**	.315**	.039	.010	.194**	.583**	.187**	.506**	-.365**	.915**	.833**	.245**	.346**
33. Factual culp S nonpen 2	.328**	.321**	.027	.006	.208**	.583**	.172**	.518**	-.378**	.926**	.846**	.236**	.347**
34. Factual culp S pen	.302**	.292**	-.005	-.005	.206**	.533**	.185**	.501**	-.357**	.831**	.876**	.214**	.366**
35. Trial fairness to S	.175**	.175**	.090*	-.057	.081	.211**	.443**	.202**	-.102*	.245**	.218**	.951**	.168**

	27	28	29	30	31	32	33	34
28. Factual culp W nonpen	-.384**	-						
29. Trial fairness to W	-.025	.122**	-					
30. Convincingness of S	-.157**	.299**	.111**	-				
31. Complainant blame of S	.745**	-.234**	-.019	-.199**	-			
32. Factual culp S nonpen 1	-.284**	.710**	.150**	.473**	-.304**	-		
33. Factual culp S nonpen 2	-.297**	.684**	.135**	.479**	-.307**	.953**	-	
34. Factual culp S pen	-.289**	.606**	.114**	.474**	-.300**	.852**	.856**	-
35. Trial fairness to S	-.052	.149**	.770**	.215**	-.056	.247**	.238**	.222**

Note. 1–23: $N = 1,029$; 24–33: $n = 575$; * $p < .05$; ** $p < .01$; communic = communication; cred = credibility; culp = culpability; infl = influences; M = complainant with moderately strong case; nonpen = non-penetrative offence; pen = penetrative offence; S = complainant with strong case; W = complainant with weak case.

4.2 The influence of the trial type on jury reasoning and verdicts

This section compared jury reasoning and decisions across different trial types to assess whether verdicts were motivated by permissible or impermissible reasoning and determine whether there was a ‘joinder effect’. Conviction rates for both the non-penetrative and penetrative offences against the focal complainant in the relationship evidence trial and the tendency evidence trial were significantly higher than those in the basic separate trial. There were no significant differences between conviction rates in the tendency evidence trial compared to the joint trial. Thus, no joinder effect was found.

Rather, the perceived culpability of the defendant and the credibility of the focal complainant increased in response to independent sources of evidence, not more evidence or the type of trial. As more independent sources of evidence were introduced to support the focal complainant’s account, his credibility increased and his evidence was accorded more weight. In line with this finding, mock jurors were more likely to blame the complainant in the basic separate trial than in any other type of trial. Similarly, ratings of the defendant’s sexual interest in boys, criminal intent and factual culpability were lowest in the separate trial and highest in the joint trial – that is, as more inculpatory evidence against the defendant was admitted – and were similar in the tendency evidence and joint trials. Jury deliberations significantly increased ratings of the defendant’s criminal intent and factual culpability in trials with relationship and tendency evidence. The perceived criminal intent of the defendant predicted the verdict at the juror and jury level, irrespective of the type of offence. In the absence of tendency evidence, juries were reluctant to convict for penetrative offences. The presence of tendency evidence increased convictions for non-penetrative and penetrative offences in both the separate and joint trials. Jury distinctions between penetrative and non-penetrative offences showed that juries reasoned about the counts separately, even for the same complainant.

Analyses of mock jurors’ recall accuracy for six multiple-choice questions about the case of the focal complainant showed that trial complexity, not trial type, predicted factual recall accuracy. Recall accuracy on these questions was greatest in the trials where only two witnesses appeared for the prosecution, and decreased as more witnesses appeared for the prosecution, in both the tendency evidence and joint trials. Mock jurors’ formal education had no effect on their factual recall accuracy. Juries were more prone to convict when more

individual jurors made errors in response to these questions, but this finding was unrelated to the type of trial, and convictions in joint trials were not predicted by these jurors' factual errors.

4.2.1 Research aim

In this section, we examined the influence of the type of trial on jury reasoning and jury decisions. In particular, we examined whether there was a 'joinder effect'; if so, the extent to which jurors and juries engaged in permissible versus impermissible reasoning; and if they did, whether this resulted in unfair prejudice to the defendant.

4.2.2 Verdict by trial type

The 10 experimental trial groups encompassed the four main types of trials compared in this project. Trial 1 was the separate trial; Trials 2, 3 and 4 presented relationship evidence in a separate trial; Trials 5 and 6 presented tendency evidence in a separate trial; and Trials 7, 8, 9 and 10 presented tendency evidence in a joint trial with three complainants. All 10 trials included two counts against the defendant, brought by the focal complainant with moderately strong evidence. One count was a charge of indecency (a non-penetrative offence) and the other was a charge of sexual intercourse (a penetrative offence).

To assess whether there was a joinder effect, we examined juror and jury ratings of the culpability of the defendant, assessed by means of a binary verdict ('guilty beyond reasonable doubt' versus 'not guilty') and by a rating of the factual culpability of the defendant on each count (recorded on a scale from 1 = very unlikely to 7 = very likely). This measure was based on the questions "How likely is it that Mark Booth masturbated Timothy Lyons's penis between 1 and 31 December 1997?" and "How likely is it that Mark Booth inserted his finger into Timothy Lyons's anus between 1 and 31 December 1997?"

We recorded verdicts for 90 groups of deliberating juries, and following deliberation, verdicts and the factual culpability of the defendant for all individual mock jurors. Table 4 shows the conviction rates at the jury level and the convictions and factual culpability ratings obtained from individual mock jurors, for all ten experimental trial groups.

Table 4. Conviction rates and mean factual culpability of the defendant for counts of the focal complainant with moderately strong evidence, by trial group

Exp	Jury	Jury verdict		Juror verdict		Factual culpability	
		Nonpen	Penetr	Nonpen	Penetr	Nonpen	Penetr
Trial	directions	(%)	(%)	(%)	(%)	(mean score, 1–7)	(mean score, 1–7)
1 Sep	Standard	11.1	0.0	19.0	9.5	4.39 (1.63)	4.12 (1.61)
2 RE	Standard	8.3	0.0	22.2	19.3	4.91 (1.56)	4.75 (1.58)
3 RE	Standard + RE	33.3	33.3	68.9	69.9	5.51 (1.61)	5.47 (1.62)
4 RE	Standard + RE, QT	10.0	0.0	23.8	19.2	4.73 (1.60)	4.59 (1.64)
5 TE	Standard	62.5	62.5	71.4	66.7	5.92 (1.36)	5.58 (1.55)
6 TE	Standard + TE	55.6	55.6	75.5	76.4	5.77 (1.46)	5.64 (1.63)
7 JT	Standard + TE	75.0	75.0	88.2	88.0	6.10 (1.45)	6.08 (1.40)
8 JT	Standard + TE, QT	87.5	87.5	90.9	88.9	6.06 (1.42)	6.00 (1.42)
9 JT	Standard + TE	100.0	100.0	98.1	95.3	6.34 (0.89)	6.34 (0.99)
10 JT	Standard	100.0	75.0	93.8	84.0	6.34 (0.93)	5.98 (1.34)

Note. Nonpen = masturbation of complainant; Penetr = digital-anal penetration; Sep = basic separate trial; RE = relationship evidence; TE = tendency evidence; JT = joint trial. Standard deviations are listed in parentheses below mean factual culpability ratings.

Inspection of the verdicts revealed that for both counts, juries' group conviction rates were on average lower than the conviction rates of the individual jurors who participated in those juries, with a gap of 8 to 9 per cent between the jury and juror verdicts in the basic separate trial, 13 to 36 per cent in the separate trial with relationship evidence, and 5 to 21 per cent in the separate trial with tendency evidence. In the 33 joint trials, this trend was not consistent once the jury conviction rate reached 100 per cent in two trials for three of the four counts on which those

juries voted. In two trials (Trials 7 and 8), as is typical, the individual juror verdicts exceeded those of the 16 juries by 1 to 13 per cent, while in two trials (Trials 9 and 10) the opposite effect emerged for three of the four verdicts rendered by 17 juries regarding the focal complainant – that is, not every individual juror agreed with those group verdicts, so the average individual conviction rates were lower, and the gap between juries’ and jurors’ conviction rates ranged from –2 to –6 per cent. In Trial 10, the individual jurors’ conviction rate for the penetrative offence exceeded that of the eight other juries by 9 per cent.

This disparity between jury and juror verdicts is in line with findings in previous mock jury studies that examined the effects of group deliberation and showed that the group deliberation process results in a more conservative or lower conviction rate compared to the conviction rate of individual jurors. While some researchers have referred to the gap between group and individual verdicts as a ‘leniency effect’¹⁵⁹ produced by deliberation, others have interpreted it as the consequence of group polarisation¹⁶⁰ or the ‘majority effect’¹⁶¹ in deliberation. This study is interesting because these differences persisted in individual jurors after group deliberation. This study took the individual juror verdict measures following deliberation, unlike most other studies that obtained individual juror verdicts either prior to the jury group deliberation, or from individual jurors who did not participate in any jury group deliberation.¹⁶²

Inspection of the factual culpability ratings showed that in all types of trials except one, the defendant’s culpability for the counts of the focal complainant for the non-penetrative offence exceeded that for the penetrative offence, although the magnitude of the difference between these ratings was not always statistically significant. In general, the jurors were more reticent to convict the defendant for the more serious penetrative offence than they were for the indecency charge. In a number of trials there were differences in jury verdicts for the two counts, indicating that jurors and juries distinguished between different counts relating to the same complainant.¹⁶³ This finding, which suggests that juries were not reasoning in a

¹⁵⁹ Robert J MacCoun and Norbert L Kerr, ‘Asymmetric Influence in Mock Jury Deliberations: Jurors’ Bias for Leniency’ (1988) 54 *Journal of Personality and Social Psychology* 21; Norbert L Kerr and Robert J MacCoun, ‘Is the Leniency Asymmetry Really Dead? Misinterpreting Asymmetry Effects in Criminal Jury Deliberation’ (2012) 15 *Group Processes and Intergroup Relations* 585.

¹⁶⁰ Dan Simon, *In Doubt: The Psychology of the Criminal Justice System* (Harvard University Press, 2012) 201.

¹⁶¹ Garold Stasser, Nancy Stella, Coleen Hanna and Adrienne Colella, ‘The Majority Effect in Jury Deliberations’ Number of Supporters Versus Number of Supporting Arguments’ (1984) 8 *Law and Psychology Review* 115.

¹⁶² Goodman-Delahunty, Martschuk and Cossins, above n 1.

¹⁶³ In Trial 10, however, the individual jurors’ mean factual culpability ratings were the same for both counts.

generalised or global way regarding the different counts, is the opposite of what many judges and practitioners anticipated juries would do in a joint trial.

Inspection of the factual culpability of the defendant by type of trial for the same offences showed that average ratings (out of a total of 7) were lowest in the basic separate trial (Trial 1) and highest in the joint trials (Trials 7 to 10). The jury conviction rates reflected a similar pattern. In other words, the extent to which the same evidence presented by the same complainant on the same issues was perceived as sufficient to support a conviction varied by type of trial. The factual culpability ratings were higher in the trials with tendency evidence, that is, in the tendency evidence trials and joint trials. This finding indicated that juries' perceptions of the defendant's guilt responded to the strength of the inculpatory evidence, and not to the type of trial *per se*. This was also contrary to what many judges and practitioners anticipate juries will do in a joint trial.

4.2.3 The influence of trial type on other dependent measures¹⁶⁴

This next section presents results for the four main types of trials (separate, relationship evidence, tendency evidence and joint) without taking into account the influence of specific jury instructions on relationship evidence or tendency evidence; that is, the juries included in the following analyses all received identical standard jury directions. Holding the jury directions constant in these trials allowed us to assess the effect of changes in the type of evidence separately from the effects of the specific jury directions on relationship evidence and tendency evidence.¹⁶⁵

The following analyses compared responses from a total of 37 juries and 398 mock jurors. Of these, nine juries viewed the separate trial ($n = 105$ jurors); 12 juries viewed the separate trial with relationship evidence trial ($n = 135$ jurors), eight juries viewed the separate trial with tendency evidence ($n = 85$ jurors) and eight juries viewed the joint trial ($n = 83$ jurors).

The comparisons across trials focuses on the allegations by the complainant with the moderately strong claim (Timothy), since the evidence from this complainant was identical in all four types of trials (separate, relationship evidence, tendency evidence and joint). We expected that as more evidence about similar events and similar complainants was admitted:

- accuracy of factual recall would decline

¹⁶⁴ Trials 1 versus 2 versus 5 versus 10.

¹⁶⁵ The influence of the jury directions is examined in Ch 4.5.

- inferences that the defendant had a sexual interest in boys would increase
- inferences that the defendant was motivated by criminal intent would increase
- ratings of the responsibility of the moderately strong complainant for the sexual abuse would decrease
- perceptions of the credibility of the moderately strong complainant would increase
- perceptions of the moderately strong complainant as convincing would increase
- the factual culpability of the defendant for offences against the moderately strong complainant would increase
- distinctions between the factual culpability of the defendant for non-penetrative and penetrative offences against the same complainant would decline
- the conviction rate for counts alleged by the moderately strong complainant would increase.

Next, for the focal complainant with moderately strong evidence (Timothy), we report the quantitative results derived from post-trial measures of factual accuracy, the defendant's sexual interest in boys, the defendant's criminal intent, victim blame, witness credibility, factual culpability and the verdict.

Accuracy of factual recall and verdict

In the post-deliberation questionnaire, mock jurors answered six multiple-choice questions about the case facts pertaining to the complainant with the moderately strong claim (Timothy), to test the hypothesis that mock jurors would be more likely to confuse case facts as more evidence about similar events or similar complainants was added in different types of trials.

Analyses revealed that mock jurors' formal education had no effect on their factual recall accuracy¹⁶⁶, indicating that mock jurors' educational status was unrelated to their memory for facts presented in the trials. However, the complexity of the type of trial did predict factual recall accuracy.¹⁶⁷ Recall accuracy was greatest and undifferentiated for the separate and relationship evidence trials in which only two witnesses appeared for the prosecution. When mock jurors in the tendency evidence and joint trials were exposed to the evidence from (a) two additional witnesses or complainants who had experienced events substantially similar to

¹⁶⁶ $\beta = -0.111$, $SE = 0.106$, $Z = -1.048$, $p = .295$.

¹⁶⁷ Relationship evidence trial versus separate trial: $\beta = 0.071$, $SE = 0.281$, $Z = 0.252$, $p = .801$; tendency evidence trial versus separate trial: $\beta = -0.839$, $SE = 0.224$, $Z = -3.746$, $p < .001$; joint trial versus separate trial: $\beta = -0.974$, $SE = 0.226$, $Z = -4.302$, $p < .001$.

those experienced by one complainant (in the separate trial with tendency evidence), and (b) two additional witnesses for the prosecution aside from the two additional complainants in the joint trial, on average they answered one more question on the multiple-choice test incorrectly. In other words, as the complexity of the evidence increased, as is shown in Figure 1, jurors' factual recall accuracy scores decreased by one. Furthermore, mock juror memory for case facts depended on the jury groups ($ICC = 0.153$), with 15.3 per cent of the variance in factual accuracy due to the jury groups. (The nature of the influence of jury deliberations on factual errors is presented in Part 4.3).

Notably, the potential for factual inaccuracy or confusion was not based on the type of trial *per se*, and was not the result of joinder. Rather, the similarity in the allegations by the three males led to the additional error, as the increase in the error rate was similar, irrespective of whether this information was presented in a separate trial, as tendency evidence, or in a joint trial as cross-admissible tendency evidence.

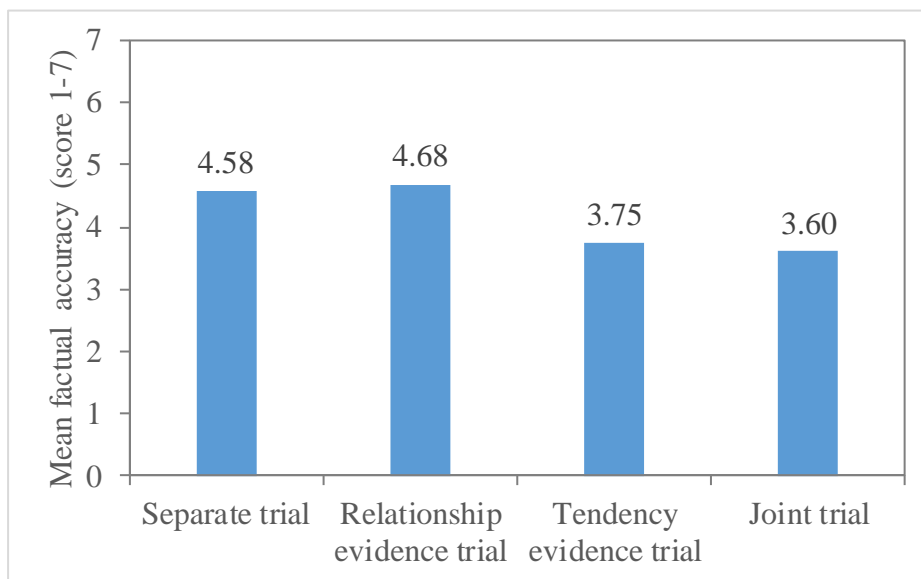


Figure 1. Mean factual accuracy, by trial type

Analyses that tested the impact of mock juror factual accuracy on verdict showed that for individual jurors, memory for case facts had no impact on verdict.¹⁶⁸ However, when jury group composition was considered, across all types of trials, the mean accuracy scores of the individual jurors in response to the six multiple-choice questions yielded a significant effect: on average, in all the trial groups, mock jurors who voted to acquit answered one more multiple-

¹⁶⁸ Non-penetrative offence: $\beta = -0.137$, $SE = 0.098$, $Z = -1.388$, $p = .165$; penetrative offence: $\beta = -0.178$, $SE = 0.093$, $Z = -1.917$, $p = .055$.

choice question correctly than did mock jurors who voted to convict (non-penetrative offence: not guilty $M = 4.65$, guilty $M = 3.78$; penetrative offence: not guilty $M = 4.60$, guilty $M = 3.78$).¹⁶⁹

On all juries, regardless of the type of trial, some jurors were better than others at keeping the facts straight. As is shown in Part 4.3, more than four-fifths of the factual errors made by individual jurors in deliberations were corrected by fellow jurors, and no uncorrected errors were instances of inter-case confusion.

Perceptions of the defendant's sexual interest in boys

Mock jurors' perceptions as to whether the accused had a sexual interest in boys was assessed by means of multi-level regression analyses conducted with jurors' *Child Sexual Abuse Knowledge* as a predictor at the juror level, and mean jury *Child Sexual Abuse Knowledge* and trial type as predictors at the jury level. The separate trial was used as the baseline group for these analyses.

Mock jurors' *Child Sexual Abuse Knowledge* had no effect on the defendant's perceived sexual interest in boys at the juror or jury level.¹⁷⁰ Perceptions were dependent on the type of trial.¹⁷¹ While the addition of relationship evidence in a separate trial did not increase mock jurors' perceptions of the defendant's sexual interest in boys, tendency evidence presented in either the separate or joint trial increased the perception that the defendant had a sexual interest in boys, as shown in Figure 2. Additional evidence from independent sources rather than additional evidence from the same complainant changed mock jurors' perceptions of the defendant's behaviour.

¹⁶⁹ Non-penetrative offence: $\beta = -3.073$, $SE = 0.828$, $Z = -3.711$, $p < .001$, not guilty: $Std Dev = 1.46$, guilty: $Std Dev = 1.55$; penetrative offence: $\beta = -2.997$, $SE = 0.916$, $Z = -3.271$, $p = .001$, not guilty: $Std Dev = 1.47$, guilty: $Std Dev = 1.55$.

¹⁷⁰ Juror level: $\beta = -0.004$, $SE = 0.011$, $Z = -0.384$, $p = .701$; jury level: $\beta = 0.018$, $SE = 0.068$, $Z = 0.269$, $p = .788$.

¹⁷¹ Relationship evidence trial versus separate trial: $\beta = 0.290$, $SE = 0.395$, $Z = 0.734$, $p = .463$; tendency evidence trial versus separate trial: $\beta = 2.318$, $SE = 0.598$, $Z = 3.876$, $p < .001$; joint trial versus separate trial: $\beta = 3.00$, $SE = 0.364$, $Z = 8.243$, $p < .001$.

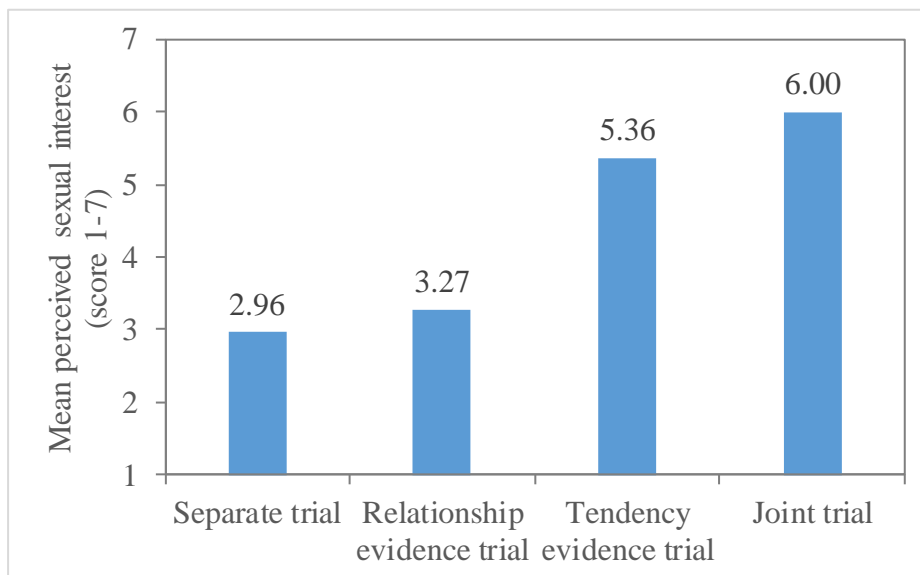


Figure 2. Perceptions of the defendant’s sexual interest in boys, by trial type

A strong effect of the influence of group deliberation emerged ($ICC = .426$); that is, 42.6 per cent of the variance in ratings of the defendant’s sexual interest in boys was attributable to the jury groups.

The criminal intent of the defendant

A composite measure of the defendant’s criminal intent was devised from responses indicating whether the defendant “abused the trust of others”, “abused his position as a coach”, “was responsible for what happened to him” and “was a risk to other boys”. We added scores reported for these items to form a single measure, called *Criminal Intent*, based on Principal Component Analyses, which revealed that these responses were loading on the same component, sharing the same variance. Results of analyses using this composite variable provided insight into how the type of trial affects inferences drawn by mock jurors and juries regarding the motivation and behaviour of the defendant.

Jury-level analyses revealed an effect for the type of trial: as more evidence for the prosecution was presented, mock jurors were more likely to infer the defendant’s *Criminal Intent*, as shown in Figure 3.¹⁷² The defendant’s *Criminal Intent* did not depend on juror or jury pre-trial

¹⁷² Separate trial versus relationship evidence trial: $\beta = 0.732$, $SE = .196$, $Z = 3.730$, $p < .001$; separate trial versus tendency evidence trial: $\beta = 1.890$, $SE = .292$, $Z = 6.484$, $p < .001$; separate trial versus joint trial: $\beta = 2.374$, $SE = .219$, $Z = 10.583$, $p < .001$.

dispositions; that is, *Child Sexual Abuse Knowledge Questionnaire (CSA-KQ)*, *Conviction Proneness (PJAQ)* or *Innate Criminality (PJAQ)*.¹⁷³

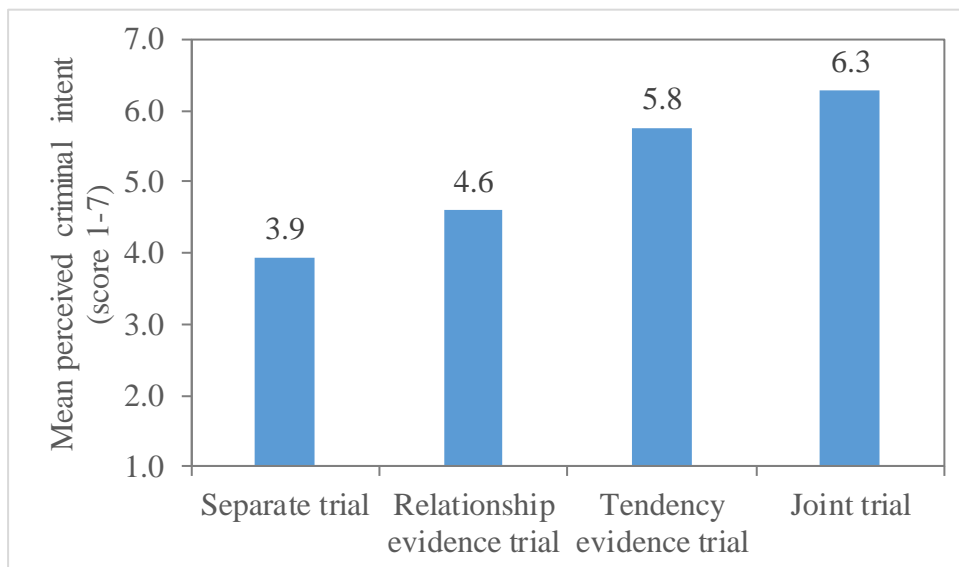


Figure 3. Perceived criminal intent of the defendant, by trial type

Multi-level regression analysis revealed a strong intra-class correlation ($ICC = .313$), indicating that a high proportion (31.3 per cent) of the inferred *Criminal Intent* was attributable to the jury groups.

We conducted further multi-level analyses to assess the influence of perceived *Criminal Intent* on verdict, after controlling for juror *Child Sexual Abuse Knowledge*. Results revealed that the defendant's perceived criminal intent predicted the verdict at both the juror and the jury level, unrelated to the offence type.¹⁷⁴ Specifically, the odds of conviction were 2.5 times greater when mock jurors perceived that the defendant had more *Criminal Intent*. Similarly to the results reported above, juror pre-trial dispositions (*Child Sexual Abuse Knowledge*) did not influence their verdicts, and this result remained constant when PJAQ factors *Conviction Proneness* and *Innate Criminality* were considered in the model.¹⁷⁵ These findings

¹⁷³ $p > .10$.

¹⁷⁴ Non-penetrative offence: Juror level: *Child Sexual Abuse Knowledge*: $\beta = 0.001$, $SE = 0.022$, $Z = 0.045$, $p = .964$, Odds Ratio = 1.001, 95% CI [0.958; 1.046]; *Criminal Intent*: $\beta = 0.835$, $SE = 0.171$, $Z = 4.886$, $p < .001$, Odds Ratio = 2.306, 95% CI [1.649; 3.223]. Jury level: *Child Sexual Abuse Knowledge*: $\beta = -0.077$, $SE = 0.157$, $Z = -0.487$, $p = .626$; *Criminal Intent*: $\beta = 2.713$, $SE = 0.565$, $Z = 4.801$, $p < .001$. Penetrative offence: Juror level: *Child Sexual Abuse Knowledge*: $\beta = 0.004$, $SE = 0.021$, $Z = 0.187$, $p = .852$, Odds Ratio = 1.004, 95% CI [0.964; 1.046]; *Criminal Intent*: $\beta = 0.937$, $SE = 0.195$, $Z = 4.801$, $p < .001$, Odds Ratio = 2.552, 95% CI [1.741; 3.741]. Jury level: *Child Sexual Abuse Knowledge*: $\beta = 0.154$, $SE = 0.124$, $Z = 1.247$, $p = .213$; *Criminal Intent*: $\beta = 2.809$, $SE = 0.659$, $Z = 4.265$, $p < .001$.

¹⁷⁵ $p > .05$. PJAQ factors *Innate Criminality* and *Conviction Proneness* did not predict verdict.

demonstrated that juror perceptions of the defendant's motivations were influenced by the evidence presented at trial, not by their own pre-trial attitudes.

Blame of the complainant with the moderately strong claim

In the post-trial questionnaire, mock jurors rated the extent to which Timothy was responsible for the alleged sexual assault. Mock jurors who viewed the basic separate trial were more likely to blame the complainant ($M = 2.47$) than jurors assessing the same allegations in trials with relationship evidence ($M = 2.17$), with tendency evidence ($M = 1.75$) or in a joint trial ($M = 2.08$).¹⁷⁶ The strongest effect emerged for a tendency evidence trial versus a separate trial; that is, the difference between perceptions that Timothy was responsible for what transpired was most extreme when comparing ratings from mock jurors in the separate trial (more responsible) than the tendency evidence trial (less responsible). There were smaller differences in the perceptions of Timothy's responsibility between the separate trial and the relationship evidence or the joint trial. Multi-level regression analysis revealed that 5.2 per cent of the variance in ratings of victim blame was attributable to the jury groups.

The credibility of the complainant with the moderately strong claim

We assessed the perception of Timothy's credibility using the two factors *Poise* and *Communication Style*, from the Observed Witness Efficacy Scale.¹⁷⁷ *Poise* measures the complainant's confidence, emotional control and anxiety management, whereas *Communication Style* refers to the complainant's verbal and non-verbal behaviour. Both of these factors are associated with the credibility of a witness on the witness stand. We conducted multi-variate, multi-level regression analyses with the two factors of the CSA-KQ (*Impact of Sexual Abuse on Children and Contextual Influences on Report*) as juror-level predictors, and type of trial as jury-level predictors.

The factors *Poise* and *Communication Style* correlated significantly with each other at the mock juror level, but were independent of each other at the jury level.¹⁷⁸ The intra-class correlation was $ICC = .082$ for *Poise* and $ICC = .055$ for *Communication Style*, indicating that the jury

¹⁷⁶ Relationship evidence trial versus separate trial: $\beta = -0.295$, $SE = 0.125$, $Z = -2.359$, $p = .018$; tendency evidence trial versus separate trial: $\beta = -0.713$, $SE = 0.132$, $Z = -5.418$, $p < .001$; joint trial versus separate trial: $\beta = -0.390$, $SE = 0.164$, $Z = -2.380$, $p = .017$. Separate trial: $Std Dev = 1.03$; relationship evidence trial: $Std Dev = 1.06$; tendency evidence trial: $Std Dev = 0.93$, joint trial: $Std Dev = 1.33$.

¹⁷⁷ Cramer, DeCoster, Neal and Brodsky, above n 18.

¹⁷⁸ Juror level: $\beta = 0.370$, $SE = 0.034$, $Z = 10.942$, $p < .001$; Jury level: $\beta = 0.001$, $SE = 0.048$, $Z = 0.013$, $p = .990$.

groups were responsible for 8.2 per cent and 5.5 per cent, respectively, of the variance in the perceived credibility of the complainant.

Analyses revealed that at the mock juror level, juror knowledge about the *Impact of Sexual Abuse on Children*¹⁷⁹ did not predict the perceived credibility of the complainant, that is, this knowledge did not affect *Poise* and *Communication Style*. However, the second factor of the CSA-KQ, mock juror knowledge about *Contextual Influences on Report*, predicted the perceived credibility of the complainant, as measured by both *Poise* and *Communication Style*.¹⁸⁰ Specifically, the more mock jurors knew about *Contextual Influences on Report*, the more likely they were to believe the complainant with respect to both factors measuring perceived complainant credibility.

When the influence of the jury groups was taken into account, the perceived credibility of the complainant was rated as equivalent in the separate trial and the relationship evidence trial.¹⁸¹ By contrast, in the tendency evidence trial¹⁸² and joint trial¹⁸³, mock jurors were more likely to rate the complainant's *Poise* and *Communication Style* more favourably, showing that jurors gave Timothy a higher credibility score when his evidence was given in the presence of other evidence about similar events from independent witnesses or complainants. Figure 4 shows the mean perceived credibility of the focal complainant as measured by the OWES factors *Poise* and *Communication Style*.

¹⁷⁹ *Poise*: $\beta = -0.008$, $SE = 0.007$, $Z = -1.010$, $p = .313$; *Communication Style*: $\beta = -0.005$, $SE = 0.007$, $Z = -0.775$, $p = .438$.

¹⁸⁰ *Poise*: $\beta = 0.022$, $SE = 0.009$, $Z = 2.341$, $p = .019$; *Communication Style*: $\beta = 0.017$, $SE = 0.009$, $Z = 2.003$, $p = .045$.

¹⁸¹ *Poise*: $\beta = 0.087$, $SE = 0.087$, $Z = 0.999$, $p = .318$; *Communication Style*: $\beta = 0.129$, $SE = 0.082$, $Z = 1.585$, $p = .113$.

¹⁸² *Poise*: $\beta = 0.356$, $SE = 0.101$, $Z = 3.530$, $p < .001$; *Communication Style*: $\beta = 0.304$, $SE = 0.081$, $Z = 3.736$, $p < .001$.

¹⁸³ *Poise*: $\beta = 0.504$, $SE = 0.085$, $Z = 5.938$, $p < .001$; *Communication Style*: $\beta = 0.403$, $SE = 0.073$, $Z = 5.492$, $p < .001$.

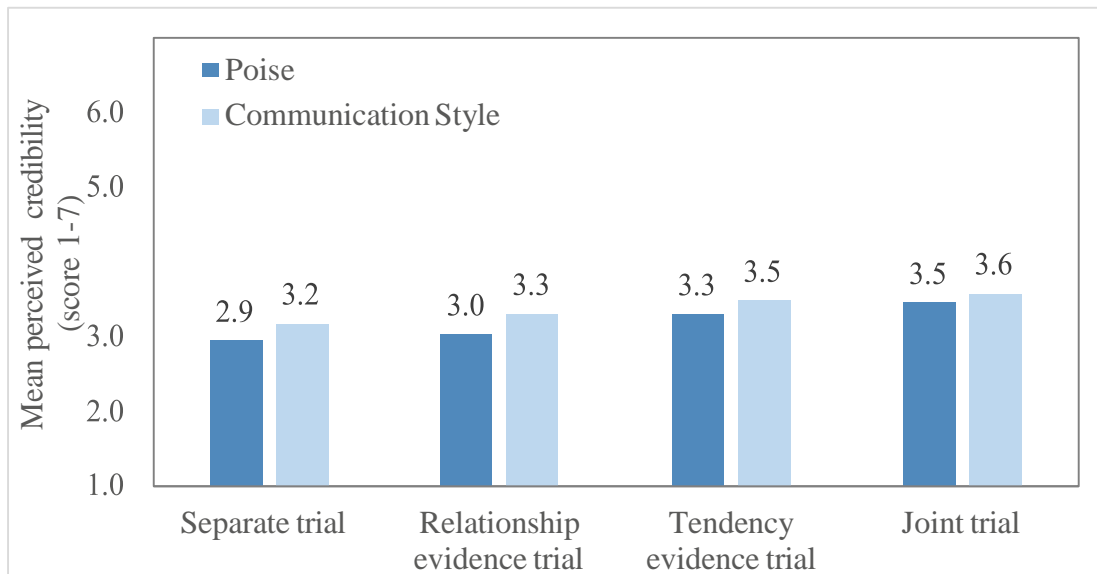


Figure 4. Mean perceived credibility of the complainant with moderately strong evidence, by trial type

The results showed that Timothy’s credibility ratings were enhanced by the evidence of two other independent witnesses or complainants who reported similar experiences with the defendant, irrespective of whether the defendant was charged with counts pertaining to those individuals, or whether they gave their evidence in a separate or joint trial. The addition of relationship evidence had no impact on Timothy’s credibility scores, as that additional information was from the same source – that is, the complainant himself. These findings indicated that when assessing Timothy’s credibility, mock jurors were extremely sensitive to the source of the information rather than the content of the information alone.

How convincing was the focal complainant?

To further explore how the type of trial affected the credibility of the focal complainant, we compared ratings of the perceived convincingness of Timothy’s evidence by type of trial.¹⁸⁴ Convincingness was used in addition to *Poise* and *Communication Style* (see above) to measure the perceived credibility of the complainant. Analyses revealed that jurors’ ratings of how convincing they perceived the complainant to be positively correlated with *Poise* and *Communication Style*, showing a medium to large effect size.¹⁸⁵ The higher the ratings of the complainant’s verbal and non-verbal behaviour (*Communication Style*), and the higher the

¹⁸⁴ Trial 1 versus 2 versus 5 versus 10.

¹⁸⁵ *Poise*: $r = .407, p < .001$; *Communication Style* $r = .442, p < .001$.

complainant's emotional control (*Poise*), the more convincing the complainant was perceived to be.

In all of the trials, Timothy's allegations and evidence remained consistent. Results confirmed that Timothy was rated as more convincing when the prosecution presented more supporting evidence:¹⁸⁶

separate trial (M = 4.03) < relationship evidence trial (M = 4.88) <
tendency evidence trial (M = 5.15) < joint trial (M = 5.87).¹⁸⁷

This analysis revealed that the context in which the evidence of the focal complainant was presented changed the weight accorded to that evidence. As more independent sources of information were introduced to support Timothy's account, his credibility increased, his evidence was accorded more weight, and jurors were more comfortable relying on his words as a source of evidence to support the verdict they reached.

These quantitative results from the individual mock jurors were reflected in mock jurors' comments in the course of the jury deliberations. In the basic separate and relationship evidence trials, mock jurors were more likely to express the belief that the evidence was simply one person's word against another's than they were in trials involving tendency evidence. Similarly, jurors were more likely to express the belief that oral evidence was not 'real' evidence in the separate and relationship evidence trials than they were in the tendency evidence and joint trials. For example, in Jury 13¹⁸⁸, Juror 12 expressed the belief that oral evidence was not evidence, although this was challenged and corrected by Juror 10:

JUROR 9: What's the evidence?

JUROR 10: The evidence is his testimony.

JUROR 12: That is not evidence, though.

JUROR 10: Yeah, it is evidence. The testimony is evidence; especially if you don't have any other evidence.

Analyses of the deliberations revealed that juror reluctance to rely on oral evidence in cases of childhood sexual abuse was not always overcome by the presence of supporting evidence from

¹⁸⁶ Relationship evidence trial versus separate trial: $\beta = 0.861$, $SE = 0.217$, $Z = 3.967$, $p < .001$; tendency evidence trial versus separate trial: $\beta = 1.128$, $SE = 0.277$, $Z = 4.076$, $p < .001$; joint trial versus separate trial: $\beta = 1.837$, $SE = 0.197$, $Z = -9.345$, $p < .001$.

¹⁸⁷ Separate trial: $Std Dev = 1.48$; relationship evidence trial: $Std Dev = 1.47$; tendency evidence trial: $Std Dev = 1.84$; joint trial: $Std Dev = 1.20$.

¹⁸⁸ Trial 2, separate trial with relationship evidence and standard jury directions.

independent witnesses. For instance, in Jury 5¹⁸⁹, Juror 5 discredited the complainant's oral evidence as 'hearsay evidence'. Similarly, in Jury 77¹⁹⁰, Juror 12 stated that there was no 'true' evidence because it was just a case of 'his word against his word'.

The factual culpability of the defendant

First, we present results for the focal complainant across all four types of trials. This is followed by results for all three witnesses/complainants.

Factual culpability ratings for the focal complainant

To explore mock jurors' reasoning in more detail, and to assess factors that predicted factual culpability on each of the counts, we used scores on the perceived convincingness of the focal complainant's evidence on each count, and scores on mock juror *Child Sexual Abuse Knowledge* – at the juror level, and as mean group scores at the jury level. In addition, type of trial was examined as a predictor of the perceived factual culpability of the defendant. These analyses were conducted separately for the two counts involving Timothy, so we could examine differences and similarities in the way mock jurors reasoned about these counts individually and as a group.

Results of these analyses revealed a strong effect for deliberation: the intra-class correlation was $ICC = .296$ for factual culpability for the non-penetrative offence, and $ICC = .286$ for the penetrative offence. This indicated that 29.6 per cent and 28.6 per cent of the variance in ratings of factual culpability was attributable to the jury groups, respectively.

Juror-level analyses revealed that the extent to which the complainant was perceived as convincing predicted factual culpability.¹⁹¹ The more mock jurors believed the complainant (Timothy), the more likely they were to perceive that the defendant was factually culpable for the alleged acts. Mock jurors' *Child Sexual Abuse Knowledge* did not predict a perception of factual culpability on either count.¹⁹²

¹⁸⁹ Trial 5, separate trial with tendency evidence and standard jury directions.

¹⁹⁰ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

¹⁹¹ Non-penetrative offence: $\beta = 0.410$, $SE = 0.058$, $Z = 7.096$, $p < .001$; penetrative offence: $\beta = 0.465$, $SE = 0.063$, $Z = 7.422$, $p < .001$.

¹⁹² Non-penetrative offence: $\beta = 0.005$, $SE = 0.008$, $Z = 0.607$, $p = .544$; penetrative offence: $\beta = 0.001$, $SE = 0.009$, $Z = 0.148$, $p = .883$.

Mean jury *Child Sexual Abuse Knowledge* predicted factual culpability of the defendant for the penetrative offence but not the non-penetrative offence¹⁹³, indicating that the more mock jurors knew about child sexual abuse the more likely they were to perceive the defendant as factually culpable for the more serious offence. Similarly, the effect of the trial type on factual culpability was different for the non-penetrative and the penetrative offences.¹⁹⁴ While the admission of relationship evidence significantly increased the perceived factual culpability for the penetrative offence, factual culpability remained the same for the non-penetrative offence. By contrast, additional witnesses in the tendency evidence trial and additional complainants in the joint trial significantly increased perceived factual culpability for both the penetrative and non-penetrative offences when compared to the separate trial and unrelated to offence type. As more evidence was presented on behalf of the Crown, the effect increased, unrelated to the offence type.

These results suggested that mock jurors processed evidence differently for the non-penetrative and the penetrative offence. Since perceptions of factual culpability were higher for the non-penetrative offence than the penetrative offence, juries with less *Child Sexual Abuse Knowledge* gave lower ratings for factual culpability for the penetrative offence than juries with higher *Child Sexual Abuse Knowledge*. Furthermore, the admission of relationship evidence in the separate trial significantly increased perceived factual culpability for the penetrative offence but not the non-penetrative offence, suggesting that juries were more reluctant to ascribe factual culpability for the more serious penetrative offence unless additional evidence was presented. In contrast, independent witnesses for the Crown increased the ratings of the factual culpability of the defendant, unrelated to the trial type – that is, regardless of whether it was a tendency evidence trial or a joint trial.

Factual culpability for each count by complainant

We also analysed mock jurors' perceptions of the defendant's factual culpability for each alleged offence separately. Table 5 presents results regarding the factual culpability of the defendant for the allegations of each complainant, by type of trial.

¹⁹³ Non-penetrative offence: $\beta = 0.012$, $SE = 0.050$, $Z = 0.236$, $p = .814$; penetrative offence: $\beta = 0.079$, $SE = 0.037$, $Z = 2.105$, $p = .035$.

¹⁹⁴ Non-penetrative offence: Separate trial versus relationship evidence trial: $\beta = 0.469$, $SE = 0.260$, $Z = 1.802$, $p = .072$; separate trial versus tendency evidence trial: $\beta = 1.470$, $SE = 0.335$, $Z = 4.384$, $p < .001$; separate trial versus joint trial: $\beta = 1.936$, $SE = 0.269$, $Z = 7.203$, $p < .001$. Penetrative offence: Separate trial versus relationship evidence trial: $\beta = 0.467$, $SE = 0.226$, $Z = 2.063$, $p = .039$; separate trial versus tendency evidence trial: $\beta = 1.241$, $SE = 0.410$, $Z = 3.030$, $p = .002$; separate trial versus joint trial: $\beta = 1.846$, $SE = 0.310$, $Z = 5.964$, $p < .001$.

When considering the tendency evidence, the separate trial and the joint trial results revealed ratings of greater factual culpability for the strong and moderately strong allegations than for the weak allegations, unrelated to the type of offence (non-penetrative or penetrative). A comparison of non-penetrative versus penetrative offences alleged by the same individual indicated that ratings of factual culpability were higher for non-penetrative offences than for penetrative offences, for both Timothy (moderately strong evidence) and Justin (strong evidence). The ratings showed that mock jurors differentiated between the factual culpability of the defendant in relation to claims of different evidential strength, and also claims of different severity (that is, non-penetrative versus penetrative offences).

The perceived factual culpability of the defendant for the counts in the moderately strong case also depended on the type of trial; factual culpability ratings increased as more inculpatory evidence against the defendant was admitted at trial: separate trial < relationship evidence trial < tendency evidence trial < joint trial.

Table 5. Factual culpability score of the defendant for each count, by complainant and type of trial

	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial
	<i>Mean (Std Dev)</i>	<i>Mean (Std Dev)</i>	<i>Mean (Std Dev)</i>	<i>Mean (Std Dev)</i>
Weak claim				
Masturbation of complainant	–	–	5.49 (1.47)	5.40 (1.69)
Moderately strong claims				
Masturbation of complainant	4.39 (1.63)	4.91 (1.56)	5.92 (1.36)	6.34 (0.93)
Digital-anal penetration	4.12 (1.61)	4.75 (1.58)	5.58 (1.55)	5.98 (1.34)
Strong claims				
Masturbation of defendant	–	–	5.86 (1.41)	6.36 (0.88)
Masturbation of complainant	–	–	5.84 (1.41)	6.42 (0.78)
Penile-oral penetration	–	–	5.62 (1.59)	5.99 (1.37)

Note. Scores ranged from 1 – 7.

Jury reasoning about the counts and verdict consistency

Chi-square analysis of jury verdicts – comparing convictions with acquittals and hung juries – revealed significant differences in conviction rates, unrelated to the type of offence.¹⁹⁵ Notably,

¹⁹⁵ Non-penetrative offence: $\chi^2 = 3.69$, $p = .055$, $\Phi = .480$; penetrative offence: $\chi^2 = 0.29$, $p = .590$, $\Phi = .135$.

differences in conviction rates between the tendency evidence trial and the joint trial were not significant. Mock jurors distinguished counts that were less serious (indecent) from more serious (penetrative offences) for the same complainant and between complainants (when the accused was charged with similar offences involving different complainants). Table 6 shows jury verdicts reflecting these distinctions.

Table 6. Jury verdicts for counts with moderately strong evidence (Timothy) by type of trial for (a) non-penetrative and (b) penetrative offence (per cent)

	Count 1: Masturbation of complainant			Count 2: Digital-anal penetration		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
Separate trial						
Relationship evidence trial	77.8	11.1	11.1	88.9	0.0	11.1
Tendency evidence trial	58.3	8.3	33.3	58.3	0.0	41.7
Joint trial	25.0	62.5	12.5	37.5	62.5	0.0
	0.0	100.0	0.0	25.0	75.0	0.0
(b) Juror verdict						
Separate trial						
Relationship evidence trial	81.0	19.0	–	90.5	9.5	–
Tendency evidence trial	77.8	22.2	–	80.7	19.3	–
Joint trial	28.6	71.4	–	33.3	66.7	–
	6.2	93.8	–	16.0	84.0	–

Note. Separate trial: nine juries, $n = 105$ jurors; relationship evidence trial: 12 juries, $n = 135$ jurors; tendency evidence trial: eight juries, $n = 85$ jurors; joint trial: eight juries, $n = 83$ jurors.

A multi-variate, multi-level logistic regression assessed the verdicts for the penetrative and non-penetrative counts by type of trial, taking into account the overall *Child Sexual Abuse Knowledge* (scores on the CSA-KQ) and the perceived consistency of the complainant's evidence. The perceived consistency of the complainant was used as a jury-level variable.

Juror-level analyses revealed that mock jurors' *Child Sexual Abuse Knowledge* did not predict verdicts on either count.¹⁹⁶ However, perceived consistency of the complainant Timothy

¹⁹⁶ Non-penetrative offence: $\beta = -0.020$, $SE = 0.023$, $Z = -0.867$, $p = .386$, Odds Ratio = 0.981, 95% CI [0.938; 1.025]; penetrative offence: $\beta = -0.007$, $SE = 0.022$, $Z = -0.313$, $p = .754$, Odds Ratio = 0.993, 95% CI [0.952; 1.036].

predicted the verdict: the odds of conviction were 2.6 time greater for the non-penetrative offence and 2.7 times greater for the penetrative offence when the mock jurors perceived the complainant as more consistent.¹⁹⁷

Jury-level analyses likewise indicated that the mean mock juror *Child Sexual Abuse Knowledge* did not predict verdict on either count¹⁹⁸; rather, the verdict depended on the type of trial. While the admission of relationship evidence did not increase convictions on either count compared to the conviction rates in the separate trial¹⁹⁹, the admission of tendency evidence increased convictions for both the non-penetrative and penetrative offences.²⁰⁰ Similarly, in the joint trial, the conviction rate for both the non-penetrative and the penetrative offence was significantly higher than in the separate trial.²⁰¹

An examination of verdict consistency between counts yielded a different pattern by type of trial. In the separate trial, most (81.0 per cent) of the individual mock jurors who voted to acquit did so on both counts, 9.5 per cent found him guilty on both counts, and 9.5 per cent voted guilty on one count and not guilty on another. In the relationship evidence trial, 77.0 per cent voted to acquit on both counts, 18.5 per cent found him guilty on both counts, and 4.0 per cent reached different verdicts on the two counts. In the tendency evidence trial, 28.6 per cent voted to acquit on both counts, 66.7 per cent found him guilty on both counts, and different verdicts were returned in 4.8 per cent of the cases. In the joint trial, 4.9 per cent of the mock jurors voted to acquit on both counts, compared to 82.7 per cent with guilty verdicts on both counts, and 12.3 per cent of the mock jurors returned a verdict of guilty on one count and not guilty on another. There were no significant differences between trial groups in the separation of the counts. Analyses of verdict consistency for Counts 1 and 2 revealed that individual mock jurors were more likely to convict the defendant on both counts when more evidence against the defendant was admitted – that is, more likely (a) in a separate trial with tendency evidence and

¹⁹⁷ Non-penetrative offence: $\beta = 0.934$, $SE = 0.167$, $Z = 5.581$, $p < .001$, Odds Ratio = 2.574, 95% CI [1.856; 3.570]; penetrative offence: $\beta = 1.010$, $SE = 0.244$, $Z = 4.136$, $p < .001$, Odds Ratio = 2.745, 95% CI [1.701; 4.430].

¹⁹⁸ Non-penetrative offence: $\beta = -0.077$, $SE = 0.071$, $Z = -1.091$, $p = .275$; penetrative offence: $\beta = -0.088$, $SE = 0.051$, $Z = -1.716$, $p = .086$.

¹⁹⁹ Non-penetrative offence: $\beta = 0.709$, $SE = 1.378$, $Z = 0.514$, $p = .607$; penetrative offence: $\beta = 1.266$, $SE = 1.252$, $Z = 1.011$, $p = .312$.

²⁰⁰ Non-penetrative offence: $\beta = 6.245$, $SE = 2.256$, $Z = 2.768$, $p = .006$; penetrative offence: $\beta = 6.835$, $SE = 2.608$, $Z = 2.620$, $p = .009$.

²⁰¹ Non-penetrative offence: $\beta = 8.165$, $SE = 1.850$, $Z = 4.413$, $p < .001$; penetrative offence: $\beta = 8.465$, $SE = 2.426$, $Z = 3.476$, $p < .001$.

in the joint trial than (b) in a separate trial.²⁰² There were no significant differences in verdict consistency when comparing a separate trial with one that included relationship evidence.

The findings on verdict were supported by ratings of the likelihood that the defendant committed the alleged acts, even where the mock jurors concluded that the claim was not proved beyond reasonable doubt.

Inspection of the conviction rates revealed that mock jurors and juries both appeared more cautious about convicting for the penetrative offence compared to the non-penetrative offence when they were prosecuted in a separate trial. However, in the tendency evidence and joint trials, the conviction rates for the penetrative offence were 62.5 per cent and 75 per cent respectively – compared to zero in the separate trials – suggesting that multiple witnesses or complainants made an allegation of penetration more plausible and credible.

The admission of relationship evidence did not increase the conviction rate for either count. Instead, it resulted in a relatively high number of hung juries, suggesting that the admission of such evidence from the same complainant did not increase the perceived trustworthiness of the source, and, in the absence of any independent supporting evidence, may have increased the ambiguity of the case facts, resulting in less unanimity one way or the other.

²⁰² Relationship evidence trial versus separate trial: $\beta = 0.708$, $SE = 1.027$, $Z = 0.690$, $p = .490$; tendency evidence trial versus separate trial: $\beta = 5.410$, $SE = 1.837$, $Z = 2.946$, $p = .003$; joint trial versus separate trial: $\beta = 6.996$, $SE = 1.498$, $Z = 4.669$, $p < .001$.

4.3 Jury reasoning by type of trial

Contrary to expectations, juries in this study were not prone to impermissible reasoning and made very few factual errors. Furthermore, juries had the capacity to self-correct the vast majority of errors when they did occur. Impermissible reasoning was more likely to occur in the separate trials without tendency evidence than in the trials *with* tendency evidence. We also found no evidence of emotional or illogical reasoning in any of the trials in which tendency evidence had been admitted.

An analysis of mock jurors' specific reasons for their decisions to convict provided negligible support for the notion that joint trials produce verdicts based on impermissible character prejudice and accumulation prejudice. As instructed by the trial judge, mock jurors used their common knowledge and experience of the world to help them understand the behaviours of the complainants and the defendant.

Factual errors: We found a low rate of errors in the jury deliberations; only 7.7 per cent of juries made more than two factual errors. Two or more factual errors were more likely to occur in jury deliberations of the joint trial – that is, the trial with the most complex evidence. When errors were made, the jury self-corrected the vast majority during the course of deliberations.

Impermissible reasoning: Juries were not prone to impermissible reasoning. Such impermissible reasoning was more common in the separate and relationship evidence trials than in the tendency and joint evidence trials.

Lower standard of proof; emotional or illogical reasoning: We found only two jurors who appeared to use a lower standard of proof than the criminal standard, and only two jurors whose verdicts were driven by emotion. None of the juries featured a juror who reasoned illogically about the evidence. None of the trials in which tendency evidence was admitted prompted any instances of these three types of impermissible reasoning.

Reasons for convicting: The reasons jurors cited as the basis for their decisions to convict provided negligible support for the notion that joint trials produce verdicts based on impermissible character prejudice and accumulation prejudice.

4.3.1 Research aim

In this section, to further examine the hypothesis that juries engage in impermissible reasoning in joint trials, and to supplement the results from the post-trial questionnaires, we analysed other sources of data, namely the transcribed jury deliberations and the written reasons given by individual jurors for their verdicts. In all instances, the purpose was to assess the prevalence of impermissible reasoning, and if impermissible reasoning was found, to determine whether it resulted in unfair prejudice to the defendant.

First, we present the quantitative outcomes of coding factual errors and impermissible reasoning in all deliberating juries by trial type. Next, we provide results of a qualitative review of individual mock jurors' reasons for their verdicts, by trial type. This is followed by an overview of jury reasoning in the deliberations in separate trials. The section concludes with a qualitative thematic analysis of jury reasoning regarding the allegations by the focal complainant in all 33 joint trials.

4.3.2 The prevalence of impermissible reasoning in jury deliberations

The jury deliberation transcripts were coded to establish whether any of the three types of potential reasoning prejudice was prevalent, and if so, whether this was caused by joinder.

Inter-case conflation of the evidence

In Part 4.2 we reported that based on post-trial responses to six multiple-choice questions, lower juror factual accuracy was associated with higher conviction rates. All of the multiple-choice questions were narrow in scope: they all pertained to the focal complainant because his evidence was common to all trial types. To explore the influence of factual accuracy on verdict in more depth (and also more broadly), factual accuracy was coded for each jury deliberation. This coding took into consideration errors made by any one of the jurors in the 90 mock juries, including legal and factual errors. Overall, the results revealed a low error rate in the jury deliberations. The proportion of juries with more than two factual errors was 7.7 per cent, although only juries that deliberated about the joint trial made three or four errors. As the evidence became more complex, more jurors made factual errors in their deliberations²⁰³, as shown in Table 7.

²⁰³ Kendall's tau_b = .291, $p < 0.005$; Spearman's rho = .333, $p < 0.001$.

Table 7. Proportion of factual errors in all jury deliberations, by trial type (per cent)

Number of errors	Overall	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial
0	36.7	55.6	38.7	52.9	21.2
1	40.0	22.2	54.8	41.2	30.3
2	15.6	22.2	6.5	5.9	27.3
3	4.4	–	–	–	12.1
4	3.3	–	–	–	9.1

Note. Basic separate trial (Trial 1) $n = 9$ juries; relationship evidence trial (trials 2, 3 and 4) $n = 31$ juries; tendency evidence trial (trials 5 and 6) $n = 17$ juries; joint trial (trials 7, 8, 9 and 10) $n = 33$ juries.

Given that the joint trial videos ran for 1 hour and 50 minutes – compared to the separate trial, which ran for 45 minutes – it is unsurprising that the working memory of jurors in the joint trials was taxed by a greater volume of information to process and a greater cognitive load, which increased their susceptibility to errors.

Juries self-corrected the majority (82.2 per cent) of the factual errors made in the course of deliberations (see Table 8). Furthermore, there was no statistically significant difference in the number of uncorrected errors between types of trials²⁰⁴; that is, juries that deliberated in the tendency evidence trial and the joint trial were not statistically more likely to fail to correct their errors than were other juries.

Table 8. Proportion of uncorrected factual errors in all jury deliberations, by trial type (per cent)

Number of errors	Overall	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial
0	82.2	100.0	87.1	82.4	72.2
1	15.6	–	12.9	17.6	21.2
2	2.2	–	–	–	6.1

Note. Basic separate trial = Trial 1; relationship evidence trial = Trials 2, 3 and 4; tendency evidence trial = Trials 5 and 6; joint trial = Trials 7, 8, 9 and 10.

²⁰⁴ $p > 0.05$

None of the uncorrected errors were instances of inter-case conflation of evidence (confusing the evidence of different complainants). Rather, these errors were intra-case errors, as juries struggled to identify which conduct by the defendant against a particular complainant was the subject of which charge, and which conduct was uncharged. This difficulty is demonstrated in the following extract from the deliberations of Jury 35²⁰⁵, concerning the three counts arising from the allegations of the strong complainant (Justin). This jury had trouble identifying the conduct that was the subject of Justin's three charges. Because of some inter-case confusion, the jury discussed an incident on the back porch about which Timothy, not Justin, had given evidence, but which was not the subject of any charge:

JUROR 2: And what were the three for Justin then? There was the one in the car -----

JUROR 4: Justin is 2, 3 and 4.

JUROR 2: Yeah, but what were they?

JUROR 4: Okay, Justin – hang on, write this down.

JUROR 2: There was the incident in the car.

JUROR 4: Yeah, there was the incident at the house in the bed.

JUROR 2: And was the other one on the patio or something.

JUROR 5: Yeah, it was on the back porch -----

JUROR 6: No, in the kitchen, the patio -----

JUROR 2: But what did the accused allegedly do?

JUROR 4: Uhm, hang on.

JUROR 2: Is that where he put ice down his pants?

(Most jurors voiced agreement).

JUROR 2: Okay.

JUROR 4: Yes, it was on the patio at the home -----

JUROR 5: Yes.

JUROR 4: He put ice down his pants and then touched him.

JUROR 2: And touched him in -----

JUROR 3: No, I don't think he touched him.

JUROR 2: I don't think he touched him.

²⁰⁵ Trial 10, joint trial with standard jury directions.

JUROR 5: No, he pointed to it.

JUROR 3: Yeah, he didn't touch him.

The jury later realised its mistake:

JUROR 3: Mmm, alright. The act of indecency towards Justin McCutchen -----

JUROR 2: So this is the one with ice.

JUROR 3: Is it? No, no that is number 3.

(Most jurors voiced agreement).

JUROR 7: That is the third kid.

JUROR 8: With Timothy.

JUROR 3: The one with Justin is the one in the bed -----

JUROR 8: And in the car.

JUROR 4: Yeah, the oral sex in the car -----

Jury 35 later confused Justin's charges a second time. While Juror 4 correctly identified that Justin had given evidence about two acts of indecency occurring at the defendant's house ("[f]ondling of the guy and the guy masturbating him") and one act of sexual intercourse occurring in the defendant's car, the jury came to the incorrect conclusion that Justin had given evidence about one act of indecency occurring at the defendant's house, and one act of indecency and one act of sexual intercourse occurring in the defendant's car. Thus, when the jury later voted, they did so with an uncorrected error (the wrong conduct for one of the counts):

JUROR 3: Count 2 is – oh, hang on one of them is the ice -----

JUROR 4: In the bathroom.

JUROR 3: ----- I don't know which one?

JUROR 7: That is the third one.

JUROR 6: Masturbation -----

(All talk at once)

JUROR 7: It is count 3 -----

JUROR 9: Counts 2 and 3.

JUROR 1: It is three counts.

JUROR 4: Justin has got three counts. One of them is -----

JUROR 3: Two are acts of indecency, one is sexual intercourse.

JUROR 4: Mmm. I actually think the sexual intercourse was the oral sex in the car -----

JUROR 3: Park thing, yep.

JUROR 4: ----- on the way to the swimming pool.

[...]

JUROR 3: So what are the two acts of indecency towards Justin?

JUROR 4: He stayed over -----

JUROR 6: Sharing the bed.

JUROR 3: One was in the bed and one was in the car.

JUROR 6: He was fondling away – forced masturbation, that’s what wrote -----

JUROR 4: The two counts of indecency are the ones that relate to “in the bed, the forced masturbation” -----

JUROR 6: Fondling, playing with -----

JUROR 4: Fondling of the guy and the guy masturbating him; they are the two acts of indecency.

JUROR 3: Two? Are you sure one -----

JUROR 7: The second one is the finger?

JUROR 4: Oh, yeah.

JUROR 3: That is Timothy -----

(All talk at once)

JUROR 5: He cried and screamed out, because of the pain.

JUROR 6: That’s the third one.

JUROR 4: I think that was Timothy. That’s why there’s two counts. That is the sexual intercourse with Timothy.

JUROR 3: So the two acts of indecency is one in the bed; one of them putting his hand on his junk in the car; and then the sexual intercourse is in the car as well.

(Most jurors voiced agreement).

Jurors’ intra-complainant errors indicated a source-monitoring error, where a person attributes a memory to an incorrect source or experience.²⁰⁶ The fact that only intra-complainant errors went uncorrected indicated the difficulties jurors faced in dealing with large amounts of similar information, particularly when a number of similar events were linked to a single complainant.

²⁰⁶ D Stephen Lindsay, ‘Children’s Source Monitoring’ in H L Westcott, G M Davies and R H C Bull (eds), *Children’s Testimony* (John Wiley and Sons, 2002) 83, 85.

Unfair prejudice by type of trial

Jury deliberations were coded to assess whether they demonstrated impermissible reasoning in one of three ways: applying a lower standard of proof; reasoning emotionally, based on the evidence; or reasoning illogically about the evidence. Overall, two important findings emerged from this analysis. First, we found that juries were not prone to impermissible reasoning, and second, we discovered that impermissible reasoning was more common in the trials *without* tendency evidence – that is, it was more common in the separate and relationship evidence trials compared to the tendency evidence and joint evidence trials.

Only two juries out of 90 contained a juror who indicated that they were basing their verdict on a lower standard of proof than the criminal standard of proof. Both of these jurors indicated that they were making their assessments of the evidence ‘on the balance’, that is, on a standard closer to the balance of probabilities.²⁰⁷

Only two juries contained a juror whose verdict was driven by emotion. One female juror based her verdict on her views about the ‘despicable’ nature of paedophiles.²⁰⁸ The other juror was a woman whose son ‘was chased by a paedophile’, hence she found the idea of acquitting the defendant very distressing:²⁰⁹

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 ----

Later:

JUROR 4: The reason behind this is because my son was chased by a paedophile ----

JUROR 6: Oh, no.

JUROR 4: ----and he thought he would get past this. Okay. So I have a very strong ----

²⁰⁷ Jury 3, Trial 2, separate trial with relationship evidence and standard jury directions; Jury 23, Trial 4, separate trial with relationship evidence, standard and context evidence jury directions, and a question trail.

²⁰⁸ Jury 57, Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

²⁰⁹ Jury 44, Trial 1, basic separate trial with standard jury directions.

JUROR 6: Oh, okay.

JUROR 9: I don't think any of us really like paedophiles -----

JUROR 9: ----- just putting it out there. None of us do.

JUROR 4: I understand that.

JUROR 9: We are not supporting him as a paedophile. What we are saying is: there is not enough evidence to say that he committed the act for this case.

None of the trials in which tendency evidence was admitted prompted any instances of these three types of impermissible reasoning. None of the juries featured a juror who reasoned illogically about the evidence.

Table 9. Proportion of all jury deliberations with impermissible reasoning, by trial type (per cent/n)

Type of impermissible reasoning	Overall	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial
Lower standard of proof	2.2 (n = 2)	–	6.5 (n = 2)	–	–
Emotion driven	2.2 (n = 2)	11.1 (n = 1)	3.2 (n = 1)	–	–
Illogical (unrelated to evidence)	–	–	–	–	–

Note. Basic separate trial = Trial 1; relationship evidence trial = Trials 2, 3 and 4; tendency evidence trial = Trials 5 and 6; joint trial = Trials 7, 8, 9 and 10.

In the separate tendency evidence and joint trials, jurors appeared to expend particular effort in reasoning more systematically and objectively where there was a greater risk of prejudice. For example, after hearing a joint trial, juror 4 in Jury 90²¹⁰ stated (with bold added for emphasis):

JUROR 4: On number 1 I am unsure because I don't know for certain. I guess that is the best way to put it. I am thinking over it, in my mind, and the fact that I am kind of going here/there, it makes me think, like, am I – I am trying to pull – I suppose, like, analyse the way that I am thinking about it. I am trying to say, 'Am I thinking that he's guilty?' **I will admit, the first – like, before the whole video started, I was thinking about it to myself. I thought, 'I bet he's guilty'.** This is before any of this thing had come out. So I didn't understand what was going to happen after or whatever. I thought we were going – yeah, I didn't really know what the hell was going to happen. And I just thought, 'Oh, I bet he's

²¹⁰ Trial 9, joint trial with four prosecution witnesses, standard and tendency evidence jury directions.

guilty'. And maybe because of all the stuff that's been on media and there was that TV show – what was it? *Devil's Playground* – that was about the priests and stuff.

JUROR 12: I missed that.

JUROR 4: And there's been so much stuff, and, like, it is actually sick to the stomach when you read another one kind of happening. **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.

4.3.3 Qualitative review of mock jurors' main reasons for their verdict

Immediately after deliberations concluded, the mock jurors in the 90 juries were asked to specify the key factor in their decision making as a short, open-ended response to the question "What was the main reason for your verdict?" A small proportion of the mock jurors (2.33 per cent; $n = 23$) responded that they had disagreed with their fellow jurors, and provided no additional substantive explanations or reasons for their decisions. These responses were excluded from the qualitative thematic analyses of the reasons for decisions to convict or acquit the defendant.

We analysed results by verdict (conviction, acquittal or a hung jury) in relation to the two charges against the focal complainant with the moderately strong evidence (Counts 1 and 2 in the separate trials, and Counts 5 and 6 in the joint trials), since this evidence and these verdicts were common to all types of trials (separate, relationship evidence, tendency evidence and joint trials), thus permitting a comparison across all four trial types.

A researcher who was blind to the experimental condition or purpose of the study collated and qualitatively analysed the reasons for mock jurors' verdicts. Responses were coded based on word frequency and similarity. The results revealed a total of 12 discrete categories of reasons (Table 10). Reasons cited in support of both a verdict to convict or a verdict to acquit the defendant are listed under 'Both'.

Table 10. Major categories of reasons for mock jurors' verdicts, by verdict type

Conviction	Acquittal	Both
Consistency of multiple independent witnesses	Evidence insufficient to prove case	Counterintuitive behaviour of defendant/victim(s)
Strong evidence/witness credibility	Weak evidence/witness credibility	
Accumulation prejudice	Lack of supporting evidence/witnesses	
Pattern of grooming behaviour		
Character prejudice		
Weak testimony of the defendant	Credible testimony of the defendant	
Tendency evidence		

Mock jurors' reasons for conviction

For jurors who chose to convict, 489 answers were codable and included in the analysis. These responses were distributed across the four trial types: eight responses from the separate trials where the conviction rate was lowest; 62 from the relationship evidence trials where the conviction rate was higher but variable; 97 from the tendency evidence trials where the majority of trials resulted in a conviction; and 322 from the joint trials where the consensus to convict was greatest.

Across all four types of trials, when a decision to convict the defendant was made, this decision centred on one of three reasons: the persuasive force of the overall consistency of the evidence against the defendant; the credibility of the witnesses who appeared for the Crown; and the pattern of grooming behaviour by the defendant.

The two most frequently cited reasons for conviction each accounted for one-third of the responses. Of the jurors who voted to convict, 34.76% ($n = 170$) mentioned the consistency of details across the evidence provided by multiple independent witnesses. Of the jurors who convicted, 34.15% ($n = 167$) mentioned the strong evidence or credibility of prosecution witnesses. The pattern of grooming behaviour engaged in by the defendant was the third most common reason, nominated in all types of trials except for the basic separate trial, and accounting for 19.02% ($n = 93$) of the decisions to convict. Together, these three reasons

accounted for almost all of the decisions to convict (87.93%). Notably, all three were examples of permissible reasoning in support of a verdict to convict the defendant.

A small proportion of responses focused on the weak testimony or lack of credibility of the defendant (3.27%). Together, the proportion of responses that fell into the remaining five categories of reasons to convict accounted for less than 9 per cent of the responses.

Table 11 shows the proportion of responses by type of trial for each reason to convict.

Table 11. Reasons for mock jurors' decisions to convict by type of trial (per cent)

Decision-making factor	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial	Total
Consistency of multiple witnesses	0.00	16.13	47.42	35.40	34.76
Strong evidence/witness credibility	75.00	53.23	23.71	32.61	34.15
Pattern of grooming behaviour	0.00	11.29	19.59	20.81	19.02
Weak testimony of the accused	0.00	9.68	2.06	2.48	3.27
Character prejudice	0.00	1.61	3.09	2.80	2.66
Tendency evidence	0.00	0.00	3.09	3.73	3.07
Defendant behaviour is counterintuitive	25.00	8.06	1.03	0.93	2.25
Accumulation prejudice	0.00	0.00	0.00	0.93	0.61
Complainant behaviour is counterintuitive	0.00	0.00	0.00	0.31	0.20
Total (n = 489)					100

Notably, the two categories of impermissible reasoning – character prejudice and accumulation prejudice – together comprised a total 3.27 per cent of the reasons cited for mock jurors' decisions to convict the defendant, and more responses were classified as indicative of character prejudice (2.66 per cent) than accumulation prejudice (0.61 per cent). It is possible that these conclusions were reached using permissible reasoning, as the coding of these written comments, provided by the mock-jurors retrospectively, after a verdict was reached, was influenced by the brevity and at times, ambiguous nature of the stated reasons for a verdict. In sum, the reasons jurors cited as the basis for their decisions to convict provided negligible support for the notion that joint trials produce verdicts based on impermissible character

prejudice and accumulation prejudice. The following paragraphs provide examples of the 12 categories of reasons for conviction.

More than one-third of all jurors who voted to convict indicated that their decision was based on the consistency of accounts provided by multiple witnesses. Responses in this category demonstrated that the similarity across different witness accounts – especially the fact that the witnesses’ reports were independent of each other – strongly motivated their decision. Some representative examples of responses in this category included:

“Three separate people who don’t know each other make very similar claims.”²¹¹

“Main reason for me was that it was a repeat pattern of offences from three complete strangers.”²¹²

The second most prevalent decision-making factor was the strength of the witnesses’ evidence, and witness credibility. Responses in this category emphasised the strength of the details in the evidence, such as information about times or places that the witness could not have known if the alleged sexual abuse had not occurred. Additionally, responses in this category emphasised the credibility of witnesses. Some representative examples of responses in this category included:

“Too much evidence, story too strong to be fabricated.”²¹³

“Strong, conclusive evidence by the affected parties and sequences of events.”²¹⁴

“Timothy described the details of the actions and where it took place.”²¹⁵

The third most cited reason in support of a decision to convict – nominated in all trials except for the basic separate trial – was the defendant’s pattern of grooming behaviour. Representative examples of responses in this category included:

“Accused was in a position of trust. He groomed the boys and had a pattern.”²¹⁶

²¹¹ Juror on Jury 28, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

²¹² Juror on Jury 40, Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

²¹³ Juror on Jury 50 Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

²¹⁴ Juror on Jury 79, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

²¹⁵ Juror on Jury 81, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

²¹⁶ Juror on Jury 78, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

“There was a strong pattern of behaviour, victim choosing, grooming process, pattern of friendship with families, afterwards always disappeared off radar...”²¹⁷

Other reasons cited in support of a decision to convict the defendant – endorsed by between 1 and 4 per cent of the mock jurors – included character prejudice, reliance on tendency evidence, the low credibility of the defendant, accumulation prejudice, and unexpected or counterintuitive behaviour by the victim(s) or the defendant.

A representative example of a response from each of these categories is presented below:

Character prejudice: “MB [Mark Booth] seemed like a homosexual.”²¹⁸

Tendency evidence: “The level of tendency evidence strengthened the whole case against him.”²¹⁹

Weak credibility of the defendant: “I believe the accused lied about a lot of things...”²²⁰

Accumulation prejudice: “Too many incidents involving the same person.”²²¹

Behaviour of accused counterintuitive: “Why did the accused not visit the mother in hospital [sic] post operation?”²²²

There were some noteworthy differences in the reasons provided by the jurors who convicted across the four types of trials. In response to the basic separate and the relationship evidence trials, the most prevalent reason for a conviction was the strong evidence and credibility of the complainant (separate trial 75 per cent, $n = 6$; relationship evidence trial 53.23 per cent, $n = 33$). In the two types of trials in which tendency evidence was admitted, where mock jurors were exposed to evidence from multiple different sources, the consistency of the evidence influenced their verdicts (tendency evidence trials 47.42 per cent, $n = 46$; joint trials 35.40 per cent, $n = 114$).

These differences were logically related to the critical features of the evidence presented in the different trial types. When presented with the testimony of only one complainant and a supporting witness in the basic separate and relationship evidence trials, mock jurors relied more extensively on their evaluations of the strength of the complainant’s evidence. When

²¹⁷ Juror on Jury 70, Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

²¹⁸ Juror on Jury 53, Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

²¹⁹ Juror on Jury 28, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

²²⁰ Juror on Jury 57, Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

²²¹ Juror on Jury 35, Trial 10, joint trial with standard jury directions.

²²² Juror on Jury 10, Trial 2, separate trial with relationship evidence and standard jury directions.

tendency evidence was presented, jurors focused on taking into consideration the evidence of all the witnesses, and on examining the consistencies or inconsistencies across different witness accounts. Figure 5 shows the major reasons for mock jurors' decisions to convict, by type of trial.

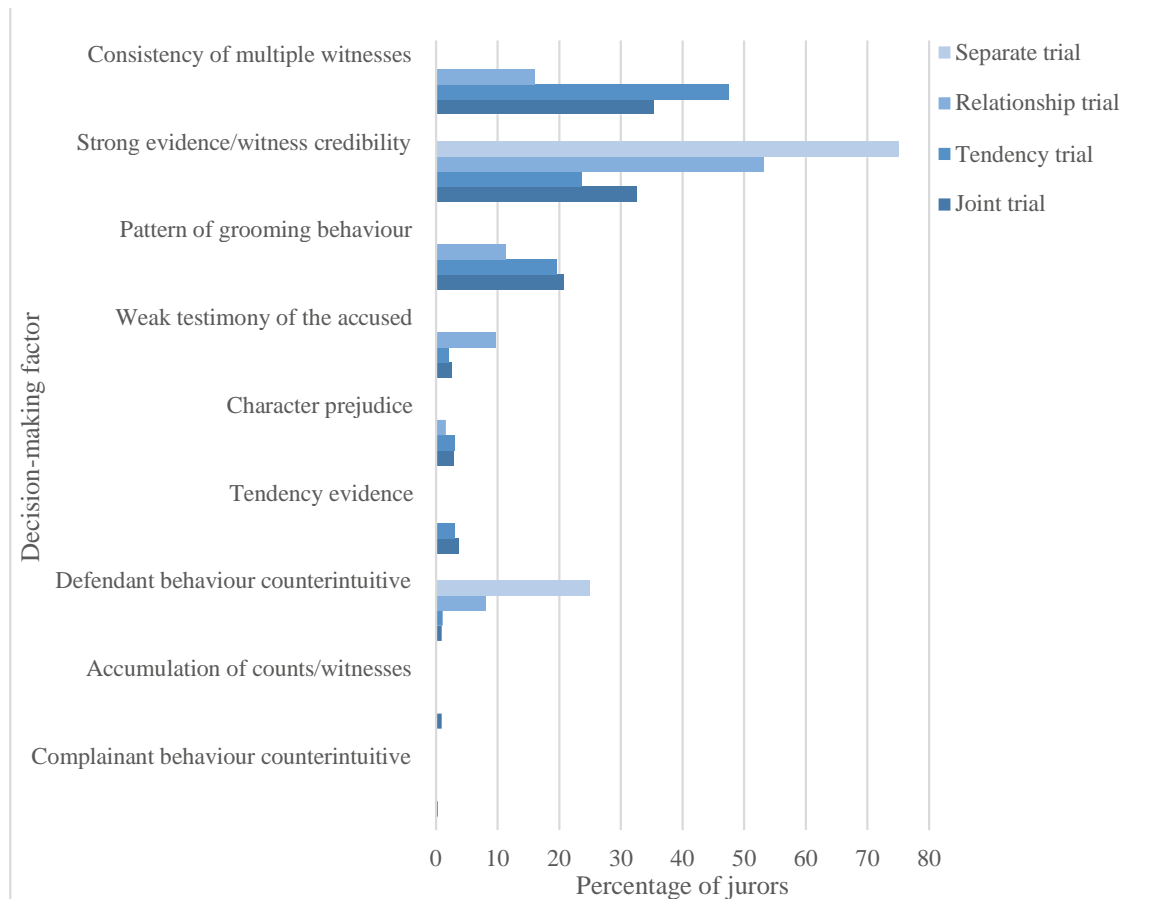


Figure 5. Key reasons for conviction, by trial type (per cent)

Mock jurors' reasons for acquitting

For jurors who voted to acquit, 302 answers were codable and included in the analysis of reasons for verdicts pertaining to the counts arising from the focal complainant's evidence. These jurors were distributed across the four trial types: 80 from the basic separate trials, 166 from the relationship evidence trials, 43 from the tendency evidence trials and 13 from the joint trials.

Across all four trial types, where the jury verdict was to acquit the defendant, the most frequently cited reason for choosing acquittal was insufficient evidence to prove the case. This reason accounted for just over three-quarters of the decisions to acquit (76.82 per cent, $n = 232$). Jurors responded by saying they required stronger forms of evidence (such as DNA evidence, or more concrete memory for details from the complainant) or that they could not

conclude that the evidence established the guilt of the defendant beyond reasonable doubt. Representative examples of responses in this category included:

“There was reasonable doubt about whether the indecent act occurred and where/when it occurred.”²²³

“Weak case did not provide evidence to prove guilt.”²²⁴

The second most prevalent reason to acquit was a lack of supporting evidence or the need for additional witness testimony (12.91 per cent, $n = 39$). Responses in this category emphasised the need for supporting evidence from other potential witnesses or family members, such as the victims’ parents. In the absence of this evidence, the oral evidence was unpersuasive – it was the word of the complainant against the word of the defendant. Representative examples of responses in this category included:

“Need more evidence and need mother to testify too – about Tim’s statement of sexual abuse.”²²⁵

“Not enough evidence presented, big gaps for witnesses (mum, other families).”²²⁶

Table 12 shows the proportion of reasons supporting a decision to acquit the defendant, in each themed category by type of trial.

Table 12. Reasons for mock jurors’ decisions to acquit by type of trial (per cent)

Decision-making factor	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial	Total
Evidence insufficient to prove case	83.75	71.69	83.72	76.92	76.82
Lack of supporting evidence/witnesses	11.25	17.46	2.33	0.00	12.91
Weak evidence/witness credibility	2.50	7.83	9.30	7.69	6.62
Credible testimony of the accused	1.25	1.81	2.33	0.00	1.66
Defendant’s behaviour counterintuitive	1.25	1.20	2.33	0.00	1.32
Consistency of multiple witnesses	0.00	0.00	0.00	7.69	0.33

²²³ Juror on Jury 4, Trial 2, separate trial with relationship evidence and standard jury directions.

²²⁴ Juror on Jury 44, Trial 1, basic separate trial with standard jury directions.

²²⁵ Juror on Jury 15, Trial 2, separate trial with relationship evidence and standard jury directions.

²²⁶ Juror on Jury 32, Trial 1, basic separate trial with standard jury directions.

Complainant's behaviour counterintuitive	0.00	0.00	0.00	7.69	0.33
Total ($n = 302$)					100

While a lack of supporting evidence was the main reason for acquittals in the separate and relationship evidence trials (separate trials 11.25 per cent, $n = 9$; relationship evidence trials 17.46 per cent, $n = 29$), the credibility of witnesses was a more important criterion in the tendency evidence and joint trials. In the tendency evidence trials, 7.69 per cent of participants ($n = 1$) indicated that witnesses' evidence and credibility was important, as did 9.3 per cent of participants in joint trials ($n = 4$). Trials with fewer witnesses and fewer counts against the accused were more likely to be seen as lacking sufficient evidence to convict.

Figure 6 shows mock jurors' reasons for their decisions to acquit the defendant, by type of trial.

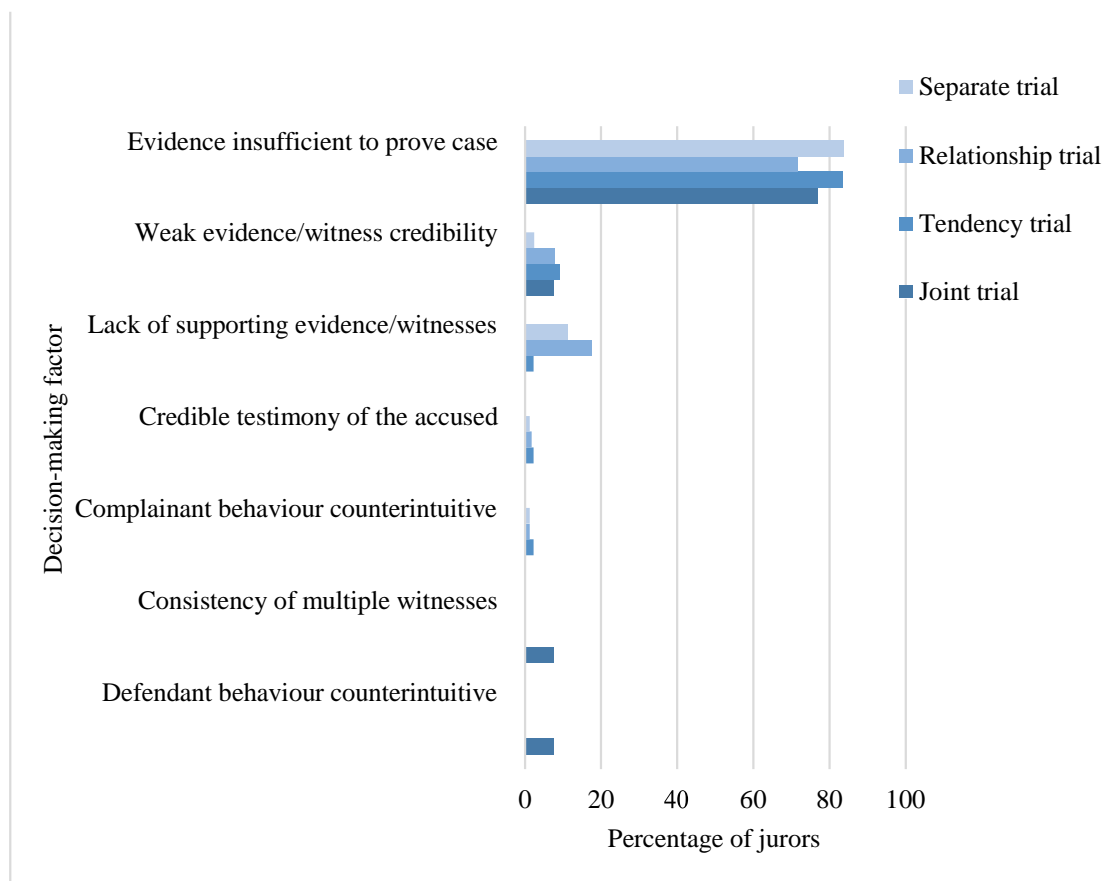


Figure 6. Key reasons for acquittal, by trial type (per cent)

Mock jurors' reasons in hung juries

In the hung juries, a total of 176 reasons were codable and included in an analysis of verdicts for the counts arising from the focal complainant's evidence. These reasons were distributed across the four trial types: 13 participants from separate trials, 108 from relationship evidence trials, 43 from tendency evidence trials and 12 from joint trials.

Across all four trial types, the most frequently cited reason for the verdict of a mock juror on a hung jury was insufficient evidence to prove the case, accounting for one-third of the decisions in this category (33.52 per cent, $n = 60$). Representative examples in this category included:

"Lack of compelling, uncontested evidence."²²⁷

"Not enough evidence to convince me of guilt."²²⁸

The second most prevalent reason – from jurors who had voted guilty – was strong evidence or high credibility of the witnesses (16.76 per cent, $n = 30$). Some representative examples of responses in this category included:

"The evidence provided was solid in my verdict he was proven guilty."²²⁹

"Evidence by victim with memories of the events after many years."²³⁰

Table 13 shows the major reasons cited for the individual verdicts of mock jurors who participated in hung juries, by type of trial.

²²⁷ Juror on Jury 19, Trial 5, separate trial with tendency evidence and standard jury directions.

²²⁸ Juror on Jury 91, Trial 4, separate trial with relationship evidence, standard and context evidence jury directions, and a question trail.

²²⁹ Juror on Jury 49, Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

²³⁰ Juror on Jury 29, Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

Table 13. Reasons for mock juror decisions on hung juries, by type of trial (per cent)

Decision-making factor	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial	Total
Evidence insufficient to prove case	53.85	39.81	13.64	28.57	33.52
Strong evidence/witness credibility	23.08	14.81	25.00	0.00	16.76
Consistency of multiple witnesses	0.00	7.41	25.00	28.57	12.85
Weak evidence/witness credibility	15.38	9.26	4.55	7.14	8.38
Pattern of grooming behaviour	7.69	6.48	15.91	0.00	8.38
Lack of supporting evidence/witnesses	0.00	9.26	2.27	7.14	6.70
Weak testimony of the accused	0.00	8.33	0.00	0.00	5.03
Character prejudice	0.00	0.93	4.55	14.29	2.79
Credible testimony of the accused	0.00	0.93	4.55	0.00	1.68
Complainant behaviour counterintuitive	0.00	0.93	0.00	7.14	1.12
Tendency evidence	0.00	0.00	2.27	7.14	1.12
Defendant behaviour counterintuitive	0.00	1.85	0.00	0.00	1.12
Accumulation prejudice	0.00	0.00	2.27	0.00	0.56
Total (<i>n</i> = 176)					100

The frequency of reasons for hung juries in each category varied across the four trial types. In separate trials, the most prevalent reason was the lack of persuasive evidence to prove the case (53.85 per cent, *n* = 7). This was true also for hung juries in relationship evidence trials, where two-fifths of the responses (39.81 per cent, *n* = 43) cited insufficient evidence. For the tendency evidence and joint trials, the consistency of facts across multiple witness accounts emerged as an important factor in juror decision making. Consistency across multiple witnesses' accounts was the most highly cited reason leading to a hung jury in tendency evidence trials (25.58 per cent, *n* = 11), presumably among individual jurors who voted to convict when other members of the jury did not. An equivalent proportion of hung jury members in joint trials cited insufficient evidence to prove the case (when they favoured acquittal) and consistency across multiple witnesses as key factors in their decision (28.57 per cent, *n* = 4) when they favoured conviction. Differences by trial type were similar to the reasons given for choosing acquittal.

Trials with fewer witnesses and counts were more likely to be seen as lacking sufficient evidence to prove the case, whereas in trials with more witnesses or tendency evidence, mock jurors were more concerned with examining consistencies across different witness accounts.

Figure 7 shows reasons for the decisions of mock jurors in hung juries, according to trial type.

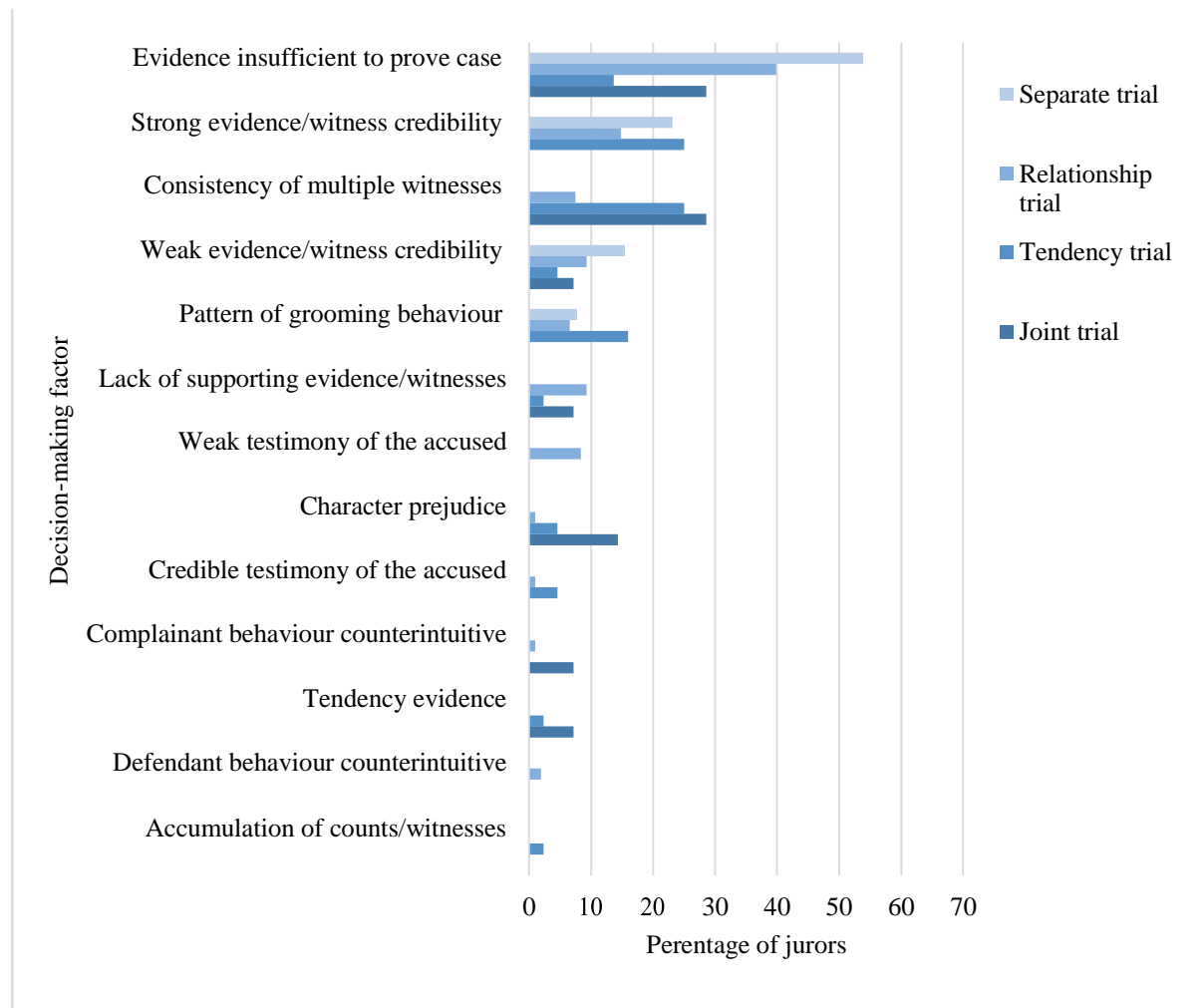


Figure 7. Key reasons for verdicts by members of hung juries, by trial type (per cent)

In sum, the reasons cited by individual mock jurors in support of the verdicts they reached provided negligible support for the proposition that conviction rates in joint trials are the consequence of impermissible reasoning due to accumulation prejudice and character prejudice. The overall rate of all instances of possibly impermissible reasoning in these categories – across all coded responses for all types of trials – was 2.16 per cent (separate trials = 0.00 per cent; relationship evidence trials = 0.60 per cent; tendency evidence trials = 2.66 per cent; joint trials = 4.01 per cent). Caution is advised in interpreting these results as in coding the verbal responses, the researchers erred on the side of over-counting examples of impermissible reasoning where the language used or expression was ambiguous.

4.3.4 Qualitative review of deliberations in separate trials

Below, we describe differences in reasoning and decision making observed in the separate trials – those that included relationship evidence and tendency evidence and those that did not. We have included excerpts of the jury deliberations to illustrate their reasoning.

Basic separate trial²³¹

Mock jurors' reasoning in the basic separate trial did not vary greatly in terms of the focus of their deliberations. As expected from the study design – which involved presenting juries in the separate trial with a moderately strong case with some unresolvable ambiguities (Trial 1) – the juries found that the 'story' of what happened to Timothy was incomplete. This resulted in a considerable amount of speculation as mock jurors struggled with their doubts about whether the defendant was guilty or not.

For example, in Jury 32²³², doubts arose because of the inconsistencies between the evidence from Timothy and the defendant about whether Timothy had run away from the defendant's home (Timothy's evidence), or whether the defendant took Timothy to visit his hospitalised mother (the defendant's evidence). Mock jurors questioned why Timothy's mother did not give evidence, why other families did not give evidence about the defendant's behaviour, and why there were no hospital records of Timothy staying overnight with his mother. They speculated about why Timothy had come forward 16 years later; what type of relationship he had with his mother, whom he had not seen for some years; and why the defendant had come back into Timothy's mother's life as her partner. Even though the defendant was not considered a credible witness (he was described as 'dismissive', 'arrogant', 'dodgy' or 'too calm'), mock jurors kept coming back to the insufficiency of evidence and the lack of 'a clincher'.

With their many questions unanswered, Juror 7 in Jury 32 recognised:

“... they have given us the perfect mock trial ... The 'what ifs', and 'what could have happened' and 'why didn't they' – it's the whole reason they have been left out.”

Similar questions and speculation were evident in Juries 33²³³, 45²³⁴, 46,²³⁵ and 71²³⁶, particularly about why Timothy's mother, whom they considered to be a key witness, had not

²³¹ Trial 1, basic separate trial with standard jury directions.

²³² Ibid.

²³³ Trial 1, basic separate trial with standard jury directions.

²³⁴ Ibid.

²³⁵ Id.

²³⁶ Id.

been called to give evidence. The evidence given by Timothy's supporting witness, Ellen Samuels, supported Timothy's account of being sexually abused in a bedroom with a TV, but did not assist these juries in resolving their doubts; the juries either dismissed this evidence as irrelevant or did not discuss it at all.

Separate trial with relationship evidence²³⁷

Generally speaking, the relationship evidence raised more questions than answers, and did not, on the face of the deliberations, appear to influence voting patterns. This was the case even though the conviction rates for Counts 1 and 2 were significantly higher than those in the juries that did not receive relationship evidence (Trial 1).

Jury 48²³⁸ provided the best example of the influence of the relationship evidence, not in terms of resolving doubts about Timothy's evidence, but in terms of providing the background or context for Timothy's allegations (the purpose of admitting the evidence). At first, the relationship evidence raised questions that could not be answered, in particular the likelihood that the defendant would have been grooming Timothy in such a public place: the equipment room at the soccer oval. With speculation that several teams would have been playing soccer on the same day as Timothy's team, it raised the possibility that anyone could walk into the equipment room at any time.

For most of their deliberation, the mock jurors in Jury 48 focused on their many doubts and questions, none of which were related to the relationship evidence. They wondered: Did Timothy ever visit Ellen Samuels's house, where it appears he was sexually abused? Did the defendant have the keys to that house? Why did Timothy's mother not give evidence? Why did Timothy delay his complaint? What movie did Timothy watch on the night of the alleged abuse? Did the defendant give truthful evidence? Why did the defendant have no further contact with Timothy and his mother after the alleged abuse? Only after discussing these issues did the jury consider the likelihood of the defendant grooming Timothy in the equipment room, and come up with their own explanation of how it could have occurred. In other words, the relationship evidence appears to have been influential because the jury was able to turn it into a plausible scenario:

JUROR 1: ... the thing is, he's a very popular coach and ... there were other activities going on with other boys at the same time, he may well have had a situation all set up nicely for

²³⁷ Trial 3, separate trial with relationship evidence, standard and context evidence jury directions.

²³⁸ Ibid.

the use of that room and the boys didn't come in when he was with person X, person Y.

So we don't know -----

(All talk at once) 58:25.

JUROR 6: There's no evidence -----

JUROR 1: No, no, but I'm just saying it is not germane to our consideration, but I'm saying these scenarios do happen. And the influence of the adult is such that they can control the flow of activity.

JUROR 2: It could be end of training. All the other kids are picked up.

JUROR 7: He did say it was later.

JUROR 2: Later, yeah.

(All talk at once) 58:43.

JUROR 3: That's what I was thinking, he's got [to] lock to the door.

[...]

JUROR 1: And it's understood "don't go in".

JUROR 3: And all the other people thought, "Oh, that's already over. Everybody's gone home and stuff". And while they are inside, nobody can hear them.

JUROR 9: Exactly.

JUROR 3: And the small little boy didn't have a clue what was going on because the adult was yeah.

JUROR 11: Just having an innocent time while -----

JUROR 5: Most sheds are also quite remote.

Despite the apparent plausibility of the relationship evidence, when a vote was taken after an hour of deliberation, only 10 out of 12 mock jurors voted guilty. While the two dissenting jurors said there was not enough irrefutable evidence to make a finding of guilt, the other 10 jurors did not use the relationship evidence to try and persuade the two dissenting jurors to change their vote, which in the end they did not.

This type of analysis in Jury 48 was not evident in other juries that received relationship evidence. In a discussion in Jury 57²³⁹ about the relationship evidence that the defendant had taken naked photographs of Timothy, Juror 5 wanted to know where the photographs were. He was convinced that a paedophile would never destroy them because they 'thrive' on such

²³⁹ Trial 7, joint trial with standard and tendency evidence jury directions.

things. Only one mock juror in Jury 57 stated explicitly that they were going to take into account the defendant's grooming behaviours and 'the context of how these acts were performed'. While the three jurors in Jury 57 who voted not guilty were eventually persuaded to change their verdict, they did not use the relationship evidence to resolve their doubts.

Similarly, in Jury 25²⁴⁰, the relationship evidence raised several questions: Did the defendant have a camera? Where was the evidence of the photos? Why did Timothy get into the defendant's bed if he thought modelling naked for the defendant in the equipment room had made him uncomfortable? No mock jurors explicitly referred to the relationship evidence as the reason for their decision to vote guilty. When one mock juror dissented, the remaining jurors eventually persuaded her to change her vote, although none of their arguments were based on the relationship evidence given by Timothy.

In Jury 49²⁴¹, the relationship evidence rang true for several mock jurors because it demonstrated that the defendant had groomed Timothy, although it was rarely referred to in their deliberation. Other jurors questioned the veracity of Timothy's relationship evidence in terms of the days on which the grooming occurred, whether other people were around and whether the door to the equipment room could have been locked. Juror 2, who had been a photo finisher in the 1990s, found the existence of the photos hard to believe because in the 1990s someone would have had to process the film, so 'iffy' photos would have been reported to the police.

Tendency evidence trial²⁴²

Although the conviction rates for Counts 1 and 2 were significantly higher in the juries that were shown tendency evidence compared to those in the separate trial that did not receive tendency evidence (Trial 1), none of the juries in the tendency evidence trial (Trial 6) engaged in impermissible propensity reasoning. For example, in Jury 27²⁴³, the mock jurors carefully evaluated the evidence of the tendency witnesses, especially Simon, who presented the weakest evidence compared to Timothy (the complainant) and Justin (the other tendency witness). Mock jurors believed that the pool – into which Simon said he had been pushed by the defendant – did not exist, and found it hard to believe that Simon had run to his mother's car in a t-shirt and no underwear. Mock jurors thought that fact would be a red flag for his mother,

²⁴⁰ Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

²⁴¹ Ibid.

²⁴² Trial 6, separate trial with tendency evidence, and standard and tendency evidence jury directions.

²⁴³ Ibid.

and wanted to know why she did not do something at the time. They also evaluated the credibility of Timothy, Simon and Justin separately. In fact, two mock jurors thought they should ignore Simon's evidence completely because of the doubts it raised. Several mock jurors recognised that the evidence of Simon, Timothy and Justin revealed a pattern of behaviour on the part of the defendant, but it was not sufficient to remove the doubts of the two jurors who voted not guilty in Jury 27.

Jury 30²⁴⁴ was dissatisfied with the lack of supporting evidence for each complainant's story, and failed to consider that each complainant's evidence could be used to support that of the other complainants. The mock jurors ignored the judge's instructions about tendency evidence as well as the instructions about the need to consider the elements of each count separately, which may have affected their decision to vote not guilty on both counts. In Jury 28²⁴⁵, most mock jurors were unsure about what amounted to tendency evidence and therefore ignored the prosecution's argument about the defendant's tendency. They voted guilty on both counts by focusing only on Timothy's evidence pertaining to the two counts.

Jury 29²⁴⁶ ignored the tendency evidence admitted in the trial, as they debated for the entire deliberation (31 minutes): whether a child of 11 years (Timothy) would or would not remember every single detail of what happened when he was sexually abused (such as the movie he watched, what house he was in); whether Ellen Samuels had or had not given the defendant a set of keys to her townhouse; why Timothy had decided to come forward so many years later; and why his mother had not believed him when he was a child. When they cast their first vote (12:1, guilty), they had only mentioned the tendency evidence once in 30 minutes of deliberation, and they completely ignored it when discussing why they had no reasonable doubt. The other mock jurors did not refer to the tendency evidence when they tried to persuade the dissenting juror to change his vote. The dissenting juror maintained his position, saying that he did not believe Timothy's evidence.

²⁴⁴ Id.

²⁴⁵ Id.

²⁴⁶ Id.

4.3.5 Thematic analysis of deliberations about the focal complainant in joint trials

This section presents a qualitative analysis of jury deliberations for the 33 joint trials (trials 7 to 10). The analysis focuses on Count 5 (indecent assault) and Count 6 (sexual intercourse) involving Timothy, the complainant with the moderately strong case.

The deliberations were differentiated by outcome, so we begin with deliberations resulting in a conviction, followed by those resulting in either a hung jury or an acquittal. The analysis is grounded within a framework of jury sensemaking (see Section 2.4).

Juries who convicted in relation to the moderately strong claims

For juries that convicted the defendant, the key issues for analysis were whether the convictions were the result of prejudice based on impermissible reasoning (factual confusion, accumulation or character bias) or on inferences logically related to the evidence presented in the trial.

Awareness of discrepant messages and evidence

In assessing Timothy's testimony and the alleged events surrounding his victimisation, the primary item of evidence that caused discussion among the mock jurors was the discrepancy in the title of the movie that Timothy said he watched with the defendant at the time the offences occurred. The trial presented inconsistent evidence on this point, and the defence suggested that Timothy's inability to be sure about the movie title was indicative of his inability to recall events more generally.

Some mock jurors inferred that Timothy's credibility was questionable based on his memory of the movie. For example, in Jury 85²⁴⁷ one mock juror said that Timothy should be able to remember the movie if the assault had actually occurred:

JUROR 6: But in his first statement to the police, he referred to the movie *Hercules*, which is a completely different movie genre. And I would think something that is disturbing as being sexually assaulted and there was a movie on, you would remember which movie it was ----

(Most jurors voiced disagreement).

JUROR 9: I don't agree with that.

²⁴⁷ Trial 7, joint trial with standard and tendency evidence jury directions.

When the uncertainty in the movie title was used to infer Timothy's lack of credibility as a witness, other mock jurors disagreed and defended the inconsistency in Timothy's evidence, as seen in the following excerpt from Jury 39:²⁴⁸

JUROR 3: Sorry, can I go on the last one for me and this gentleman over here, where we have got 'not guilty'? The reason I sort of believe that was, one, there were so many holes in Timothy's story. He was, I have noted here the movie wasn't out. He was changing. He added in the second residence. It was so made up. It was like, 'Oh, wait, no'. It was almost like he backtracked.

JUROR 7: It's 20 years ago and he was a kid.

(Most jurors voiced agreement).

...

JUROR 12: I just think it is quite difficult for a child and admittedly the first fellow was he did change. You know, he wasn't sure about the movie and all that, but that is quite normal. If you asked me what happened 20 years ago, even if it was a frightening thing, the details wouldn't be particularly good.

(Most jurors voiced agreement).

Later the discussion continued:

JUROR 10: In the judge's summary, they discuss the fact that the court will accept that evidence will change, the longer that the case before it is brought to court. And it is okay for us to accept that their memory because [of their] age may not be that good.

...

JUROR 12: The thing about the evidence, too, some of the dates were fuzzy. Some of the like, that film. But the description of what happened to him was clear as a bell.

...

JUROR 8: Does it matter what movie he watched?

JUROR 6: No -----

For the most part, however, mock jurors were forgiving of Timothy's inconsistencies about the title of the movie, seeing it as a function of both the passage of time and the inability of any child to remember such specific details. For example, in Jury 63²⁴⁹ one person noted:

²⁴⁸ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

²⁴⁹ Trial 7, joint trial with standard and tendency evidence jury directions.

JUROR 4: But as a child, you would watch lots of movies; you are confused. You would not know which movie you are -----

And in Jury 83²⁵⁰, a similar justification was made:

JUROR 2: I mean, there is the discrepancy of the film but that doesn't cause so much of an issue for me.

...

JUROR 9: I mean, you could mistake the video -----

(Most jurors voiced agreement).

JUROR 9: ----- it is that many years ago.

Jury 62²⁵¹ followed a similar line of reasoning:

JUROR 2: Don't have issues with the movie; you know, two different movies -----

(Most jurors voiced agreement).

JUROR 5: I don't remember a movie that I watched last week. I don't think a child -----

JUROR 8: I can't remember a movie that I watched last week.

JUROR 9: I don't think any of us could remember a movie that we watched back when we were 10, like, on an exact date.

Most mock jurors forgave the inconsistencies in Timothy's evidence, such that they became irrelevant to Timothy's credibility and the question of guilt. Interestingly, this was not so in the case of Simon's testimony, where the question of whether the defendant's townhouse had a pool or not was seen as central to Simon's credibility and the determination of guilt. The issue of the perceived credibility of the witnesses is discussed further below.

In general, the juries were quite accurate in their discussions of the evidence. As discussed above, while there was a level of confusion in some of the juries – primarily in terms of which incidents were associated with which complainants – the deliberative process of discussion, questioning and clarification nearly always resulted in an accurate recall of the evidence.

Drawing on common experience and knowledge to resolve disparities

To assist in their understanding of the evidence and in order to resolve disparities, mock jurors often drew upon their own experiences and knowledge. This was particularly the case when

²⁵⁰ Ibid.

²⁵¹ Trial 10, joint trial with standard jury directions.

attempting to understand how the defendant came to be a soccer coach and in a position to abuse the boys. For example, Jury 75²⁵² discussed possible reasons for the defendant to have befriended the boys in such a close, personal way:

JUROR 10: So it could have been anywhere within that period. But a child to recall that as an adult, I think looking at what the defence provided to ask us to make it beyond reasonable doubt, there weren't any there wasn't anything else besides that. And there wasn't any explanation given by Mr Booth or by his defence as to why he befriended these children, besides him just being a nice person. There wasn't any other reason. Now, I used to coach rugby because I used to be a teacher; and I never, you know, there were lots of children. I did it for three years. And there was no time, just from my personal experience, although the children were all different that I had any reason to take them or become involved with them in a personal situation. And there was no reasonable explanation for why he did that, to those particular three boys.

...

JUROR 10: As a teacher, I remember I used to pat boys on the – I taught a Christian Brothers Boy school and I used to pat them on the head and I had to stop doing that. So I just can't pat or touch them anywhere.

Similarly, Jury 39²⁵³ discussed experiences of soccer teams to question why a man without children in the team would be the coach, as a way of understanding the defendant's opportunity to commit the offences:

JUROR 5: So of under 12 ... generally, my kids have been involved in sports all their life, and there's always a parent on the team, coaching the team; until they get older and then they need certain skills. But at that age, there's always some parents there or there's someone else there.

Jury 69²⁵⁴ also noted the presence of parents as a way of trying to make sense of the defendant's role as coach:

JUROR 3: Our son has someone coaching him, an older man coaching him this year, who has no child in the team; but he was, uhm, obtained through the club, by someone – I think he's related to someone who is running the club. And that was because something else fell through. So it was very fortunate that he's very good. But he doesn't take the kids to

²⁵² Trial 7, joint trial with standard and tendency evidence jury directions.

²⁵³ Trial 10, joint trial with standard jury directions.

²⁵⁴ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

McDonald's after the game and isolate himself with the children. There's always the parents around. These days, this is why the rules are stricter -----

JUROR 9: Absolutely.

JUROR 3: ----- "You must have a parent at soccer. You must have a parent at training."

At times, personal experiences of traumatic events were used to understand the inconsistencies in Timothy's evidence regarding the title of the movie. For example, two mock jurors in Jury 38²⁵⁵ noted:

JUROR 1: I had something happen to me at that age as well. Even I don't remember much; because it is just something that you block out. And there's only very small aspects of it that I remember.

...

JUROR 4: I had a pretty traumatic event happen to me at age 13 and I cannot remember, for the life of me, a whole month period.

JUROR 9: Okay. There you go. It is the age. Maybe when you are older you retain more.

Similarly, in Jury 62:²⁵⁶

JUROR 8: Okay. Well, a recent traumatic event I had with my mum passing away and I sat in the lounge room all night watching TV and I have no idea what I watched. But I remember my mum passing away. So, you know, they can remember a traumatic event but not necessarily -----

JUROR 9: The minor details.

In some juries, people's knowledge of other criminal cases was used to assist in understanding the nature of the sexual offending, as in Jury 83:²⁵⁷

JUROR 6: Yeah, so you start to get away with it, "Gees, I can up it". And then, also, I did this thing when I was studying about a serial killer, John Wayne Gacy. And his taste in killing boys got as he got away with it, he got older the boys were younger and they started getting older; and the crimes are more gruesome. This is kind of a similar thing where he got away with one sexual act and it was quite minimal; and then it got worse and worse and bigger and bigger. And that's why, kind of, had the verdict of 'guilty' for the rest of them, because it just seems his taste got worse as he got away with it.

²⁵⁵ Trial 10, joint trial with standard jury directions.

²⁵⁶ Ibid.

²⁵⁷ Trial 7, joint trial with standard and tendency evidence jury directions.

Mock jurors also used their experiences and knowledge to help make sense of the evidence about whether the defendant had been given a set of keys to Ellen Samuels's townhouse. Timothy had testified that he had been abused in the defendant's 'other house', so the question of the defendant's access to Ellen Samuels's townhouse was crucial because Timothy had remembered a television on top of a chest of drawers. This description matched the main bedroom in Ellen Samuels's home. While the defendant claimed that he did not have a set of keys to her townhouse, a number of juries did not accept this version of events. For example, in Jury 77²⁵⁸ one person noted:

JUROR 12: But, see, Ellen said that she often gave keys to Mr Booth; whereas when Mr Booth was questioned, he said that he's never had the keys.

JUROR 8: He said he put the letters under the door.

JUROR 12: Yeah, I think that area should have been challenged.

JUROR 6: Yes, that's right. Because I collect a lot of mails for neighbours and I can tell you most of the time you can't get under the door.

Overall, people's common experience and knowledge was brought to bear to help attribute meaning to the evidence, especially around resolving inconsistencies and understanding motive and opportunity.

Testing the plausibility of the competing narratives

The key issue for the juries appeared to be whether the complainants and the defendant were credible. Juries assessed credibility based on a number of factors, including the complainants' motivation to report to the police, the accuracy of their recall of events, and the presence or absence of supporting evidence.

In Simon's case, the most commonly invoked issues raising suspicion about the veracity of his evidence were the question of financial gain in the form of victim's compensation and the lack of evidence about the existence of a pool at the defendant's property. By contrast, juries realised that Justin's allegations were supported by his school friend's and his mother's evidence of a change in his behaviour after the alleged abuse.

Mock jurors appeared to have conflicting views about Timothy's credibility. On the one hand, some mock jurors were initially skeptical about his inability to recall the title of the movie that he watched at the time he was allegedly abused. On the other hand, the supporting evidence

²⁵⁸ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

from Ellen Samuels about the description of her bedroom at the time supported Timothy's story. The following excerpt from Jury 69²⁵⁹ illustrates this:

JUROR (?) 3: Does anybody think that it is not true just because -----

JUROR 2: No, because the witness said it, yeah?

(Most jurors voiced agreement).

JUROR 6: She would have given him the spare set of keys, because he came into her house to put her mail in.

JUROR 13: And why would she lie?

(All talk at once) 42:20.

JUROR 10: She had no reason to lie.

JUROR 7: Doesn't have a motive, either, and the fact that he didn't vary when they are all trying to press him; like the fact that he didn't, in fact, have a chest drawer or a TV in his bedroom. And he was -----

JUROR 2: He was confident about it.

JUROR 7: Yeah, and the fact that he was so confident about it, it's like he definitely remembers that. And there's proof because we have the lady -----

JUROR 10: The witness. And I think the different movie doesn't make any -----

(Most jurors voiced agreement).

JUROR 6: I would say it is a time lapse.

JUROR 7: Definitely-----

Jury 70²⁶⁰ also illustrates the importance of Ellen Samuels's evidence in strengthening Timothy's credibility:

JUROR 1: I thought the naming of the movie was irrelevant -----

(Most jurors voiced agreement).

[...]

JUROR 4: Because that's the whole thing, discredit -----

(All talk at once) 20:33.

JUROR 6: It was good that they had the -----

²⁵⁹ Ibid.

²⁶⁰ Ibid.

JUROR 9: The lady that came in.

JUROR 6: They said “There was a TV...”

Jury 77²⁶¹ used similar reasoning:

JUROR 11: That there was a TV on top of a cabinet which is quite specific to a bedroom.

(Most jurors voiced agreement).

JUROR 11: And it matched.

(Most jurors voiced agreement).

JUROR 11: So that, for me, just threw it all out. “Okay, that really happened” -----

(Most jurors voiced agreement).

JUROR 11: ----- “Thank God they brought in that credible witness.”

JUROR 2: Yeah.

JUROR 6: There’s no motive/angle either way.

JUROR 8: That’s why they brought her in as a witness, because it would strengthen the prosecution.

JUROR 4: It did.

JUROR 2: Which it did.

Similarly, Jury 62²⁶² placed significant weight on the supporting evidence of Ellen Samuels:

JUROR 9: And the whole TV thing, like, what’s happened -----

JUROR 3: Yeah, that was unbelievable, until she made it believable and then it was like, “Oh, that corroborates that”.

For Jury 61²⁶³, Timothy’s unusual behavior following the alleged abuse was enough to convince mock jurors that his evidence was credible:

JUROR 9: Timothy was the ‘finger in the bum’ boy. Again, I found his evidence credible. I found the circumstances surrounding his evidence, in that he didn’t go back anymore, didn’t see him anymore, credible -----

...

JUROR 9: The fact that he was supposed to stay three days at the coach’s house and he only stayed one night and then ran away -----

²⁶¹ Id.

²⁶² Trial 10, joint trial with standard jury directions.

²⁶³ Ibid.

JUROR 8: That's right, yeah.

JUROR 9: ----- and didn't go back again, it really -----

JUROR 8: To an empty house.

JUROR 9: That's right, with mum in hospital.

For some mock jurors, the presence of multiple independent victims also seemed to lend credibility to the claims by the three complainants in the joint trial. For example, in Jury 64:²⁶⁴

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

JUROR 1: I just think he would have no motivational 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.

Similarly, Jury 51²⁶⁵ found that the most convincing evidence of guilt was that there was more than one complainant, and that there were similarities between the allegations. One juror believed that "you don't get three people coming out with similar stories if there isn't something there". The fact that the victims did not know each other and had no prior involvement with each other also bolstered the jury's opinion of guilt. One juror explained that "the first one (Simon) seemed to be a little bit dodgy, but after, the second and third ones made me believe":

JUROR 5: I think one corroborated the other.

...

JUROR 5: You are not going to get three people that will come out and accuse a person like that without there being some basis to it.

...

JUROR 3: The pattern, the pattern of behaviour. His MO is virtually the same with all three, yeah.

...

JUROR 11: The other connection that was consistent between all three boys is that he was stepping in as a father figure -----

JUROR 3: As a father, yeah.

²⁶⁴ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁶⁵ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

JUROR 11: ----- as a father. So he's obviously preying on that -----

JUROR 3: He found a weakness that he can explain, yeah -----

JUROR 11: ----- in a vulnerable family situation. So he's obviously fulfilling that role of-----

JUROR 3: Yep.

Jury 62²⁶⁶ used a similar line of reasoning, placing substantial weight on the presence of multiple complainants:

JUROR 1: I was just going to make some general statements about what we all just listened to, and that is that overall, there were three separate complaints. These guys didn't know each other. So I find that very compelling against the accused because they did not collaborate; there was no reason to collaborate or collusion. And their motivations were basically individually based, and their own individual experience. While the circumstances that I have heard might be scrambled, you have got to understand that this was 15 years earlier; and while whether he saw a movie look, at the end of the day, these were kids.

...

JUROR 3: Absolutely. The three together makes it very, very concrete -----

Ultimately, Timothy's evidence appeared to be given substantial weight with his credibility enhanced by the supporting evidence given by Ellen Samuels and the other complainants so that the inconsistencies in his evidence about the movie title appear to have been dismissed.

Jurors' communication about their assessments and evaluations

For the most part, juries appeared to be respectful and helpful in their deliberations, working together to construct a narrative upon which they could mostly agree. The average duration of each deliberation for the juries who convicted was 34 minutes. Deliberation generally moved systematically through the various counts, notwithstanding that some juries spent time on correcting factual confusion when it arose.

Reaching a verdict

The juries appeared to take the judicial instructions very seriously and worked assiduously to apply them as best they could. Discussions of the meaning of 'beyond reasonable doubt' took place in a number of juries in relation to the evidence of all three complainants, as mock jurors attempted to clarify how much doubt could be considered 'reasonable'. Ultimately, juries

²⁶⁶ Trial 10, joint trial with standard jury directions.

appeared to apply a common-sense understanding of the concept, without trying too hard to quantify the level of certainty required for conviction.

On the other hand, the use of tendency evidence – and interpretations of the tendency evidence direction – appeared to be more difficult for juries, particularly in relation to Simon, who was the first complainant to give evidence in the joint trial, as illustrated by Jury 83:²⁶⁷

JUROR 8: Which is the tendency evidence for the first one? There seems to be a pattern, the sexual interest in boys under 12; engagement in sexual activities and using a position of authority as a soccer coach to gain access. If you insert tendency evidence in the first one – I'm not sure how that works, actually.

...

JUROR 5: Yeah, I see what you mean, you can't use the tendency because it is the first one. It's tricky.

Jury 54²⁶⁸ experienced similar difficulties, also struggling to interpret the judge's directions about the use of tendency evidence:

JUROR 9: I think he's guilty on this count.

JUROR 3: I think the few discrepancies bother the case.

JUROR 9: Can we go to tendency on this, too?

JUROR 3: Yeah, we can.

JUROR 1: Can we go to tendency?

JUROR 9: If that is the case, then I have got no doubt.

JUROR 3: Because we have got full conviction of the other case; so we can go to tendency.

...

JUROR 6: I understand what they are saying in tendency but I don't think that we can bring up the past ones as evidence ... that is why tendency is a really vague term. The fact that she let us use it is interesting.

(All talk at once) 39:41

JUROR 8: Because that is previous records. You can't bring up previous records but this is the same case.

JUROR 1: Yes, that's right.

²⁶⁷ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁶⁸ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

(All talk at once) 39:46.

JUROR 11: It goes towards the pattern of behaviour. That is what determines the evidence.

(Most jurors voiced agreement).

Jury 68²⁶⁹ also found the application of the tendency evidence directions to be complicated:

JUROR 4: But you have got two kids there, that you believe their stories; and one who doesn't. So that means that the guy is still guilty -----

JUROR 12: Yeah, of course.

JUROR 4: ----- so, therefore, there is a 50 per cent probability that he did on the first case as well -----

JUROR 11: No.

JUROR 5: I think you use tendency evidence, if he's done two crimes -----

JUROR 11: No, that shouldn't come into existence.

...

JUROR 9: So I think we are kind of crossing over. I agree with you, if the first one was by itself with nothing else, I don't think there's enough; but as the judge said, you can infer from the other cases. It's not – yeah, while there are a few inconsistencies with the story, it's not completely different to what happened to the other two. So it is not unreasonable to think that it could have happened. You are never going to know 100 per cent anyway ----- Uhm -----

Ultimately, Jury 68 convicted on Counts 2-6, but was hung on Count 1, as they were uncertain about applying the tendency evidence to the first count.

Despite difficulties in its application, Jury 52²⁷⁰ noted the importance of tendency evidence in these sorts of cases where substantial time has passed:

JUROR 1: 20 years on, it is really one person's word against another. And that's why this pattern thing becomes important -----

JUROR 5: It is the pattern.

JUROR 1: ----- and that's why credibility becomes important. You have to make a decision on that basis.

²⁶⁹ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

²⁷⁰ Trial 8, joint trial with standard and tendency evidence jury directions and a question trail.

However, as Jury 53²⁷¹ noted, just because there appears to be a pattern, it does not necessarily mean that every charge is proved:

JUROR 11: Serial of things, much the same thing, for much the same reasons, at two-year intervals; it does suggest -----

JUROR 4: Yes, same sort of -----

JUROR 2(?): Same patterns. I thought there was also a consistency in the phrases used by Mark Booth. He said to two boys, “What’s down here?” or “What have you got down there?”

And

JUROR 12: He’s serial – he sort of has the pattern of a serial sex offender. He gets the trust of the parents -----

(Most jurors voiced agreement).

...

JUROR 4: Well, it all follows a pattern but because someone’s done one, it doesn’t mean that they have done any more -----

JUROR 12: I know.

JUROR 4: ---- it could be – one-off ----- But these three victims, they don’t know each other -----

(Most jurors voiced agreement).

JUROR 4: ----- and it is very similar. They all -----

JUROR 11: Yes, of the same operation; two-year intervals; much the same situation.

JUROR 4: That in itself -----

JUROR 11: Is very suspicious.

In Jury 39²⁷², one mock juror spent some time explaining the use of tendency evidence:

JUROR 10: Do I mind if I raise two things? First of all, part of what we are allowed to work with is tendency evidence, which means that we are allowed to look at him and go, “We have got three people who have got a complaint against him”. So there is a pattern forming.

JUROR 7: Exactly.

²⁷¹ Ibid.

²⁷² Id.

JUROR 10: If you have a think about the defence, they did not argue against or prove or disprove that there was any tendency there, the fact that there were three people there. He has specific evidence for each person to try and dispute it, but he never really came up with any reason as to “how come there’s three people that have got a complaint against this soccer coach?” So that’s my first point. So we are allowed to use the tendency evidence here. I still struggle now, I put that aside because I still struggle with the first count, which is where – and I’m not saying that there was a pool or wasn’t a pool. But my point is that the prosecution didn’t prove one way or the other whether there could have been a pool. So that’s why I have a doubt, where I can’t go “guilty” on that one.

In *Jury 41*²⁷³, the self-correcting nature of juries became evident, as one mock juror reminded another of the need to focus on the evidence:

JUROR 11: Do we believe he has a tendency to ... to that level?

JUROR 10: I do, yes.

JUROR 2: And that’s where we can use that.

JUROR 7: Yes, we have to use our gut feeling and what -----

JUROR 4: Using our gut feeling that -----

JUROR 10: I’m not using my gut feeling. I am using the evidence, corroborating, to actually form that he has a tendency; which is what we are allowed to do. There is no gut feeling.

Hung juries and juries who acquitted in relation to the moderately strong claims

For juries that acquitted or were hung, the key issue for analysis is whether their decisions were the result of inferences that were logically related to the evidence presented in the trial.

Awareness of inconsistent evidence

Two of the juries returned a conviction on Count 5 (masturbation, non-penetrative) and an acquittal on Count 6 (penetration). For *Jury 60*²⁷⁴, the main reason for the acquittal appeared to be a lack of credibility in relation to Timothy’s evidence:

JUROR 8: When he was cross-examined, he ended up saying, “I’m not sure”. So his final thing that we have seen him say is, “I don’t know. I’m not sure”. So now we are kind of in a position where, if we are going to take what he said, we don’t know what movie he’s

²⁷³ *Id.*

²⁷⁴ *Trial 10*, joint trial with standard jury directions.

seen. So if there's anything to get out of that, it's it brings into question his ability to recall what happened.

Testing the plausibility of the competing narratives

For Jury 60, a key consideration driving the acquittal on Count 6 (penetration) appeared to be a belief that the defendant wouldn't go 'that far' and take the 'extra step' of sexual intercourse:

JUROR 6: With all this, even though we agree that Mr Booth is a predator with indecent assault, with the mother being in hospital and the child being that upset, I couldn't see Mr Booth going that extra step to sexual intercourse ----

Although there was no actual evidence offered for the above belief, it was sufficiently strong for the jury to return a verdict of not guilty for the sexual intercourse charge.

Jurors' communication about their assessments and evaluation

In Jury 67²⁷⁵, which convicted on Count 5 (masturbation, non-penetrative) but was hung on Count 6 (penetration), some mock jurors found the evidence insufficient to establish guilt beyond a reasonable doubt:

JUROR 5: As I said, I have probably got – I believe he's probably done it but, again, the law ----

JUROR 11: Says 'beyond reasonable doubt'.

JUROR 5: No, no, but I believe he's done it ----

JUROR 6: But you feel there's not enough evidence.

JUROR 5: ---- but the evidence has to support that. Just what I feel and what can be supported by evidence, they are two different things.

JUROR 11: But they haven't shown us enough evidence, I think.

(All talk at once) 39:49

JUROR 11: 'Beyond reasonable doubt'; that is all it is.

JUROR 5: But, remember, you have got the evidence it is not what you feel; it is the evidence.

JUROR 6: But I feel we have got to make a decision on whatever evidence they have given us.

(Most jurors voiced agreement).

²⁷⁵ Trial 9, joint trial with four prosecution witnesses, standard and tendency evidence jury directions.

JUROR 1: It should be ‘not guilty’ with regards to what you are saying -----

JUROR 5: Exactly right. If you are looking at it from a strict perspective, and not an emotional perspective, or what you know, I think it’s highly possible that this guy’s done it. Does the evidence support that? In my view, I don’t think the evidence [does] support it.

JUROR 11: But does the evidence not support that?

JUROR 5: But guess what, he doesn’t have to -----

JUROR 11: It goes both ways.

JUROR 5: No, no, it doesn’t. It doesn’t. And that is the whole point. It is the Crown’s case to prove it beyond reasonable doubt. It is the Crown’s obligation. That’s the first opening statement that they make; it is the Crown’s obligation to put in enough evidence, to put it beyond reasonable doubt. And that’s where they say the law is an arse, sometimes. There is plenty of people that have done crimes and have got off on technicalities because of the evidence, or the chain of evidence is not right.

Reaching a verdict

Jury 60²⁷⁶ attempted to build a picture of the defendant and the type of offences that he typically committed. The jury accepted that he had committed the indecent assault against Timothy and convicted him, but did not believe that the alleged sexual intercourse took place as it was ‘out of character’:

JUROR 6: That is the only reason that – you know, a lot of this has been opportunistic for him, Mr Booth. I don’t think – what he’s doing is bad but we are talking about from ‘indecent assault’ to ‘sexual intercourse’. And we have already said “not guilty” on one lot of sexual intercourse. Again, I don’t think that -----

JUROR 5: “Not guilty.”

JUROR 6: ----- under those circumstances, he would go that far.

...

JUROR 8: It’s sort of out of character as well for him, with all other acts -----

JUROR 6: That’s what I think as well.

JUROR 8: ----- everything else is touching and -----

²⁷⁶ Trial 10, joint trial with standard jury directions.

JUROR 9: Yeah. But you have got to take into account that there were a lot of years in between this. You don't know how many other acts he's done -----

JUROR 3: Yeah.

JUROR 9: ----- to what happened at that stage.

...

JUROR 7: It doesn't really matter. Any sort of perpetrator can actually do that, though. You can't – it seems like you are all just coming to an agreement on a person not doing the act because -----

JUROR 3: You know what, if this was a real case – well, I think he's guilty. Okay, I have no doubt. I think he was doing it before. He preys on – he picks his prey. He looks at kids -----

JUROR 9: I agree.

JUROR 3: ----- that don't have dads around or travelling. I think he picks his mark. And I think that he was probably escalating. But that is my opinion -----

JUROR 9: You can't prove that.

JUROR 10: You can't prove that.

JUROR 3: ----- but I can't prove that.

Jury 67²⁷⁷ adopted a similar line of reasoning and convicted on Count 5 but was hung on Count 6. This jury agreed that the pattern in the defendant's behaviour negated any reasonable doubt that they might have. However, jurors also suggested that sexual intercourse was not consistent with the pattern, which led to the inability of the jury to reach a unanimous verdict for Count 6.

²⁷⁷ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

Case Study 1: Decisions to convict and acquit for the weak claim²⁷⁸

Juries attending a joint trial of three complainants were required to return a total of six verdicts. This case study investigated jurors' deliberation process in a joint trial to assess their susceptibility to impermissible reasoning triggered by the joinder of multiple counts. The trials selected for this analysis were Trials 7 and 8. Given that Count 1 involved the weakest complaint of sexual assault, there was also a question as to whether the juries that voted to convict for Count 1 engaged in impermissible tendency reasoning.

Permissible tendency reasoning

In the joint trials, all juries were informed that they could use tendency reasoning according to the specific instructions in the judge's tendency evidence direction. The direction provided to the juries is summarised below (as it pertained to Count 1):

Here, the Crown asserts that at the time of the alleged offences, the accused:

1. had a tendency to have a sexual interest in young boys under the age of 12;
2. had a tendency to engage in sexual activities with young boys under the age of 12;
3. had a tendency to use his position of authority as a soccer coach to gain access to young boys under the age of 12 so that he could engage in sexual activity with them.

How can you use this evidence? The Crown relies upon this evidence to prove beyond reasonable doubt that the accused had a sexual interest in each complainant and was willing to act upon it in the way that each complainant alleges. The Crown argues that you will find the accused's sexual interest proved beyond reasonable doubt and therefore you can use it to prove the allegations in the indictment beyond reasonable doubt. The Crown says that you will be satisfied that the accused had a sexual interest in Justin McCutcheon on the basis of the acts of a sexual nature committed against Justin McCutcheon. Finally, the Crown says that you will be satisfied that the accused had a sexual interest in Timothy Lyons on the basis of the acts of a sexual nature committed against Timothy Lyons.

Before you can use the evidence of these other acts of a sexual nature in the way the Crown asks you to do so, you must make two findings beyond reasonable doubt. If you decide that one or more of the acts *against Justin McCutcheon* is proved beyond reasonable doubt and you can infer or conclude beyond reasonable doubt that the accused had a sexual

²⁷⁸ Inspections of the conviction rates in joint trials: Trial 7 versus Trial 8 showed that more juries rendering a verdict without a question trail voted to convict the defendant on Count 1 (37.5 per cent) than juries who received a question trail (12.5 per cent); however, the question trail had no significant influence on verdict, as is discussed in Part 5.

interest in Justin McCutcheon, you may use that fact in determining whether the accused committed the offences against Simon Rutter and Timothy Lyons. And finally, if you decide that one or more of the acts *against Timothy Lyons* is proved beyond reasonable doubt and you can infer or conclude beyond reasonable doubt that the accused had a sexual interest in Timothy Lyons, you may use that fact in determining whether the accused committed the offences against Simon Rutter or Justin McCutcheon.

The evidence must not be used in any other way. It would be completely wrong to reason that, because the accused has committed one offence or is guilty of one piece of misconduct, he is therefore generally a person of bad character and for that reason must have committed all the offences charged. That is not the purpose of the evidence being placed before you and you must not reason in that way.

You cannot use the tendency evidence in any way prejudicial to the accused unless you accept the Crown's argument that it shows that the accused had a sexual interest in, for example, Simon Rutter which therefore makes it more likely that the accused committed the other offences charged against him, that is those involving Justin McCutcheon and Timothy Lyons. Remember that you are required to find that the elements of each specific charge are proved beyond reasonable doubt before you can find the accused guilty of that charge.

According to this direction, juries could either decide that the evidence revealed a tendency (or sexual interest) on the part of the accused, or that it did not. In relation to Count 1, if jurors found beyond reasonable doubt that the defendant was guilty of the sexual acts against Justin and/or Timothy, *and* they could “infer or conclude beyond reasonable doubt that the accused had a sexual interest” in Justin and/or Timothy, they could then use that finding “in determining whether the accused committed the offence against Simon Rutter” (Count 1).

Thus, either of the above approaches amounts to permissible jury reasoning. The juries could use the tendency evidence in the way they were directed, or they could decide that the defendant did not have the specified tendency

Extracts from the deliberations of the 16 juries who viewed a joint trial are included in the following case analysis. Most (75 per cent) of these juries voted to acquit the defendant on Count 1, whether or not they received a question trail to assist their decision making. Of the four juries that voted to convict on Count 1, three reached this verdict without the aid of a

question trail (Juries 64²⁷⁹, 66²⁸⁰ and 75²⁸¹), while one jury reached this decision with the aid of a question trail (Jury 51²⁸²).

*Juries that convicted on Count 1*²⁸³

First, we describe the reasoning of the three juries that deliberated without a question trail. The 11 jurors in Jury 64²⁸⁴ approached the task in a verdict-driven manner, spending only 16 minutes deliberating and providing few reasons for their decision. Jurors began their deliberations with a show of hands concerning Count 1; 10 jurors voted guilty and one juror dissented. Next, they discussed the evidence pertaining to Count 1, in particular Juror 7's uncertainty about the existence of a swimming pool, since the evidence presented at trial about the swimming pool was contradictory. Simon had testified that he was indecently assaulted after he had been pushed into the defendant's pool. However, the defendant testified that his home did not have a swimming pool, placing the credibility of the complainant in issue. Other jurors also volunteered their doubts about other aspects of Simon's evidence – whether or not there had been a thunderstorm and whether Simon had run to his mother's car dressed only in a t-shirt.

There was very little discussion as to why the jury initially voted guilty for Count 1. With the dissenting juror holding out to acquit, the jury moved on to discuss the other counts. It voted guilty for Counts 2, 3 and 4 because of the credibility of Justin's evidence, which was supported by his mother and best friend. Similarly, the jury voted guilty for Counts 5 and 6, finding no reason why Timothy would have made up his evidence. When the jurors returned to Count 1, they re-read the judge's tendency direction, and decided they could infer a pattern of behaviour on the part of the accused:

JUROR 4: I agree. I think – what did they say – the pattern of behaviour can be inferred.

(Most jurors voiced agreement).

JUROR 6: Yep. And once we have established it for pattern of behaviour, we can then look at the other question ----

JUROR 11: So I think we can infer that there was an interest ----

After a short discussion of the defendant's behaviours, the dissenting juror was asked whether

²⁷⁹ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁸⁰ Ibid.

²⁸¹ Id.

²⁸² Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

²⁸³ Juries 64, 75, 66 and 51.

²⁸⁴ Trial 7, joint trial with standard and tendency evidence jury directions.

she was still voting not guilty for Count 1:

JUROR 6: ... So given this evidence and the pattern of behaviour, are you still sticking with a not guilty verdict?

JUROR 7: Uhm, I would have liked to have seen a bit more evidence in that, but I think given all the other facts, about the case, I would go 'guilty'.

Finally, this jury was not dissuaded by the defence allegation that Simon had fabricated his evidence to obtain victim's compensation:

JUROR 5: The victim's compensatory thing threw me a little. I don't know if -----

JUROR 9: They threw that in to make us feel like, "Oh, he's not guilty. This guy is doing a victim's claim. He's lying". And that's what they tried to do.

JUROR 11: That is the only difference -----

(Most jurors voiced agreement).

The deliberation of Jury 64²⁸⁵ revealed that the jury had little incentive to deliberate, taking a total of 16 minutes to deal with six counts. Although the deliberation did not disclose why 10 jurors voted guilty for Count 1 on the first show of hands, there was no indication that they engaged in impermissible propensity reasoning. In fact, when they returned to Count 1 and the dissenting juror's reasons for voting not guilty, all jurors were satisfied beyond reasonable doubt of the sexual acts committed against Justin and Timothy *before* they concluded that the accused had a sexual interest in Justin and Timothy. This permissible use of the tendency evidence persuaded Juror 7 to change her vote.

Jury 75²⁸⁶, comprising 13 jurors, began by discussing the evidence of the three complainants. It decided, initially, that Justin was the only credible complainant, and that Simon's evidence in relation to Count 1 was the weakest. Jury 75 deliberated in relation to Count 1 only after considering all the other counts and making findings of guilt in relation to Counts 2 to 6. When discussing Simon's evidence, the mock jurors identified several doubts about the time and place of the allegation, the existence of a swimming pool and Simon's possible motivation to fabricate his evidence (victim's compensation). As a result, Jury 75 found Simon's evidence insufficient to prove the elements of Count 1 beyond reasonable doubt:

JUROR 3: That is that time and place thing. You know, I don't think there was enough proof about what – the actual place.

²⁸⁵ Ibid.

²⁸⁶ Id.

JUROR 9: The evidence by Simon didn't justify -----

JUROR 3: I don't think we can say it's beyond reasonable doubt. Well, I can't. I have doubts about it.

In order to resolve their doubts, one juror wondered if they could use the tendency evidence:

JUROR 9: Could you use the tendency evidence, as you have convicted that – or you are possibly saying 'guilty' on those other counts, of 2, 3, 4 -----

JUROR 3: That we think he had a sexual interest in the other two boys, so -----

JUROR 9: And that you can say there was a pattern of behaviour, that he was grooming these young boys.

Because the jury was unsure whether they were allowed to consider the defendant's pattern of behaviour, they re-read the judge's tendency direction, but after misinterpreting the direction, they concluded they could not use the tendency evidence in relation to Count 1. After further deliberation, some jurors were persuaded that Simon was telling the truth:

JUROR 3: It was a less severe act as well. He was rubbing him down with towels, etcetera; whereas the others, he actually went further.

(Most jurors voiced agreement).

JUROR 10: If Simon was going to lie about it, he could have made up a much more -----

JUROR 9: Elaborate.

JUROR 10: ----- serious, elaborate story, like – instead of just handling. I shouldn't say "just"; a lesser offence of sexual -----

However Juror 4 was not convinced:

No, I still don't. I think that, to me, the issue of the pool is still pretty fundamental; that the person is basing it on, "I was pushed into a pool. Took me out". There's got to be a pool there in the place that he's at. And it was not proven that there was a place.

Nonetheless, Juror 3 recognised the similarities in the descriptions given by the three complainants to explain the defendant's grooming behaviours and his justifications for the alleged abuse:

JUROR 3: And the boys used the same language, "Having fun, joking around; this is what men do". So they were using the same words about what he had said -----

(Most jurors voiced agreement).

JUROR 10: He said that as well, the first witness. Simon used the same -----

JUROR 3: They have used the same language about how he spoke to the -----

JUROR 10: That's right

JUROR 3: So that makes that tendency stuff more believable.

Although Jury 75²⁸⁷ reconsidered whether the tendency evidence was sufficient to overcome their doubts, with no resolution as to the retrospective application of the tendency direction, they realised that the defence barrister's tactics during Simon's cross-examination had 'planted' doubts in their minds about Simon's credibility:

JUROR 10: Simon also denies it and the defence has a tendency to infer guilt – infer in the question. They don't care about what your answer is. They will ask an incriminating question, knowing that you will deny it. But once that question has been posed, then the jury -----

JUROR 12: You can't forget about it.

JUROR 10: ----- the jury can't forget the question and that puts that doubt in your mind. And that's what he was doing all the way. And that's where he said, "Are you after the money?" "No, I'm not." But, of course, the jury is now thinking, "Oh, yeah, he hasn't got a regular job" ----

JUROR 7: That's right.

JUROR 6: And the other two guys were okay. They had reasonable jobs.

JUROR (?): Very good-----

(Most jurors voiced agreement).

JUROR 10: But is that a reason to say "not guilty" because he's got casual work?

JUROR (?): No.

After dismissing their doubts about Simon's credibility, a sticking point for Jury 75 was their uncertainty as to whether the alleged act of indecency after the pool incident had occurred. Jury 75 decided that an act of indecency had occurred the first time Simon and the defendant showered together. Simon had said "I wouldn't let him dry my genitals" after the first shower. Jury 75 voted guilty in relation to Count 1, although the prosecutor had stated that Count 1 pertained to an act of indecency in the bathroom after Simon had been pushed into a swimming pool by the defendant.²⁸⁸

²⁸⁷ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁸⁸ The Crown's opening remarks informed the jury that "After a shower, the accused offered to dry Simon with a towel which progressed to the accused stroking Simon's penis as the accused pressed his body against him". The Crown's closing remarks stated that Count 1 was an allegation that "the accused masturbated Simon's penis

Jury 66²⁸⁹ began its deliberations in a similar verdict-driven manner, with several jurors announcing that the defendant was guilty of all six counts before any discussion of the evidence. The discussion went back and forth between a group of jurors who were convinced of the defendant's guilt and two jurors who had doubts. Although the jury deliberation was not clear-cut, because of the degree of over-talking, some dominant jurors appear to have engaged in impermissible tendency reasoning because of their global approach in evaluating the evidence, that is, before they had discussed the elements of each count separately, as they had been directed to do by the trial judge. The group in favour of conviction recognised that there was a pattern of behaviour by the defendant and similarities in what the defendant had told the complainants (such as, "that's what men do") and in relation to the fact that two complainants had shared the defendant's bed. One dissenting juror disagreed, arguing that the Crown had not proved "any of this". At this early point in the deliberation, the dissenting juror's observations were correct because the jury was only permitted to use the patterns or similarities once they had made findings beyond reasonable doubt that the defendant had a sexual interest in one or more of the complainants. Only then could they use that finding to "infer or conclude beyond reasonable doubt that the accused had a sexual interest" in another complainant.

Most of the deliberations by Jury 66²⁹⁰ focused on the dissenting jurors' doubts, with the dominant group trying to remove those doubts. Only towards the end of the deliberation did the jury consider each count separately in order to render verdicts. The jury voted guilty on all counts separately, although when they delivered their verdict for Count 1, they did reiterate an earlier discussion on credibility to review why they reached that verdict.

We can infer from the deliberation that after discussing the evidence, this jury took a global approach to the counts and considered each count in the context of every other count. Even though the jury recognised that Justin was the most credible complainant, the similarities in the evidence of all three complainants appeared to be the deciding factor. One juror based her verdict on the fact that there was "too much in common" between the complainants' accounts for the defendant to not be guilty. This may have amounted to impermissible reasoning since the juror reasoned that "if there's one person, you think, 'OK', two people you think, 'Oh!', three people, you know –", indicating that she was basing her verdict on the cumulative impact

in his bathroom in August 1993, as he pressed his body against Simon". The jury directions specified that Count 1 was an allegation of indecency but did not describe the event with any more particularity.

²⁸⁹ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁹⁰ Ibid.

of the evidence rather than considering each count separately. However, there was no clear evidence that the jury, as a whole, reasoned in this way.

With a question trail to assist it, Jury 51²⁹¹ compared the evidence of the three complainants. The mock jurors decided that Simon's evidence about Count 1 was "a little bit dodgy" (Juror 1) compared to that of Justin and Timothy, because of the unresolved issue about the existence of a swimming pool, and the somewhat implausible evidence from Simon that he had left the defendant's house wearing only a t-shirt. Nonetheless, the jurors considered that the evidence given by each complainant supported the evidence of the others:

JUROR 5: You are not going to get three people that will come out and accuse a person like that without there being some basis to it -----

JUROR 12: ----- But the thing that absolutely clinched it for me was Simon Rutter who had never met the three boys, from my understanding, and they wouldn't have been in the courtroom when this was happening – had never met each other. They didn't have a chance to corroborate. But Simon Rutter and Justin both quoted him as saying something "This is whatever men do" -----

The jury also considered that the defendant, when giving evidence, was trying to hide something, and that his demeanour in the witness box was not that of an 'innocent' person who would have been more 'shocked' and 'devastated' when giving evidence. Because the complainants did not know each other before the trial, and because the alleged events happened in different years, the jury considered that the defendant was guilty of all counts, including Count 1:

JUROR 5: I think they corroborated each other, because they didn't know each other. You know, to get three people come out, one supporting the other -----

JUROR 3: It speaks -----

JUROR 5: ----- and I found them all to be credible.

JUROR 3: Especially the same pattern of behaviour for each one of the kids.

JUROR 5: Yes, exactly.

It appeared that this jury used the fact that three complainants had made similar allegations against the defendant to engage in a form of improbability reasoning., identifying that the three separate complaints, who did not know each other, increased the credibility of each other's accounts. They reasoned that it was improbable that three complainants, independently, would

²⁹¹ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

make similar allegations of sexual abuse and be lying:

JUROR 12: ----- I don't know, it's pretty difficult. As someone said, he could be innocent but there would be too many coincidences. These guys met each other in a pub and got together and said, "Oh, did you belong to this soccer club?", or whatever. There's too many coincidences.

JUROR 4: They weren't smart enough to make that kind of stuff up. That would take a lot of conniving, and you would have to be a real – you know, to get all those things -----

JUROR 5: Anyway, nothing came up that they could have known each other.

Each of the four juries that voted guilty on Count 1 in the joint trials did so for different reasons. Although Jury 64²⁹² conducted a speedy deliberation, the mock jurors correctly interpreted the judge's tendency direction, and engaged in permissible tendency reasoning, as directed by the trial judge, which ended with the majority persuading the dissenting juror to change her vote from not guilty to guilty for Count 1.

By contrast, Jury 75²⁹³ did not engage in tendency reasoning because they misunderstood the tendency direction. As the jurors had found the defendant guilty in relation to the counts involving Justin and Timothy, they were entitled to "infer or conclude beyond reasonable doubt that the accused had a sexual interest" in Justin and/or Timothy, and to use that finding "in determining whether the accused committed the offences against Simon Rutter". Instead, Jury 75's finding of guilt in relation to Count 1 was based on an incorrect interpretation of the facts in the case, as it decided that an act of indecency had occurred during the first showering incident even though Simon said nothing had happened.

Juries that acquitted in relation to Count 1²⁹⁴

Of the 12 juries that voted to acquit the defendant on Count 1, seven were given a question trail, while the rest deliberated without one.

The members of Jury 83²⁹⁵ began consideration of the evidence for Count 1 with a discussion of their doubts regarding Simon's evidence, including the fact that, compared to Justin and Timothy, Simon was unemotional in the witness box. Initially, the jury voted not guilty for Count 1 but decided not to record that vote until they had considered all the other counts as Juror 2 explained:

²⁹² Trial 7, joint trial with standard and tendency evidence jury directions.

²⁹³ Ibid.

²⁹⁴ Juries 39, 40, 41, 52, 53, 54, 63, 76, 83, 84, 85 and 86.

²⁹⁵ Trial 7, joint trial with standard and tendency evidence jury directions.

Well, I quite like to not put the verdict down until we discuss all of them ... because we still have to consider each case, case by case; but at the same time there is the tendency evidence. I am pretty convinced that he has this tendency. So I will put my hand up to say “I think not guilty”, but I’m not going to put that down yet, to be honest.

The jury’s discussion of the three counts involving Justin was very short, compared to their lengthy discussion of Count 1. No doubts were raised about Justin’s evidence; the jury was convinced by Justin’s behaviour change as a child and by his supporting witnesses. All agreed that the defendant was guilty on Counts 2, 3 and 4, with no discussion of the separate elements of each offence.

The jury’s deliberation of Counts 5 and 6 was much longer, reflecting the number of ambiguities in Timothy’s evidence, compared to Justin’s evidence. Jury 83²⁹⁶ dismissed Timothy’s mistake about the movie he watched. They were impressed by the fact that he remembered a bedroom with a TV on top of a chest of drawers, and by the evidence of Ellen Samuels who confirmed Timothy’s description of her bedroom. Deciding that Ellen had no reason to lie, the jury voted guilty for Counts 5 and 6. When Jury 83 returned to Count 1, they listed the various doubts raised by Simon’s evidence:

JUROR 6: There’s a few things; the pool, the date and the mum seeing him half naked -----

JUROR 5: ----- Yeah, yeah, right, ‘half naked’. And what was the other one, sorry?

JUROR 6: Pool and got the dates wrong; July and May -----

JUROR 5: The dates. But the dates are a really big deal?

JUROR 6: I don’t know that’s what we are disputing. It might be the pool and the mum seeing -----

JUROR 5: And that’s my only problem.

JUROR 4: ----- could be confused about a pool, where it was; it could have been a lap pool or water fountain, who knows -----

JUROR 5: Yeah, or a blow-up pool, for all we know.

JUROR 4: ----- could anyone in this room honestly say you would pick someone up half naked -----

JUROR 5: God, no.

JUROR 4: ----- and not notice?

Despite further deliberation, during which the jury appeared to believe Simon’s story that he

²⁹⁶ Ibid.

had run out of the defendant's house dressed only in a t-shirt after being sexually abused, Jury 83²⁹⁷ decided it needed more evidence to convict on Count 1:

JUROR 3: ----- I am kind of keeping in mind that this guy is innocent until proven otherwise. So maybe that's what I am struggling with. If he's guilty, I am looking for something to hang my hat on, then that's a different story. But if you kind of approach it as "he's totally innocent, he's done nothing wrong", I want to feel really confident that this definitely happened before -----

JUROR 5: I agree.

JUROR 2: I agree as well ... I'm not confident saying he's guilty with this count.

In particular, the jury decided that they could not use the tendency evidence because Simon was the first complainant:

JUROR 1: "The Crown asserts that at the time of the alleged offence the accused had a tendency to have sexual interest in young boys." We have got no evidence to say that he had that tendency at the time -----

JUROR 8: Oh, at the time.

JUROR 1: You know what I mean; because it was the first one, we don't have any evidence to say that he had those tendencies at the time.

Although some jurors later decided, after re-reading the judge's tendency direction, that they *could* use the tendency evidence for Count 1, they believed that there were too many uncertainties about Simon's evidence to vote guilty as Juror 3 explained:

Like, you can kind of get creepy feelings from people without something necessarily had happened. To my mind, that is still, like, in the realm of possibility; that he's seen this guy [the defendant] and he's remembered an incident or maybe there was something else, like an argument. Like, I am not saying that he's deliberately going off money or deliberately trying to lie, but there's enough kind of holes and no one to kind of support him ... I know that sounds terrible because he's just a child at the time and you would hope that whatever he says is true, but there just are, like – to my mind – and putting myself in the zone of this defendant, he may be guilty of a lot but that doesn't mean he's guilty of everything.

Like Jury 83, Jury 84²⁹⁸ also identified that the main problem with Simon's evidence was the pool:

²⁹⁷ Trial 7, joint trial with standard and tendency evidence jury directions.

²⁹⁸ Ibid.

JUROR 5: With him, I believe, like, immediately I went “yes” but the pool thing is the problem, isn’t it? ... So that is the time and place for him, is the one thing that is missing.

This jury also had concerns about the motives of the defendant in befriending boys from a soccer team:

JUROR 1: Having been a coach myself, I know what it’s like to single out five kids and take five children -----

JUROR 5: You don’t do that.

JUROR 1: ----- you don’t do that. You have to take the whole team, if you were a manager or caretaker. I can talk from experience ... taking kids; you don’t do it. It is looking for a problem.

JUROR 5: But I think that’s why someone ... goes into and gets 11 boys and then picks out that weak one; the one that’s vulnerable; or even two. And then works it down to the one that he can actually manipulate.

Despite recognising these issues, Jury 84²⁹⁹ was more concerned about the unsupported parts of Simon’s evidence: the existence of a pool, the thunderstorm, whether Simon had actually run out of the defendant’s house half-naked, and the likelihood that the defendant would have grabbed Simon’s crotch in front of other boys and parents. These doubts resulted in considerable speculation:

JUROR 7: My only concern is the pool because even in the townhouse, there’s very limited room for a pool in the backyard. There’s certain requirements you have to have a certain backyard of a certain size to get a pool. And the other thing was the thunderstorm, the defence said that the majority of that particular month was ‘no thunderstorm’. Usually, with winter, it is a very dry ... because there’s not enough evaporation for the rain ...

(All talk at once) ...

JUROR 8: And the mother is coming to the son naked, "Where are the clothes?" So-----

JUROR 7: I would have confronted him at that stage.

JUROR 8: And mother happened to turn up; so doing it in such a situation where somebody can turn up ...

JUROR 1: You know, parents can – “kid’s so excited to see me now”. It could be that as well.

JUROR 8: And naked, running up to?

²⁹⁹ Id.

JUROR 1: Well, you don't know if he had a towel.

JUROR 8: He was wearing a T-shirt.

JUROR 1: He had something on, so -----

JUROR 7: But only a t-shirt ----- That's what he said, so -----

After seven jurors decided to assume, as a fact, that there was no pool in the defendant's backyard at the relevant time, they could not agree beyond reasonable doubt that the 'time and place' element of Count 1 was satisfied. Unlike other juries, this jury did not consider the issue of tendency evidence at all. By contrast, the jury voted guilty for Counts 2, 3 and 4 (although one juror dissented on Count 4) with almost no analysis of Justin's evidence. As Juror 5 observed, "It's got everything. It's got the mother even noticing and – everything". Similarly, Jury 84³⁰⁰ undertook very little analysis of Timothy's evidence, although they did not deliver unanimous verdicts for Counts 5 and 6.

Jury 85³⁰¹ began with a global discussion of the evidence they found to be the least credible. The deliberation rambled back and forth between different parts of the evidence of all the witnesses, including the defendant. When the jurors eventually decided to consider the counts, they decided to do Count 1 last, but without giving clear reasons.

For Counts 2, 3 and 4, the jury was convinced of Justin's credibility as a result of his emotion in the witness box and the two supporting witnesses who gave evidence of his change in behaviour. As a result, they voted guilty on all three counts. There was a much longer discussion of Timothy's evidence (which the jury perceived as less strong and more ambiguous than Justin's evidence) and speculation about why Timothy's mother failed to give evidence (compared to Justin's mother). All agreed that the tipping point was the evidence of Ellen Samuels, who confirmed that she had a townhouse with a TV in the bedroom to which the defendant had access at the time Timothy alleged he was abused. The jury voted guilty for Counts 5 and 6. By contrast, the jury was less convinced by Simon's evidence:

JUROR 1: ... but there is nothing strong in the Simon Rutter case ...

JUROR 12: There is no witnesses to the showers.

Because of the inconsistencies in Simon's testimony and the lack of supporting witnesses, the jury spent much more time speculating about whether certain things happened or did not

³⁰⁰ Trial 7, joint trial with standard and tendency evidence jury directions.

³⁰¹ Ibid.

happen. For example:

JUROR 10: Maybe his whole life was stuffed up because of that event.

JUROR 1: We don't know. Maybe he was stuffed up – he says he's a trouble child. Appears to be lonely. Who knows. I don't know whether it is because of ----

JUROR 3: Yeah, that money is there for victims of crime. He didn't put the system in place. You can't blame the bloke for taking advantage of ----

JUROR 1: I just have something unsettling with this Simon Rutter ...

JUROR 6: This is the one who couldn't tell the difference between summer and winter.

...

JUROR 1: I don't find him as gullible as the other cases. There is just something there which isn't going down favourably ----

JUROR 2: It is just one word against another.

(Most jurors voiced agreement).

With the jury unable to reach unanimous agreement on Count 1, they discussed their main doubts:

JUROR 11: So it is the dates, the pool, the thunderstorm, the running outside with no clothes, and then the mum.

Even after considering the defendant's tendency, half of the jurors in Jury 85 did not consider this evidence was sufficient to remove their doubts about Simon's evidence:

JUROR 9: I have no doubt because if he's guilty on the other cases ... he probably abused him, but I don't think beyond reasonable doubt that I can say that at that – I want to but I don't think that's what they want us to do ...

JUROR 3: If she [the mother] had said, "Yep, he ran out in a t-shirt", well, then that's evidence ----

JUROR 1: We have got ... no evidence to say that someone recognised the emotional stress he was under; coming out with a t-shirt from the house, there's nothing there.

JUROR 9: I suppose the bottom line is, yes, perhaps he was indecently assaulted, but to what extent is the issue and did it occur as per the script, that he's saying? He may have been assaulted in some way and it's hurting, and no doubt; and this guy should rot. But it's, uhm ... there's not enough.

JUROR 1: It is not as strong as the other two.

In the end, the jury was split evenly on Count 1, with six members voting guilty and six voting not guilty.

Jury 86³⁰² began with a discussion of their doubts about Count 1. Although Juror 9 recognised a pattern of behaviour (“the whole pattern is saying ‘this guy is doing repeated sexual assault to these boys’, and these are 10-year-old boys”) jurors focused on the gaps in Simon’s evidence: whether a swimming pool had existed; whether there had been a thunderstorm as alleged by Simon but denied by the defendant; and why no one came forward to support Simon’s allegation that the defendant had a habit of grabbing the crotches of boys in the soccer team. Some jurors believed that Simon had made up his allegations and then convinced himself they were true:

JUROR 8: You could convince yourself that it happened.

JUROR 6: That is true, it can happen.

JUROR 8: You can hear it happen to somebody else and ----

JUROR 6: And then you think about it or fantasise about it sometimes ----

JUROR 8: Yeah.

JUROR 6: And then it becomes a ----

JUROR 5: Reality.

JUROR 6: ---- reality.

JUROR 8: Particularly if there’s money at the end of it.

Jury 86 was dominated by Juror 8 who repeatedly said, in relation to all the counts, that there was ‘no evidence’, and decided that the complainants’ accounts of sexual abuse were insufficient to prove the offences beyond reasonable doubt. Other jurors agreed and decided that five out of six counts (including Count 1) could not be proved beyond reasonable doubt. Towards the end of their deliberation, the jury was in such a hurry to finish that the discussion ended in a rather shambolic way, with several jurors confused about what they had decided:

JUROR 4: So the first one we said “not guilty”? And everyone was pretty comfortable with that at the end, because they felt that they couldn’t do it with enough reasonable doubt

...

JUROR 4: Number 2, we did ----

JUROR 6: Not guilty.

³⁰² Id.

JUROR 7: I thought we did 'guilty'.

JUROR 6: Guilty.

JUROR 1: Number 2, 'not guilty'.

JUROR 4: 'Not guilty'? Because there wasn't enough reasonable doubt? 3 and 4?

JUROR 6: 3 and 4. 'Guilty'.

JUROR 1: 3, 'guilty'.

JUROR 2: 3 'guilty'; 4 'not guilty' ...

JUROR 2: 5, 'not guilty'.

JUROR 1: 4/5 'not guilty'.

JUROR 7: To me, I would do it the other way around ----

JUROR 6: Number 5 was 'guilty', is that what you are saying?

JUROR 7: Yes, totally. But ----

JUROR 4: So you think 4 should be 'guilty'?

JUROR 7: I think 3 should be 'not guilty' and 4 should be 'guilty'. ...

JUROR 6: We are hungry and we have to go (laughs).

JUROR 7: No, I care. I care ... yeah, I really care. Actually, I probably think 'guilty', 'guilty' ----

JUROR 4: On 3 and 4?

JUROR 7: Yep, because there's – in that logic, there is ----

JUROR 6: Okay, who says 'guilty', 'guilty' on 3 and 4 ...

(Five people raised their hands).

JUROR 6: I don't. Who says 'guilty' or not guilty' for 3 and 4 – 'guilty' or 'not guilty'? ...

JUROR 6: 4 was 'not guilty' ... 5 and 6 were 'not guilty' ...

JUROR 4: Only 3 is 'guilty' at the moment

JUROR 12: Well, I thought that was 'guilty' and I said why but I think everyone thought it was 'not guilty', so.

JUROR 4: But I guess it is a unanimous verdict, so if you feel really strongly about it.

JUROR 12: ... well, I just thought the witness was quite reliable and I didn't think there was any reason really why there was doubt. But I mean, if other people are saying that there was reasonable doubt, then ----

JUROR 7: That is for 4, correct?

JUROR 4: Yeah.

JUROR 6: So 1 was 'not guilty'. The only 'guilty' was number 3; is that right?

(Most jurors voiced agreement).

In the end, Jury 86³⁰³ only delivered one guilty verdict, for Count 3, with the obstructionist approach of Juror 8 and the time limit on the deliberation preventing them from undertaking a proper consideration of the evidence and the counts.

Jury 39³⁰⁴ voted guilty on all counts except Count 1. The not guilty verdict for Count 1 arose out of the jury's doubts about the existence of a swimming pool. By contrast, Justin's evidence in relation to Counts 2, 3 and 4 did not raise any doubts and almost no deliberation so, with a show of hands, the jury voted guilty for all three. For Counts 5 and 6 the jury was split, 10 jurors to two in favour of guilty, with little deliberation. In other words, Timothy's evidence, compared to Justin's, raised enough doubt for two jurors to vote not guilty and to argue that Timothy's evidence "was so made up" (Juror 3). After further discussion about the defendant's habit of picking vulnerable children and having no connection with them to start with, Juror 3 decided to change his vote to guilty for Counts 5 and 6.

Compared to Justin and Timothy, Simon's story "didn't really add up" (Juror 5), and more than half of the jury's deliberation was devoted to Count 1. Although Juror 10 noted that they were allowed to use tendency evidence, it was not sufficient to remove the doubts:

First of all, part of what we are allowed to work with is tendency evidence, which means that we are allowed to look at him and go, "We have got three people who have got a complaint against him". So there is a pattern forming ... If you have a think about the defence ... he never really came up with any reason as to "how come there's three people that have got a complaint against this soccer coach?" So that's my first point. So we are allowed to use the tendency evidence here. I still struggle – now, I put that aside because I still struggle with the first count, which is where – and I'm not saying that there was a pool or wasn't a pool. But my point is that the prosecution didn't prove one way or the other whether there could have been a pool. So that's why I have a doubt, where I can't go "guilty" on that one.

³⁰³ Id.

³⁰⁴ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

While other jurors agreed that there was a pattern and a tendency, Simon's evidence for half of the jurors "just didn't stack up" (Juror 5). For Count 1, the jury was split six one way and six the other.

Jury 40³⁰⁵ closely followed the questions in the question trail. For Count 1 they decided that they were not satisfied that the first element had been established beyond reasonable doubt; that is, that the events had occurred between 1 March and 1 December 1993, even though they recognised the similarities in the evidence of each of the complainants. In other words, Jury 40 decided that Simon's inconsistencies about when he had joined the soccer team (March or July) meant that they could not be certain beyond reasonable doubt that the alleged act of indecency had occurred during the relevant period of time, thus failing to recognise that both March and July were within the period specified in the question trail.

Jury 41³⁰⁶ also began their deliberation by following the question trail. Like Jury 40, Jury 41 was not satisfied that the Crown had proved the first element of Count 1 beyond reasonable doubt; that is, that the events had occurred between 1 March and 1 December 1993. The jury also held doubts about the existence of a swimming pool as alleged by Simon, and Simon's recollection of a thunderstorm; they accepted the defence argument that Simon was "going for the money" (Juror 1).

Jury 52³⁰⁷ began their deliberation in relation to Count 1 with questions about the existence of the swimming pool that Simon alleged he was thrown into. This meant that the jury was not satisfied beyond reasonable doubt that the Crown had proved that the defendant had committed an act of indecency against Simon, and thus voted not guilty on Count 1. However, after finding the defendant guilty in relation to Counts 2 to 4, Jury 52 recognised that they were involved in a 'process' to do with the defendant's pattern of behaviour:

JUROR 7: But how can you say 'not guilty' for 1 but 'guilty' for all the others?

JUROR 1: You could possibly infer 'guilty' for one -----

JUROR 3: At the end of the process.

JUROR 1: ----- at the end of the process ... The way I saw it: if you just had the facts on Count 1, at least I would have said 'not guilty'.

Nonetheless, contrary to the judge's tendency direction, the jury decided they could not re-visit

³⁰⁵ Ibid.

³⁰⁶ Id.

³⁰⁷ Id.

Count 1 based on the recognition of a pattern of behaviour by the defendant.

Like other juries, Jury 53³⁰⁸ expressed several doubts about the prosecution's case in relation to Count 1, including whether Simon was motivated by money, the 'vagueness' of his evidence, the lack of evidence from Simon's mother, and the existence of a swimming pool. As Juror 12 said, "he's the earliest case and he really has nothing to back him up, like Justin had his friend and his mother to back him up. Simon Rutter pretty [much] has no one". Although some jurors considered that the defendant was guilty of Count 1, the jury was not convinced beyond reasonable doubt:

JUROR 12: I think he's probably guilty but -----

JUROR 7: We don't have the -----

JUROR 12: No, you are missing a bit there.

JUROR 8: Yeah ...

JUROR 12: And it's because it's beyond reasonable doubt. I think he's guilty of it.

JUROR 7: So do I.

JUROR 8: But there are reasonable doubts because of the inconsistency ...

JUROR 12: For that reason -----

JUROR 7: We can't convict ...

JUROR 8: ... Yes, we might agree that he committed an act of indecency and you can get that from the evidence or from your impression of the evidence; and that the act was indecent and that they were under 16. But the time and place, if there was no pool at the place and he said it happened when he got pushed into the pool, then in this particular charge I think you have to say 'not guilty'.

Similarly, Jury 54³⁰⁹ was 'thrown' by the swimming pool issue and could not accept that Simon was pushed into a pool as he alleged. Because it could not reach a decision on Count 1, Jury 54 considered all the other counts first, voting guilty for Counts 2 to 6. The jury then considered the applicability of the tendency evidence to Simon's allegation of sexual abuse, although the jurors were confused about how it could apply to Count 1:

JUROR 9: Does tendency have to follow a first act or can it because this is the first thing?

JUROR 3: This is the first thing.

³⁰⁸ Id.

³⁰⁹ Id.

JUROR 9: The earliest ... Can tendency only follow after or can tendency -----

JUROR 1: It could be after.

JUROR 6: ... I don't know ... how they have given us such solid point indicators in a certain way [in the tendency direction], like you said. How do they know at that time of the alleged offence he had a tendency to have sexual interest in young boys under 12? ...

JUROR 11: It is looking at the entire three cases.

JUROR 6: No, that one is prior to the three cases.

JUROR 9: I don't think you can use tendency for this first one because you are going back in time.

JUROR 3: Yeah.

JUROR 9: You can't go back in time. It is the first thing, it is '93 and the other one is '95 and '97.

JUROR 3: Oh, that's a point ...

JUROR 9: Yep, so it's like you can't use tendency -----

As a result, they were unable to resolve their doubts; nine jurors voted not guilty on Count 1 and three jurors voted guilty.

Jury 63³¹⁰ began with a discussion of the elements of Count 1 with many of the jurors deciding the defendant was guilty. With further deliberation, it emerged that several jurors had doubts about the existence of a swimming pool at the defendant's townhouse, and were adamant that the prosecution had not proved the pool's existence beyond reasonable doubt. After Jury 63 voted guilty in relation to Counts 2 to 6, they considered whether they could use the tendency evidence to vote guilty in relation to Count 1:

JUROR 9: So now we have tendency -----

JUROR 4: We have enough now to go back to Count 1 and say he's guilty.

JUROR 9: Now, you have got tendency evidence; you have got two of them.

JUROR 4: Count 1 is 'guilty'.

JUROR 9: Yes, because you have -----

JUROR 13: Hang on, I still don't agree with that -----

JUROR 9: Not 1; you can't go back to 1.

³¹⁰ Trial 7, joint trial with standard and tendency evidence jury directions.

JUROR 4: Can't go back to 1? Sorry. Well, we have already convicted him on Counts 2, 3, 4, 5 and 6. The only count that we can't convict him on, with a bit of reasonable doubt is, Count 1 -----

JUROR 9: It legitimises the two that you do have evidence for ... Because you have got two, it is a pattern, which gives you more, a stronger basis.

Although the jury re-read the judge's tendency direction, they decided the tendency evidence did not resolve their doubts about the existence of the pool, although there were some uncertainties about how to use the direction. Even though several jurors thought the defendant was guilty in relation to Count 1, the jury voted not guilty:

JUROR 8: You have to go on the evidence.

JUROR 4: You have to go on the evidence. Otherwise you will be on the jury for the next few months.

JUROR 5: Okay, 'not guilty', but he's guilty.

JUROR 8: Yeah, I think he's guilty -----

Jury 76³¹¹ also began its deliberation by identifying the doubts in relation to Simon's evidence – in particular, the inconsistency in the dates when Simon said he joined the soccer team and the existence of the swimming pool, which resulted in a vote of not guilty for Count 1 early in the deliberation. Unlike most other juries, Jury 76 could not agree that the defendant was guilty on the other counts; 10 members voted guilty for Counts 2 and 3 (three not guilty), and 11 voted not guilty (two guilty) on Count 4. For Count 5, nine voted guilty (four not guilty) and for Count 6, seven voted not guilty (six guilty). Because the jury did not find the defendant guilty on any of the counts, they were not able to consider tendency evidence in the trial as identified by Juror 13:

You can only accept tendency evidence if you think at least one of them is guilty. Then you can say, "Yeah, that establishes a pattern" well, not a pattern, but, yes, he has a sexual interest in boys, if you think if you find him guilty of indecently assaulting a boy. But, together, we haven't found any single one of them beyond reasonable doubt. Therefore, you can't establish tendency, in my mind, if you can't prove one, 'you can't prove any one in individually, but because they are altogether', therefore there is tendency. The only tendency that I think they have proven is 'there's a tendency for this man to be accused of sexual assault', not that he actually did it.

³¹¹ Trial 8, joint trial with standard and tendency evidence jury directions and a question trail.

There was considerable variation in the reasons why the above 12 juries voted not guilty for Count 1. Four juries (39³¹², 83³¹³, 85³¹⁴ and 63³¹⁵) recognised that their guilty verdicts for Counts 2 to 6 meant that they could consider whether the tendency evidence should be taken into account for Count 1. Three juries ignored the tendency evidence altogether (84³¹⁶, 86³¹⁷ and 53³¹⁸) while for others the issue of tendency evidence was irrelevant because they were not satisfied beyond reasonable doubt of the first element of Count 1 – that is, the time and place (40³¹⁹, 41³²⁰ and 76³²¹).

For those juries that did recognise the issue of tendency evidence, two juries misunderstood the retrospective application of the evidence. Jury 52³²² incorrectly decided that they could not re-consider their verdict in relation to Count 1 by taking into account the defendant’s pattern of behaviour after finding the defendant guilty for Counts 2 to 6. Similarly, Jury 54³²³ incorrectly decided that they could not use the tendency evidence in relation to their verdict for Count 1 because Simon’s allegation was the ‘first one’ and a tendency could not apply ‘backwards’.

By contrast, Juries 63 and 83 were not prepared to use the tendency evidence to remove the number of doubts they had about Simon’s evidence, even though this was a permissible form of reasoning. Jury 85 was also troubled by the doubts they had regarding Simon’s evidence. Although the jurors thought it possible that Simon had been sexually assaulted, they decided they needed more evidence before they could convict. After considering the defendant’s tendency, half of the jurors in Jury 85 did not consider this evidence was sufficient to remove their doubts about Simon’s evidence.

Although Jury 84 had as many doubts about Simon’s evidence as Jury 83, they did not consider the issue of tendency evidence. Their decision that there was no swimming pool meant they could not be satisfied that the first element of Count 1 had been proved beyond reasonable doubt. Similarly, Juries 39 and 53 decided that there was insufficient evidence to prove the

³¹² Ibid.

³¹³ Trial 7, joint trial with standard and tendency evidence jury directions.

³¹⁴ Ibid

³¹⁵ Id.

³¹⁶ Id

³¹⁷ Id.

³¹⁸ Trial 8, joint trial with standard and tendency evidence jury directions, and a question trail.

³¹⁹ Ibid.

³²⁰ Id.

³²¹ Id.

³²² Id.

³²³ Id.

existence of a swimming pool. Juries 40 and 41 decided that Simon's inconsistencies about when he had joined the soccer team (March or July) meant that the first element of Count 1 could not be satisfied.

Differences between juries that convicted and the juries that acquitted

Many of the juries that acquitted the defendant on Count 1 did not scrutinise their doubts in relation to Simon's evidence as closely as the juries that voted to convict; they did not consider the validity of the doubts raised by the defence or scrutinise the defendant's credibility as closely as they did Simon's.

The deliberation transcripts disclosed that many juries either ignored or misunderstood the lengthy tendency direction. Of the juries that considered the tendency evidence, only Jury 64 correctly followed the trial judge's directions regarding the use of the tendency evidence in coming to their verdict to convict the defendant on Count 1.

4.4 Influence of the number of counts and number of witnesses on jury reasoning and decision making

This section tested the hypothesis that juries in a joint trial use the overall number of charges or witnesses to determine the guilt of the defendant. The results provided no support for the hypothesis that accumulation prejudice against the defendant triggered impermissible reasoning. This conclusion was based on a separate analysis of the accumulation of counts and accumulation of witnesses. Together, the quantitative and qualitative analyses on each issue confirmed that jurors and juries made logical and appropriate distinctions between the same types of offence alleged by different complainants, based on the strength of evidence. The findings demonstrated that mock jurors based their judgment of the defendant's culpability on their assessments of the complainants' credibility, not the overall number of counts and witnesses. There was no confirmation of reliance on accumulation in a joint trial, as there was no significant increase in conviction rates, or in judgments of the defendant's factual culpability for the focal complainant in trials with six versus two counts. Similarly, adding two prosecution witnesses in a joint trial did not increase conviction rates, and most notably did not elevate the conviction rate for the complainant with the weak claim. In addition, mock juror ratings of victim blame did not vary in response to increases in the number of Crown witnesses, as might be expected if jurors were improperly accumulating the evidence. Rather, victim blame was predicted by individual mock jurors' misconceptions about child sexual abuse. Results of the coding of the deliberations revealed no impermissible reasoning or reduction in the onus of proof in trials with more counts. Juries made more factual errors as they were exposed to more witnesses and their cognitive load increased, but there were no differences in the rate of uncorrected or persistent errors across the various trials – that is, the results in separate and joint trials were similar. A case study of deliberations in a joint trial showed that juries in trials with six counts devoted most available deliberation time to the weak claim – where the disparities in evidence were greatest – controverting the view that juries would gloss over these differences in a joint trial. A further case study of deliberations in a joint trial confirmed that none of the jury decisions to convict or acquit were based on impermissible reasoning about the tendency evidence.

4.4.1 Research aim

Courts have hypothesised that the number of counts and witnesses against a defendant in a joint trial may create a cumulative impression of guilt, prompting juries to use impermissible

reasoning based on the accumulation of witnesses. In other words, juries are presumed to use the overall number of charges or witnesses in a trial as a heuristic or peripheral cue about the culpability of the defendant, and thus fail to centrally process or analyse the evidence.

To examine jury susceptibility to impermissible reasoning by accumulation prejudice in joint trials, a series of analyses tested the impact on jury reasoning of the total number of counts and the total number of witnesses in a trial.

First, to test the influence of the number of counts, we compared groups of juries that heard separate versus joint trials in which the evidence was similar but the defendant was charged with either two or six counts.³²⁴ In other words, we compared (a) the separate trial with tendency evidence, in which the defendant was charged with two counts of child sexual assault³²⁵ with (b) the joint trial, in which the defendant was charged with six counts of child sexual assault.³²⁶ In both trials, the three male witnesses/complainants (Simon, Timothy and Justin) provided identical evidence, but in the joint trials, two additional supporting witnesses appeared on behalf of the complainant with the strong case (Justin).³²⁷

Second, to test the influence of the number of witnesses, we compared groups of juries that heard joint trials in which the evidence was similar but the Crown called a total of four versus six witnesses. For this purpose, in Trial 9 we omitted evidence from two supporting witnesses for the complainant with the strong claims (Justin) in Trial 7. Both witnesses had verified that at the time of the alleged abuse, Justin displayed uncharacteristic behaviour that indicated he was distressed, and one of the witnesses confirmed that Justin had told him about the abuse at the time it occurred. In both trials, 7 and 9, the juries received the same jury directions, including a tendency evidence direction.

We tested jury reasoning by accumulation of the counts with a total of 33 juries and 373 mock jurors: 17 juries ($n = 197$ jurors) viewed the separate trial with tendency evidence (Trials 5 and 6), and 16 juries ($n = 176$ jurors) viewed the joint trial (Trials 7 and 10).³²⁸ The analyses compared jury ratings on:

³²⁴ Trial 6 versus Trial 9.

³²⁵ Trials 5 and 6.

³²⁶ Trials 7 and 10.

³²⁷ The additional two witnesses in Trial 7 yielded no significant differences, as is shown below, so no confounding influence affected these comparisons.

³²⁸ In half of the trials in each group, the judge gave standard jury directions plus specific directions on the appropriate use of tendency evidence, and in half of the trials, only standard instructions were given. These analyses were conducted without taking into account the specific jury directions on tendency evidence, since their effects were controlled across the overall number of trials.

- factual culpability of the defendant for non-penetrative and penetrative offences of the three witnesses/complainants (Simon, Timothy and Justin)
- convictions for counts alleged by the focal complainant with the moderately strong case.

In addition, a case study of jury deliberations in Trials 6 and 9 (Case Study 2) compared jury reasoning about the tendency evidence when the number of counts differed. These trials were appropriate for comparison because in both trials, the Crown called the same number of witnesses (four) and the jury groups received identical jury directions – that is, standard directions plus specialised directions on the use of tendency evidence. As such, these comparisons were not confounded by differences in the number of witnesses across trial groups.

We tested jury reasoning by accumulation of the witnesses in a number of ways. First, we analysed the results of quantitatively coding all jury deliberations, to examine the relationship between factual accuracy in deliberations and changes in the number of Crown witnesses. We compared factual accuracy in trials with two Crown witnesses (Trials 1, 2, 3 and 4) against trials with four Crown witnesses (Trials 5, 6 and 9) and trials with six Crown witnesses (Trials 7, 8 and 10).

Next, we quantitatively tested jury reasoning by accumulation of the witnesses in a total of 17 juries with 201 mock jurors who viewed a joint trial: eight juries ($n = 93$ jurors) attended a joint trial in which the Crown called a total of six witnesses; and nine juries ($n = 108$ jurors) viewed a joint trial in which the Crown called a total of four witnesses. The analyses compared jury ratings on:

- victim blame for cases of different evidential strength
- the defendant's factual culpability on six counts
- conviction rates for six counts

In addition, a case study of jury deliberations in Trials 7 and 9 (Case Study 3) compared jury reasoning in a joint trial when the number of Crown witnesses differed. These trials were appropriate for comparison because the Crown called four witnesses in Trial 7, and Trial 9 was identical except that the Crown called six witnesses. Both jury groups received the same jury directions: standard directions plus specialised directions on the use of tendency evidence. As

such, these comparisons were not confounded by differences in the number of counts across trial groups.

4.4.2 What is the influence of the number of counts?

Factual culpability of the defendant

Mock jurors in both trials rated the perceived factual culpability of the defendant for each of the four non-penetrative and two penetrative counts of child sexual abuse. Separate two-level regression analyses explored how mock jurors' *Child Sexual Abuse Knowledge*, their perceived convincingness of the complainants and the type of trial influenced findings of culpability. The perceived factual culpability of the defendant for the weak, moderately strong and strong counts is presented separately by count. Figure 8 shows the mean perceived factual culpability of the defendant for each alleged act, organised by trial type. The four allegations described by the complainants with weak and strong evidence were presented as counts only in the joint trial. If juries relied on more peripheral cues – such as the number of counts – in their reasoning and decision making, we expected that they would make fewer distinctions between complainants in the joint trial with six counts than in the separate trial with two counts, and that ratings of factual culpability for all three complainants in the joint trial would all be similar, including the defendant's culpability for the count by the complainant with the weak case.

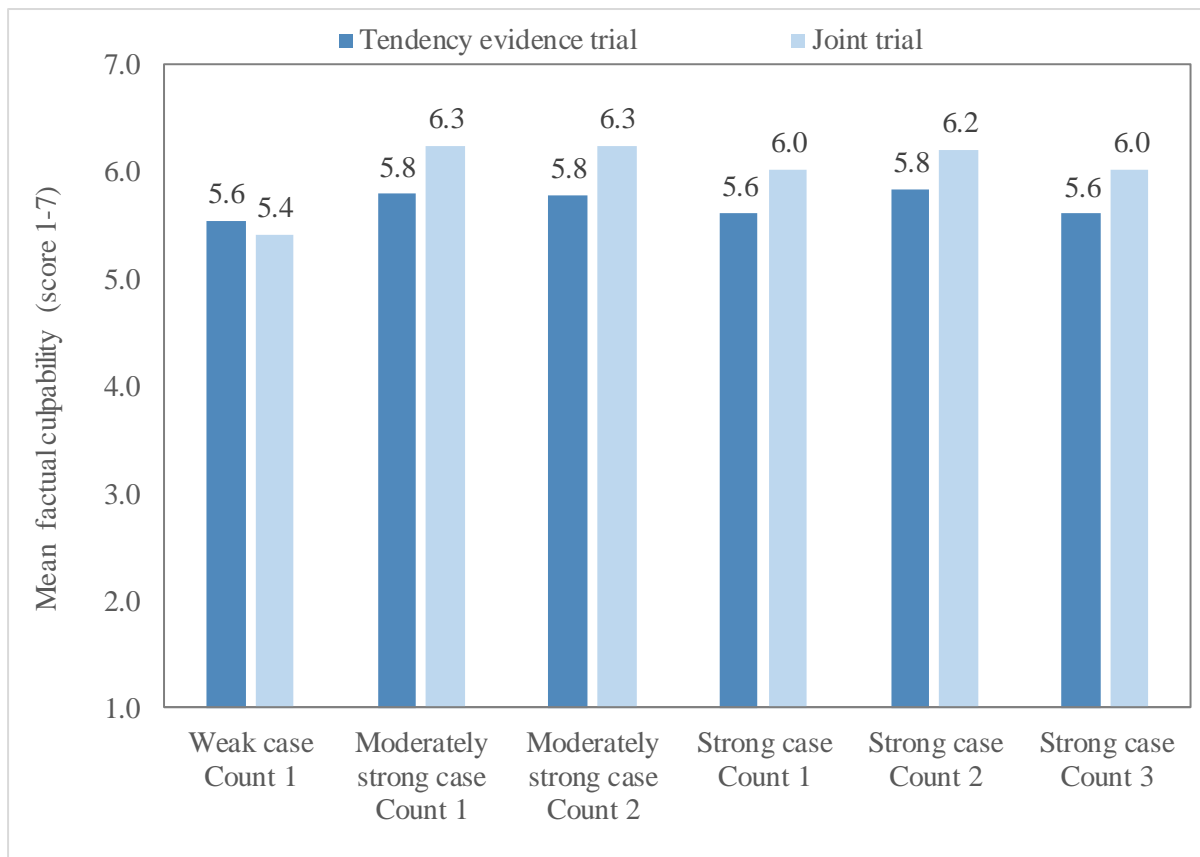


Figure 8. Mean perceived factual culpability for each allegation, by case strength and trial type

Factual culpability for the weak claim

The intra-class correlation revealed that the jury groups accounted for 14 per cent of the variance ($ICC = .140$), so multi-level analyses were necessary and the dependent measures could not be treated as independent.

Mock juror and mean jury *Child Sexual Abuse Knowledge* did not predict the finding of factual culpability of the defendant for the weak case.³²⁹ Ratings of the extent to which the complainant was perceived as convincing were significant, indicating that the more convincing Simon appeared, the more likely it was that mock jurors would find the defendant factually culpable.³³⁰ The total number of counts in the trial had no impact on the factual culpability of the defendant in the case involving the complainant with weak evidence.³³¹ Regardless of mock juror *Child Sexual Abuse Knowledge*, the perceived convincingness of the complainant was the

³²⁹ Juror level: $\beta = 0.009$, $SE = 0.009$, $Z = 0.953$, $p = .341$; jury level: $\beta = -0.024$, $SE = 0.066$, $Z = -0.373$, $p = .702$.

³³⁰ $\beta = 0.327$, $SE = 0.069$, $Z = 4.769$, $p < .001$.

³³¹ $\beta = -0.142$, $SE = 0.260$, $Z = -0.546$, $p = .585$.

only measure in this model that predicted the finding of factual culpability for the defendant in the weak case.

Factual culpability for the moderately strong claim

The intra-class correlation indicated a strong effect of group deliberation, that is, the jury groups were responsible for 23.0 per cent and 30.2 per cent of the variance in factual culpability of the defendant for the non-penetrative and penetrative offences, respectively.

Mock jurors and mean jury *Child Sexual Abuse Knowledge* did not predict factual culpability for the counts pertaining to the moderately strong case.³³² Ratings of the extent to which the complainant was perceived as convincing were significant, indicating that the more convincing Timothy appeared, the more likely mock jurors were to rate the defendant as factually culpable for both counts involving Timothy.³³³ No effect for the type of trial – that is, the total number of counts – emerged. Mock jurors' ratings of the defendant's factual culpability did not differ, regardless of the number of charges.³³⁴

Factual culpability for the strong claim

About one fifth of the variance in factual culpability of the defendant was due to the jury groups (masturbation of defendant: $ICC = .209$; masturbation of complainant: $ICC = .209$; oral-penile penetration: $ICC = .250$).

Mock juror or mean jury *Child Sexual Abuse Knowledge* did not predict the factual culpability of the defendant for the three counts involving Justin.³³⁵ By contrast, ratings of the extent to which the complainant was perceived as convincing were significant. The more convincing Justin appeared, the more likely mock jurors were to rate the defendant as factually culpable.³³⁶

³³² Non-penetrative offence: Juror level: $\beta = 0.011$, $SE = 0.008$, $Z = 1.414$, $p = .157$; jury level: $\beta = 0.054$, $SE = 0.057$, $Z = 0.942$, $p = .346$. Penetrative offence: Juror level: $\beta = 0.008$, $SE = 0.008$, $Z = 1.010$, $p = .313$; jury level: $\beta = 0.091$, $SE = 0.065$, $Z = 1.392$, $p = .164$.

³³³ Non-penetrative offence: $\beta = 0.377$, $SE = 0.059$, $Z = 6.387$, $p < .001$; penetrative offence: $\beta = 0.374$, $SE = 0.063$, $Z = 5.943$, $p < .001$.

³³⁴ Non-penetrative offence: $\beta = 0.430$, $SE = 0.238$, $Z = 1.808$, $p = .071$; penetrative offence: $\beta = 0.489$, $SE = 0.293$, $Z = 1.668$, $p = .095$.

³³⁵ Juror level: Masturbation of defendant: $\beta = 0.011$, $SE = 0.008$, $Z = 1.396$, $p = .163$; masturbation of complainant: $\beta = 0.011$, $SE = 0.008$, $Z = 1.325$, $p = .185$; oral-penile penetration: $\beta = 0.016$, $SE = 0.009$, $Z = 1.858$, $p = .063$. Jury level: Masturbation of defendant: $\beta = 0.045$, $SE = 0.055$, $Z = 0.829$, $p = .407$; masturbation of complainant: $\beta = 0.025$, $SE = 0.055$, $Z = 0.459$, $p = .646$; oral-penile penetration: $\beta = 0.071$, $SE = 0.062$, $Z = 1.148$, $p = .251$.

³³⁶ Masturbation of defendant: $\beta = 0.356$, $SE = 0.059$, $Z = 6.009$, $p < .001$; masturbation of complainant: $\beta = 0.368$, $SE = 0.063$, $Z = 5.844$, $p < .001$; oral-penile penetration: $\beta = 0.411$, $SE = 0.068$, $Z = 6.035$, $p < .001$.

Furthermore, the type of trial predicted factual culpability. Mock jurors perceived the defendant as more factually culpable for the two non-penetrative counts involving Justin when this evidence was assessed in a joint trial with six counts and two additional witnesses than when it was assessed in a separate trial with two counts.³³⁷

The influence of the total number of counts on verdict

To assess the influence on conviction rates of the total number of counts against the defendant, we considered verdicts rendered for the focal complainant with the moderately strong evidence (Timothy), as these were the counts common to both the separate trial with tendency evidence and the joint trial.

The conviction rate by juries in the joint trial was slightly higher than in the separate trial with tendency evidence: 87.5 per cent and 75 per cent versus 58.8 per cent, as shown in Table 14. Less than one-fifth of the juries in the separate trial with tendency evidence trial were hung (17.6 per cent and 11.8 per cent). Chi-square analyses of jury verdicts – convictions versus acquittals plus hung juries – revealed no difference in conviction rates in the tendency evidence trial compared to the joint trial.³³⁸

Table 14. Jury and mock juror verdicts for the non-penetrative and penetrative offences of the moderately strong complainant, by type of trial (per cent)

	Count 1: Masturbation of complainant			Count 2: Digital-anal penetration		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
Tendency evidence trial	23.5	58.8	17.6	29.4	58.8	11.8
Joint trial	12.5	87.5	0.0	25.0	75.0	0.0
(b) Juror verdict						
Tendency evidence trial	26.3	73.7	–	27.8	72.2	–
Joint trial	9.2	90.8	–	13.9	86.1	–

Note. Tendency evidence trial: 17 juries, $n = 197$ jurors; joint trial: 16 juries, $n = 176$ jurors.

³³⁷ Masturbation of defendant: $\beta = 0.486$, $SE = 0.228$, $Z = 2.133$, $p = .033$; masturbation of complainant: $\beta = 0.497$, $SE = 0.229$, $Z = 2.170$, $p = .030$; oral-penile penetration: $\beta = 0.454$, $SE = 0.280$, $Z = 1.624$, $p = .104$.

³³⁸ Non-penetrative offence: $\chi^2 = 5.14$, $p = .023$, $\Phi = .535$; penetrative offence: $\chi^2 = 5.14$, $p = .023$, $\Phi = .535$.

Two-level regression analysis was conducted on mock juror verdicts. Mock juror *Child Sexual Abuse Knowledge* predicted the verdict for the penetrative offense only.³³⁹ The extent to which the complainant was convincing also predicted the verdict:³⁴⁰ mock jurors who perceived the complainant as more convincing were twice as likely to convict the defendant compared to mock jurors who did not. We found no effect on verdict of mean jury *Child Sexual Abuse Knowledge*³⁴¹ or trial type³⁴² for either count shown in Table 14. Hence, the number of charges had no effect on the verdict.

³³⁹ Non-penetrative offence: $\beta = 0.052$, $SE = 0.029$, $Z = 1.765$, $p = .077$, Odds Ratio = 1.053, 95% CI [0.994; 1.115]; penetrative offence: $\beta = 0.056$, $SE = 0.028$, $Z = 2.030$, $p = .042$, Odds Ratio = 1.058, 95% CI [0.985; 1.116].

³⁴⁰ Non-penetrative offence: $\beta = 0.803$, $SE = 0.165$, $Z = 4.862$, $p < .001$, Odds Ratio = 2.233, 95% CI [1.615; 3.086]; penetrative offence: $\beta = 0.808$, $SE = 0.155$, $Z = 5.220$, $p < .001$, Odds Ratio = 2.243, 95% CI [1.565; 3.038].

³⁴¹ Non-penetrative offence: $\beta = 0.022$, $SE = 0.331$, $Z = 0.066$, $p = .947$; penetrative offence: $\beta = 0.076$, $SE = 0.404$, $Z = 0.189$, $p = .850$.

³⁴² Non-penetrative offence: $\beta = 2.470$, $SE = 1.405$, $Z = 1.759$, $p = .079$; penetrative offence: $\beta = 2.570$, $SE = 1.696$, $Z = 1.515$, $p = .130$.

Case Study 2: Deliberations about tendency evidence

This qualitative analysis compared how the number of counts affected jury reasoning in deliberations. Trials 6 and 9 were selected for comparison as both trials included four Crown witnesses, while the number of counts against the defendant differed (two versus six, respectively).

The impact of four fewer counts for juries to consider meant that the time spent by mock jurors scrutinising Timothy's evidence for Counts 1 and 2 in the trial with two counts (Trial 6) was greater than the time spent on Timothy's evidence for Counts 5 and 6 in the trial with six counts (Trial 9), these being identical counts in both trials. In other words, in the trial with two counts, the ambiguities in Timothy's evidence consumed more deliberation time than they did in the joint trial where juries had four additional counts to consider. With six counts to consider in the joint trial, the amount of deliberation time per count was limited, and deliberation time was divided among the evidence of the three complainants, although not necessarily equally. Not only did all jury deliberations in the joint trial begin with a consideration of the evidence of the complainant in the weak case (Simon), but most of the deliberation time was spent analysing Simon's evidence (which gave rise to Count 1). This was because, of the three complainants, Simon's evidence created more unresolved questions and doubts for juries.

By contrast, most deliberations in the separate trial with tendency evidence began with a consideration of Timothy's evidence, and there was much less scrutiny of the evidence from the two other witnesses (Simon and Justin). As Juror 3 in Jury 27³⁴³ said, "But [Timothy's] the one that we have to focus on". Another feature of the deliberations in the separate trial with tendency evidence was a lack of clarity about the significance of Justin's and Simon's evidence, as if mock jurors did not understand how they would ultimately use their evidence. Mock jurors tended to discuss the evidence from Simon, Timothy and Justin together, jumping from one to the other, rather than making a clear delineation between the evidence, as they did when deliberating about the joint trial.

The following two exchanges illustrate the difference in the degree of scrutiny that was applied to Timothy's evidence in the separate trial with tendency evidence versus the joint trial. In Jury 27,³⁴⁴ the mock jurors are questioning the finer details of Timothy's evidence:

³⁴³ Trial 6, separate trial with tendency evidence, standard and tendency evidence jury directions.

³⁴⁴ Ibid.

JUROR 2: There is more doubt. How would Timothy know that Mark had two places, if he had only been there once? How can you say, “We went to his other place”?

JUROR 10: He didn’t know necessarily but we know -----

JUROR 2: But he said that.

JUROR 10: But we know [for] certain from Mrs Samuels.

JUROR 2: How do you know that she was saying the truth? How do you know they weren’t colluding? No one – asked -----

JUROR 10: Okay, okay, I agree with you ... but the defence did not raise that. The defence kept quiet about Mrs Samuels ...

JUROR 2: You completely overlooked the fact that I said “how would Timothy know that Mark had two places?”

JUROR 10: How would he not know if he’s been to the one place first and not the other place? He’s been to both places, he would know about both places.

JUROR 2: He didn’t say that.

JUROR 10: Yes, he did.

JUROR 2: Did he? Anyone else? Did anyone know that he had been to two places?

JUROR 6: I don’t recall him mentioning how many times he had been to Mark’s place -----

JUROR 10: But he had been more than once.

JUROR 6: Well, I don’t recall him saying that.

JUROR 3: Yeah, I don’t think -----

JUROR 10: Okay.

JUROR 6: I do recall – I did write down that he went for three days; or what intended to be three days -----

JUROR 10: Yes, he did. He ran away.

JUROR 6: ----- and he stayed for one night. But I don’t have written down, and I don’t think it is mentioned, whether he had been there prior to that ... two properties.

Compared to the above extract, when Jury 68³⁴⁵ discussed Timothy’s evidence in the joint trial, the mock jurors gave it little scrutiny. By the time the jurors came to discuss the specific charges relating to Timothy, they had already made general observations on a number of issues including the plausibility of children remembering significant events (for example, a sexual

³⁴⁵ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

assault) while forgetting peripheral details (such as location, time and dates); their disbelief in the defence's claim that the complainants were motivated for financial and/or personal reasons to make a false complaint; and their belief it was counterintuitive for the defendant to be coaching without a child of his own playing on the team. They had also already discussed Simon's evidence (failing to reach a unanimous decision) and Justin's evidence (voting guilty on all three counts), and had touched on the use of tendency evidence. Accordingly, when they came to Timothy, they said very little before proceeding to apply the tendency evidence ("Plus, you know, it is three separate counts as well") and finding the defendant guilty:

JUROR 2: Count 5: "act of indecency towards a person under the age of 16 years, namely Timothy Lyons, between 1st and 31st December 1997; guilty or not guilty?"

JUROR 8: So that was when they watched the movie -----

JUROR 12: The sleepover.

(Most jurors voiced agreement).

JUROR 4: Guilty.

(Most jurors voiced agreement).

JUROR 4: He knew there was a TV on the cupboard.

JUROR 8: Yeah.

JUROR 4: "I didn't have a TV." The woman said that she did have one; so he [the defendant] was lying. Guilty.

(Most jurors voiced agreement) ...

JUROR 2: And sexual intercourse of Timothy Lyons -----

JUROR 4: Guilty.

JUROR 12: In a sleepover, nobody knows what happened apart from those two.

JUROR 4: That's right. So it is hearsay and he's making the complaint and the other guy looks like he's got form. So, therefore, you would probably have to go with the child's -----

JUROR 12: Plus, you know, it is three separate counts as well.

JUROR 4: Yeah, so therefore, guilty.

One of the ways in which jury susceptibility to the effect of accumulation prejudice is believed to operate is by lowering the standard of proof, and the transcribed deliberations were coded to identify instances where juries adopted this strategy. Examination of the deliberations of juries in Trials 6 and 9 revealed no evidence of this strategy in either group. However, jurors on Jury

88³⁴⁶, deliberating on six counts, thought that they could apply a lower standard of proof to the accused because of the delay in complaint. Due to a misunderstanding of the judge's direction on delay in complaint, this was not a by-product of the number of counts before the jury in that trial:

JUROR 11: I think the biggest tension for me is the delay in the complaint and it's pretty much what the judge – this is my interpretation of what the judge said – is that you have got to apply a lower standard of guilt, a level of guilt, than you normally would.

...

JUROR 1: Well, you have to look at the evidence. There is the Briginshaw standard, which is the famous case, where it was – where something is really high, like rape/murder, you look at a high level. What you are saying is quite correct. We are coming down to, like, risk ----

JUROR 11: Yeah.

JUROR 1: ---- factor. So it is determination of risk, "Is the person capable of doing this? Is this person not capable of doing this?" And that is sort of what we are looking at, with what you are saying.

JUROR 11: Yeah, a lot of the evidence that the complainant has produced, they couldn't remember the dates; they couldn't remember the place ----

JUROR 12: But there are two ----

JUROR 11: Well, that's fine, but the point I am making is that this is about someone's guilt, that we are deciding ----

JUROR 12: Someone's life, yes.

JUROR 11: ---- and that last section, the judge talked about – the delay in complaint says that we are meant to apply lower standard, essentially. That's my interpretation, unless anyone else ----

JUROR 1: No, you are right.

³⁴⁶ Id.

4.4.3 What is the influence of the number of witnesses on jury reasoning?

Factual accuracy in deliberations

Preliminarily, we examined the extent to which factual accuracy was related to the overall number of witnesses in a trial. The factual recall outcomes reported in Part 4.2 were derived from multiple-choice questions about the case of the complainant with the moderately strong evidence, since that evidence was common to all trials. Coding the transcripts of jury deliberations allowed us to supplement those analyses by examining all factual recall errors jurors made in the course of group deliberations.

We conducted an analysis to determine whether juries that observed more prosecution witnesses were more likely to confuse the facts than juries that observed fewer witnesses. We compared error rates in jury deliberations in Trials 1, 2, 3 and 4 (hearing prosecution evidence from two witnesses, Timothy and Ellen) with those of the juries in Trials 5, 6 and 9 (who heard prosecution evidence from four witnesses, Timothy, Ellen and two tendency witnesses, Simon and Justin) and with jurors in Trials 7, 8 and 10 (who heard evidence from Timothy and Ellen; the two tendency witnesses/complainants, Simon and Justin; and two additional supporting witnesses for Justin – a total of six witnesses).

Overall, in more than one-third of the juries (36.6 per cent), none of the jurors made a factual error. The proportion of error-free juries attending joint trials with six prosecution witnesses was equivalent to that of juries attending separate trials with two or four prosecution witnesses. In 40 per cent of the trials, the deliberations revealed one factual error; more than twice as many of these errors arose in separate trials (47 per cent) than in joint trials (21 per cent). In 15.5 per cent of the juries, two factual errors were noted, and 7.7 per cent of all juries made three or four factual errors in deliberation.

As is shown in Figure 9, more jurors attending trials with six prosecution witnesses made between two and four factual errors than jurors attending trials with two or four prosecution witnesses. As more witnesses were introduced into the trial (whether it was a separate or a joint trial) and the trial evidence increased in complexity, mock jurors were more likely to make factual recall errors in deliberation.³⁴⁷

³⁴⁷ Kendall's Tau-b = -0.53, $p = .01$.

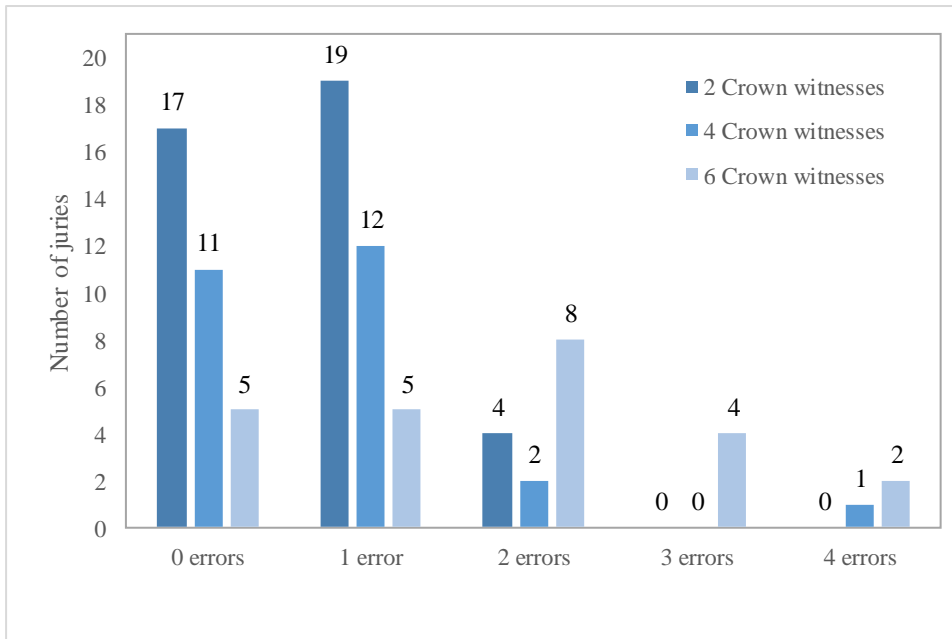


Figure 9. Number of factual recall errors in deliberation, by total number of prosecution witnesses

The deliberation transcripts also provided an opportunity to assess the extent to which peers corrected one juror’s factual recall errors. There was no statistically significant difference in the number of uncorrected or persistent errors depending on the type of trial. So while individual jurors who were exposed to more witnesses had less accurate factual recall than those exposed to less evidence from fewer witnesses, their fellow jurors were capable of policing these instances, and these errors were corrected within the jury groups.

How does the number of witnesses influence jury decision making?

To test whether juror reasoning in a joint trial differed based on the total number of witnesses, we varied the total number of prosecution witnesses in a joint trial from four to six.³⁴⁸

The analysis reported below includes responses from a total of 17 juries and 201 mock jurors who viewed the joint trial. In all, eight juries ($n = 93$ jurors) attended a joint trial in which the Crown called a total of six witnesses, and nine juries ($n = 108$ jurors) viewed a joint trial in which the Crown called a total of four witnesses. The two supporting witnesses for the complainant with the strong claims (Justin) were omitted in Trial 9. Both witnesses had verified that at the time of the alleged abuse, Justin displayed uncharacteristic behaviour that indicated he was distressed, and one of the witnesses confirmed that Justin had told him about the abuse

³⁴⁸ Trial 7 versus Trial 9.

at the time it occurred. In both trials, the juries received the same jury directions, including a tendency evidence direction.

As a preliminary matter, we conducted analyses to ensure the equivalence of the two trials at the outset. There were no differences in mock juror or jury pre-trial biases between the two trials on any pre-trial measures (the CSA-KQ, FEEBS or PJAQ, including the separate factors within each of these measures). Similarly, participants' gender distribution, age and education were comparable in the two trials. Furthermore, mock jurors and juries in the two groups did not differ in their recollection of case facts presented during the trial, as measured by the six multiple choice questions on the post-trial questionnaire.

To examine whether the juries were influenced by the total number of witnesses, we compared both juries' ratings of:

- victim blame
- the defendant's factual culpability for the offences
- convictions for all offences.

Victim blame for counts of different strength

To examine whether the juries engaged in more peripheral or heuristic reasoning in a trial with more prosecution witnesses, we examined the inferences mock jurors and juries made about the extent to which each complainant was responsible for what happened to him. If juries used more peripheral cues – such as the number of witnesses – in their reasoning and decision making, we expected that they would make fewer distinctions between complainants in the joint trial with six witnesses than in the joint trial with four witnesses, and that ratings of victim blame would all be similarly lower in the trial with six witnesses. We conducted separate multi-level regression analyses for each complainant to test the effect on victim blame of (a) mock juror and mean jury *Child Sexual Abuse Knowledge*, and (b) the number of prosecution witnesses.

Victim blame for the weak claim

Mock jurors' *Child Sexual Abuse Knowledge* predicted victim blame in the weak case. Specifically, mock jurors with more accurate *Child Sexual Abuse Knowledge* were less likely than their counterparts with less accurate knowledge to blame the victim for the alleged offences.³⁴⁹ However, at the jury level, neither the mean jury *Child Sexual Abuse Knowledge*

³⁴⁹ $\beta = -0.033$, $SE = .009$, $Z = -3.670$, $p < .001$.

nor the number of witnesses for the prosecution influenced victim blame (joint trial with four Crown witnesses: $M = 2.15$; joint trial with six Crown witnesses: $M = 2.23$).³⁵⁰ The intra-class correlation indicated that less than 1 per cent of the variance was attributable to the jury group ($ICC = .007$).

Victim blame for the moderately strong claim

Mock jurors' *Child Sexual Abuse Knowledge* had a significantly negative effect on victim blame in the moderately strong case, such that jurors with more accurate *Child Sexual Abuse Knowledge* were less likely to blame the victim.³⁵¹ This effect was not significant at the jury level. The number of witnesses for the prosecution did not predict victim blame in the moderately strong case (joint trial with four Crown witnesses: $M = 1.87$; joint trial with six Crown witnesses: $M = 1.97$).³⁵² The intra-class correlation was .007, indicating that less than 1 per cent of the variance was attributable to the jury groups.

Victim blame for the strong claims

Mock jurors with greater *Child Sexual Abuse Knowledge* were less likely to blame the complainant with the strong case³⁵³, although this effect disappeared at the jury level.³⁵⁴ Similarly, the number of supporting witnesses called by the prosecution in the strong case did not predict victim blame.³⁵⁵ Accordingly, victim blame did not differ between the joint trial with four Crown witnesses ($M = 1.87$) and the joint trial with six Crown witnesses ($M = 2.05$).³⁵⁶ The intra-class correlation for the strong case was .003, indicating that less than 1 per cent of the variance was attributable to the jury groups.

In other words, regardless of how many supporting witnesses the prosecution called in the strong case, mock jurors were equally likely to blame the complainant. Only mock jurors' *Child Sexual Abuse Knowledge*, which seemed to be unrelated to the jury groups (as was indicated by the intra-class correlation) significantly decreased the extent of victim blame.

³⁵⁰ Child Sexual Abuse Knowledge: $\beta = 0.018$, $SE = 0.039$, $Z = 0.474$, $p = .636$; number of witnesses: $\beta = 0.014$, $SE = 0.147$, $Z = 0.093$, $p = .926$. Joint trial with four Crown witnesses: $Std Dev = 1.23$; joint trial with six Crown witnesses: $Std Dev = 1.21$.

³⁵¹ Juror level: $\beta = -0.038$, $SE = 0.008$, $Z = -4.923$, $p < .001$; jury level: $\beta = 0.017$, $SE = 0.028$, $Z = 0.614$, $p = .539$.

³⁵² $\beta = 0.022$, $SE = 0.147$, $Z = 0.148$, $p = .883$. Joint trial with four Crown witnesses: $Std Dev = 1.09$; joint trial with six Crown witnesses: $Std Dev = 1.23$.

³⁵³ $\beta = -0.023$, $SE = 0.009$, $Z = -2.414$, $p = .016$.

³⁵⁴ $\beta = -0.008$, $SE = 0.035$, $Z = -0.236$, $p = .814$.

³⁵⁵ $\beta = 0.106$, $SE = 0.144$, $Z = 0.740$, $p = .460$.

³⁵⁶ Joint trial with four Crown witnesses: $Std Dev = 1.30$; joint trial with six Crown witnesses: $Std Dev = 1.40$.

Factual culpability of the defendant

We examined the factual culpability of the defendant and then verdicts on all six counts to assess whether the total number of Crown witnesses had an impact on jury reasoning and decisions.

Figure 10 shows the factual culpability ratings of the defendant for each of the six non-penetrative and penetrative counts of sexual abuse in Trial 6 versus Trial 9.

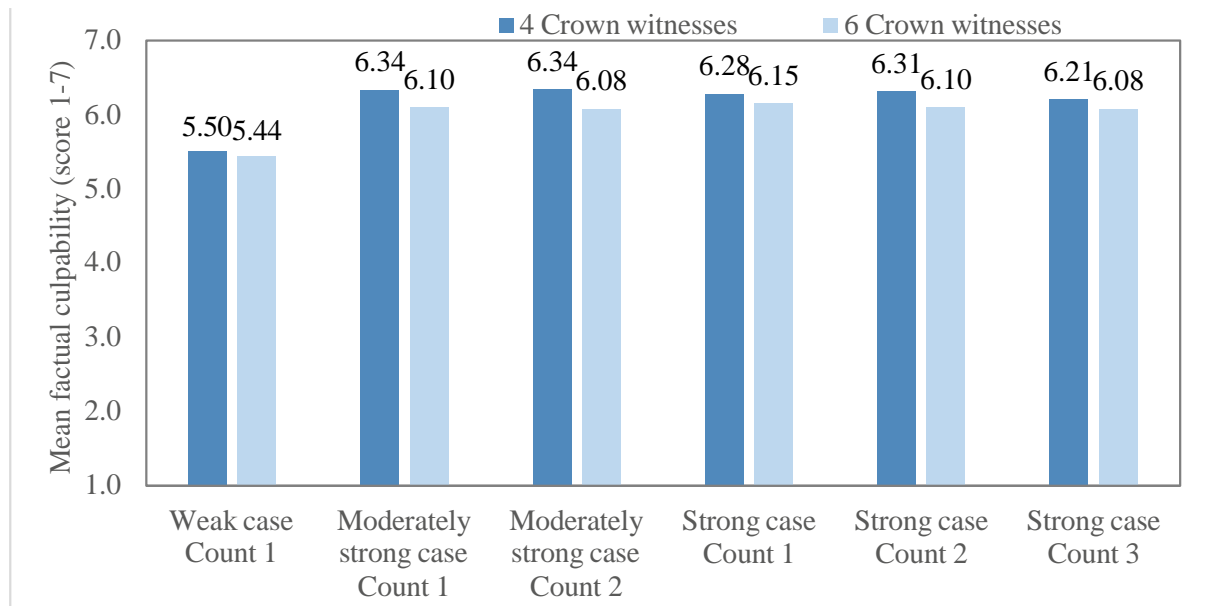


Figure 10. Mean perceived factual culpability in the joint trial with four versus six witnesses.

Results are presented separately for the two-level regression analyses testing the influence of mock jurors' *Child Sexual Abuse Knowledge*, the perceived convincingness of the complainants, and type of trial on perceived factual culpability for the weak, moderately strong and strong cases.

Factual culpability for the weak claim

A total of 15.8 per cent of the variance in the perceived factual culpability of the defendant for the non-penetrative offence was attributed to the jury groups and remained unexplained by other measures.

Mock jurors' *Child Sexual Abuse Knowledge* did not predict the factual culpability of the defendant at the juror or the jury level.³⁵⁷ However, the extent to which the complainant, Simon, was perceived to be convincing did predict perceptions of factual culpability.³⁵⁸ The more mock jurors perceived the complainant as convincing, the more likely they were to perceive the defendant as factually culpable. The total number of Crown witnesses had no impact on the perceived factual culpability of the defendant for the non-penetrative count involving Simon.³⁵⁹

Factual culpability for the moderately strong claims

Two-level analysis indicated that 15.5 per cent of the variance for the non-penetrative offence and 12.9 per cent of the variance for the penetrative offence was unexplained and could be attributed to the jury groups.

Mock jurors' *Child Sexual Abuse Knowledge* did not predict the factual culpability of the defendant at either the juror or the jury levels, unrelated to the type of offence.³⁶⁰ However, factual culpability was predicted by how convincing mock jurors perceived the complainant, Timothy, to be. The more mock jurors perceived the complainant to be convincing, the more likely they were to perceive the defendant as factually culpable on both counts (a non-penetrative and a penetrative offence).³⁶¹ The number of Crown witnesses had no impact on the factual culpability of the defendant in this case.³⁶²

Factual culpability for the strong claims

Less than 10 per cent of the variance for the non-penetrative offences (8.4 per cent) and the penetrative offence (6.2 per cent) could be attributed to the jury groups, indicating that juries had a relatively small impact on the perceived factual culpability of the defendant for the three counts involving Justin. Mock jurors' *Child Sexual Abuse Knowledge* did not predict the factual culpability of the defendant on any of the three counts.³⁶³ This effect remained non-significant

³⁵⁷ Juror level: $\beta = 0.023$, $SE = 0.013$, $Z = 1.711$, $p = .087$; jury level: $\beta = -0.027$, $SE = 0.074$, $Z = -1.716$, $p = .086$.

³⁵⁸ $\beta = 0.402$, $SE = 0.069$, $Z = 5.876$, $p < .001$.

³⁵⁹ $\beta = -0.155$, $SE = 0.334$, $Z = -0.464$, $p = .643$.

³⁶⁰ Non-penetrative offence: Juror level: $\beta = 0.014$, $SE = 0.010$, $Z = 1.302$, $p = .193$; jury level: $\beta = 0.007$, $SE = 0.057$, $Z = 0.122$, $p = .903$. Penetrative offence: Juror level: $\beta = 0.014$, $SE = 0.008$, $Z = 1.660$, $p = .097$; jury level: $\beta = 0.010$, $SE = 0.051$, $Z = 0.203$, $p = .839$.

³⁶¹ Non-penetrative offence: $\beta = 0.270$, $SE = 0.085$, $Z = 3.198$, $p = .001$; penetrative offence: $\beta = 0.232$, $SE = 0.081$, $Z = 2.846$, $p = .004$.

³⁶² Non-penetrative offence: $\beta = -0.284$, $SE = 0.258$, $Z = -1.103$, $p = .270$; penetrative offence: $\beta = -0.293$, $SE = 0.244$, $Z = -1.200$, $p = .230$.

³⁶³ Masturbation of defendant: $\beta = 0.012$, $SE = 0.011$, $Z = 1.047$, $p = .295$; masturbation of complainant: $\beta = 0.012$, $SE = 0.011$, $Z = 1.135$, $p = .256$; oral-penile penetration: $\beta = 0.017$, $SE = 0.010$, $Z = 1.725$, $p = .084$.

at the jury level based on the mean jury *Child Sexual Abuse Knowledge* score.³⁶⁴ Perceived convincingness was again the only significant predictor of factual culpability; the more mock jurors perceived Justin to be convincing, the more they found the defendant to be factually culpable on all three counts.³⁶⁵ The two additional Crown witnesses called to support Justin's evidence had no independent impact on the factual culpability of the defendant for the three counts.³⁶⁶

Verdicts for counts of different strength

A comparison of the proportion of guilty verdicts returned for each count in the joint trials with either four or six prosecution witnesses indicated that individual mock juror verdicts on all six counts were higher in the trial with fewer witnesses than they were in the trial with more witnesses. The overall proportion of guilty verdicts was greater for the complainants with moderately strong and strong evidence than for the complainant with the weak evidence, as shown in Table 15. Moreover, the overall conviction rate was lower for penetrative offences than for non-penetrative offences. Juries treated the allegations of each complainant separately. Chi-square analysis of the jury verdicts comparing convictions with acquittals plus hung juries revealed no significant differences in the conviction rates of juries in trials with four versus six prosecution witnesses, unrelated to the case strength or the type of offence.³⁶⁷

³⁶⁴ Masturbation of defendant: $\beta = -0.018$, $SE = 0.051$, $Z = -0.361$, $p = .718$; masturbation of complainant: $\beta = -0.023$, $SE = 0.049$, $Z = -0.483$, $p = .629$; oral-penile penetration: $\beta = 0.000$, $SE = 0.044$, $Z = -0.010$, $p = .992$.

³⁶⁵ Masturbation of defendant: $\beta = 0.296$, $SE = 0.090$, $Z = 3.292$, $p = .001$; masturbation of complainant: $\beta = 0.296$, $SE = 0.096$, $Z = 3.082$, $p = .002$; oral-penile penetration: $\beta = 0.294$, $SE = 0.086$, $Z = 3.405$, $p = .001$.

³⁶⁶ Masturbation of defendant: $\beta = -0.165$, $SE = 0.219$, $Z = -0.756$, $p = .450$; masturbation of complainant: $\beta = -0.233$, $SE = 0.212$, $Z = -1.097$, $p = .273$; oral-penile penetration: $\beta = -0.206$, $SE = 0.212$, $Z = -0.971$, $p = .331$.

³⁶⁷ $p > .05$.

Table 15. Jury and mock juror verdicts for each count, in joint trials with six or four Crown witnesses (per cent)

	Joint trial 6 Crown witnesses			Joint trial 4 Crown witnesses		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
Weak: masturbation of complainant	62.5	37.5	0.0	33.3	22.2	44.4
Moderate: Masturbation of complainant	25.0	75.0	0.0	0.0	100.0	0.0
Moderate: Digital- anal penetration	25.0	75.0	0.0	0.0	100.0	0.0
Strong: Masturbation of defendant	12.5	87.5	0.0	0.0	100.0	0.0
Strong: Masturbation of complainant	0.0	100.0	0.0	0.0	100.0	0.0
Strong: Oral-penile penetration	25.0	75.0	0.0	0.0	100.0	0.0
(b) Juror verdict						
Weak: Masturbation of complainant	46.7	53.3	–	38.3	61.7	–
Moderate: Masturbation of complainant	11.8	88.2	–	1.9	98.1	–
Moderate: Digital- anal penetration	12.0	88.0	–	4.7	95.3	–
Strong: Masturbation of defendant	11.8	88.2	–	1.9	98.1	–
Strong: Masturbation of complainant	3.3	96.7	–	1.9	98.1	–
Strong: Oral-penile penetration	12.9	87.1	–	4.7	95.3	–

Note. Joint trial with six Crown witnesses: eight juries, $n = 93$ jurors; joint trial with four Crown witnesses: nine juries, $n = 108$ jurors.

Further analyses assessed the extent to which individual mock jurors' verdicts were consistent or distinguished between the allegations of the three complainants. These analyses measured the degree to which a mock jurors voted the same way in relation to similar counts. Conviction rates for the same offence type – namely, masturbation of the complainant – revealed that individual mock jurors distinguished between the same offence type depending on the strength of the evidence presented. A total of 35.7 per cent of the mock jurors reached different verdicts regarding the same offence type (non-penetrative) alleged by the complainants with the moderately strong versus the weak cases. Forty per cent of individual mock jurors reached

different verdicts regarding the same offence type (non-penetrative) alleged by the complainants with the strong versus the weak cases. Finally, 4 per cent of the mock jurors reached different verdicts regarding the same type of non-penetrative offence alleged by the complainant with the strong versus the moderately strong claims. In sum, in the context of a joint trial, mock jurors distinguished between similar allegations of different evidential strength, and were sensitive to variations in the strength of the evidence presented by the weak, moderate and strong claims respectively.

The small proportion of ‘not guilty’ verdicts for the counts pertaining to the moderately strong and strong evidence ($n = 1$ to $n = 12$) precluded logistic regression analyses for these counts. We conducted a multi-level logistic regression analysis only for the weak case.

Verdict for the weak claim

Neither mock juror³⁶⁸ nor mean jury³⁶⁹ *Child Sexual Abuse Knowledge* predicted the verdict for Count 1. Ratings of the complainant’s perceived convincingness significantly predicted verdicts for this claim.³⁷⁰ The odds of conviction were 1.7 times greater when mock jurors perceived the complainant as more convincing. In line with perceived factual culpability, the number of witnesses called by the Crown (four versus six) had no influence on the verdict for Count 1.³⁷¹

³⁶⁸ $\beta = 0.000$, $SE = 0.024$, $Z = 0.002$, $p = .998$, Odds Ratio = 1.000, 95% CI [0.954; 1.048].

³⁶⁹ $\beta = -0.291$, $SE = 0.236$, $Z = -1.235$, $p = .217$.

³⁷⁰ $\beta = 0.543$, $SE = 0.120$, $Z = 4.535$, $p < .001$, Odds Ratio = 1.722, 95% CI [1.361; 2.178].

³⁷¹ $\beta = -0.528$, $SE = 1.179$, $Z = -0.448$, $p = .654$.

Case Study 3: Deliberations about a joint trial³⁷²

In the trial with four Crown witnesses³⁷³, the conviction rate was not significantly different from that in the trial with six Crown witnesses³⁷⁴. This is a counterintuitive finding, since the latter contained the evidence of two additional supporting witnesses for the complainant with the strong case (Justin). We expected that removing the evidence of the two supporting witnesses for this complainant would reduce the strength of Justin's allegations against the defendant. Beginning with Trial 7, we examined deliberations in joint trials with four versus six Crown witnesses, to assess their reasoning in the different joint trials.

Deliberations in a joint trial with six Crown witnesses³⁷⁵

Contrary to lawyers' and judges' expectations, as were outlined in Chapter 2, there was no evidence of impermissible propensity reasoning in the joint trial deliberations. Each jury in Trial 7 considered the six counts separately, and all began with Count 1 (except for Jury 85³⁷⁶ which decided Count 1 was too hard to start with). It was only after each jury had considered each count separately that they turned to the issue of tendency evidence, if at all, as per the judge's tendency direction.

During their deliberations, each jury considered only the evidence that pertained to each particular count. Count 1 took up most of the deliberation time. Jurors expressed a great deal of dissatisfaction about the weaknesses in the evidence of the complainant with the weak case (Simon) and the lack of evidence supporting his allegations – in particular, Simon's uncertainty about the date he was in the soccer team, the pool he was pushed into (the defendant denied he had a pool), the thunderstorm (disputed by the defence), whether it was feasible that Simon had escaped the defendant's home by running out with only a t-shirt on, and the fact that Simon's mother did not give evidence.

Some juries did not consider the tendency evidence issue at all, while others did not understand how tendency evidence worked despite the judge's tendency direction. In particular, the existence of two other complainants was not enough to persuade most juries that the defendant was guilty of the count involving Simon (Count 1). For example, Jury 83³⁷⁷ initially decided

³⁷² Trials 7 versus Trial 9.

³⁷³ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

³⁷⁴ Trial 7, joint trial with standard and tendency evidence jury directions.

³⁷⁵ Ibid.

³⁷⁶ Id.

³⁷⁷ Id.

that it could not use the evidence of the defendant's sexual interest in Timothy and Justin in relation to Count 1 because the defendant did not have a tendency at the time of the first incident. Jury 75 also raised this issue.³⁷⁸ After re-reading the judge's tendency direction, the members of Jury 83 realised they *could* use the tendency evidence with respect to Count 1, but decided that that evidence was not sufficient to remove their reasonable doubt regarding the commission of the sexual act that was the subject of Count 1.

Jury 84³⁷⁹ also disregarded the tendency evidence; Juror 4 recognised that "we have put in a few 'not guilty' verdicts so it's hard to get that tendency evidence in". While six jurors in Jury 85³⁸⁰ recognised that there was a pattern of behaviour by the defendant, it was not sufficient to remove their doubt regarding Simon's evidence, so six jurors voted 'not guilty' for Count 1. By contrast, in Jury 64³⁸¹ the tendency evidence was sufficient to remove the doubt of the one dissenting juror for Count 1, and the jury delivered a unanimous 'guilty' verdict.

Jury 75³⁸² provides a good example of how juries avoided propensity reasoning:

JUROR 9: The evidence by Simon didn't justify ----

JUROR 3: I don't think we can say it's beyond reasonable doubt ...

JUROR 9: Could you use the tendency evidence, as you have convicted that – or you are possibly saying 'guilty' on those other counts, of 2, 3, 4 ----

JUROR 3: That we think he had a sexual interest in the other two boys, so ----

JUROR 9: And that you can say there was a pattern of behaviour, that he was grooming these young boys.

JUROR 12: But it says here 'not to do that'.

... [THE JURY RE-READS THE JUDGE'S TENDENCY DIRECTION]

JUROR 3: But he didn't have a pattern because that was the first one.

Jury 75 decided they could only consider the tendency evidence after deciding that the time, place and sexual act pertaining to Count 1 were proved beyond reasonable doubt, rather than using the evidence of the defendant's sexual interest in Timothy and Justin "in determining whether the accused committed the offences against Simon Rutter". The tendency evidence

³⁷⁸ Id.

³⁷⁹ Id.

³⁸⁰ Id.

³⁸¹ Id.

³⁸² Id.

was irrelevant to their final decision because after considering each separate element of Count 1, Jury 75 decided the defendant was guilty of sexually assaulting Simon.

*Deliberations in a joint trial with four Crown witnesses*³⁸³

Deliberations in the trial with four Crown witnesses also revealed that mock jurors distinguished between the strength of each complainant's evidence. Juries interpreted Simon's evidence as the weakest compared to the evidence provided by Justin and Timothy. For example, in Jury 67³⁸⁴, Juror 12 observed that "As the witnesses went on, they kind of slightly became more credible" and "emotional". For this jury, Simon's story had too many holes: the non-existent pool, the thunderstorm and running to his mother's car half-dressed. Similarly, Jury 70³⁸⁵ very quickly decided that the only complainant whose story created a 'seed of doubt' was Simon, because of the non-existent swimming pool and the fact that he was confused about when he joined the soccer team (March or August). As a result, the jury voted 'not guilty' on Count 1, with very little reference to the other complainants' evidence – the tendency evidence did not influence their decision. Jury 70 determined that because Simon's allegations were about events that took place after a shower and lunch – whereas Justin and Timothy's allegations were about events at sleepovers at the defendant's house – the evidence about these events was not sufficiently similar to establish a pattern of behaviour by the defendant.

In the absence of the two supporting witnesses, Justin's case remained strong since none of his evidence gave rise to the ambiguities that characterised Simon's evidence. There were several *factual* aspects of Simon's evidence about which the defence raised doubts, but the defence was not able to do so in relation to Justin's evidence. For example, after discussing the sexual acts involving Justin (Counts 2, 3 and 4), Jury 70 voted guilty with little deliberation, deciding that they had 'no doubt' in believing Justin's evidence as a result of the defendant's grooming, the secrecy surrounding the acts of abuse, Justin's fear and his marked change in behaviour. Jury 70 was also influenced by the fact that the defendant had cut off contact with each family after the abuse and, with no family of his own, had no apparent reason to coach boys' soccer teams.

In Jury 87³⁸⁶, Juror 11 was more explicit about the differences between Simon's and Justin's evidence: "because this guy hasn't got a steady job, whereas the other guy is a financial

³⁸³ Trial 9, joint trial with four prosecution witnesses, and standard and tendency evidence jury directions.

³⁸⁴ *Ibid.*

³⁸⁵ *Id.*

³⁸⁶ *Id.*

director, we are going to believe him more”. Compared to the evidence given by Simon, Justin’s description of sexual abuse was ‘so typical’ of jurors’ common sense knowledge about child sexual abuse as the following exchange shows:

JUROR 9: I have to say that Justin’s description of what took place, and the things that were said/done, and the things that he said that Mark said, it just rang true to me that that ... was the sort of rubbish that would occur in that situation ...

JUROR 7: That’s what you hear time and time again, actually, that conversation; that ‘it’s okay’.

...

JUROR 9: The interesting thing about it was, when the defence said, “What did you do?” to the witness, and he said was – one of the things was, “I froze. I didn’t know what to do. I just lay there ...”

JUROR 7: That is just typical.

JUROR 9: It is so believable. ... When someone’s of authority and power and influence, that’s what they do.

JUROR 7: Of course.

JUROR 5: That’s very natural.

Jury 69³⁸⁷ made similar comments about the believability of Justin’s evidence, with Juror 6 commenting “There wasn’t anything factual that was disputed with him”. As a result, Jury 69 was more willing to dismiss the doubts that the defence tried to raise in relation to Justin’s evidence compared to Simon’s evidence, thus dismissing the defence allegation that Justin was jealous of the defendant’s friendship with his father.

Compared to Simon’s evidence, juries made fewer comments about the ambiguities in relation to Timothy’s evidence (his mistake regarding the movie he watched with the defendant, whether he was abused in a bedroom with a the TV on a chest of drawers, and the defendant’s second house) because Timothy had a supporting witness who gave evidence that she owned a house with a TV on a chest of drawers in the bedroom to which the defendant had access, and because mock jurors generally accepted that a child of 11 could misremember the name of a movie. For example, in Jury 87³⁸⁸, Timothy’s evidence was considered to be ‘the most compelling’ because he ran away after being sexually abused and locked the doors to his house.

³⁸⁷ Id.

³⁸⁸ Id.

When cross-examined, Timothy was 'pretty sure' about being sexually abused in a bedroom with a TV, which, for this jury, amounted to 'overwhelming evidence' that the act of abuse occurred.

4.5 The influence of jury directions on jury reasoning and decision making

In this section we examine whether there is any support for judicial assumptions about the effectiveness of jury directions in reducing impermissible reasoning. Contrary to those assumptions, the findings are in line with a large body of empirical studies demonstrating the ineffectiveness of most jury directions. Few differences arose in our systematic analysis of jury reasoning and decisions in relationship evidence, tendency evidence and joint trials, where juries received standard versus specific directions on the use of relationship evidence and tendency evidence. The relationship evidence direction was more effective than the tendency evidence direction, which produced no apparent benefits, irrespective of whether it was provided in a separate or a joint trial. Analysis of the deliberations revealed that error rates in using the context evidence and the tendency evidence were unaffected by the presence of these directions. Jurors devoted more time in deliberation to discussing the standard of ‘beyond reasonable doubt’ than tendency directions. There was no indication that the tendency direction benefited juries’ reasoning processes.

The context evidence direction helped juries overcome their reluctance to convict for the penetrative offence, but conviction rates for the non-penetrative offence were unaffected – although factual culpability ratings on both counts increased significantly in the presence of the direction. Consistent with the findings reported in Section 4.2, the convincingness of the complainant predicted conviction rates. In addition, higher *Child Sexual Abuse Knowledge* on the part of individual jurors – rather than jury directions – predicted convictions for penetrative offences in trials with tendency evidence. In both separate and joint trials with tendency evidence, the judge’s tendency evidence direction had no significant influence on the verdict, or the perceived criminal intent or factual culpability of the defendant.

Jurors’ self-reported measures revealed that juries who received context directions as opposed to standard directions perceived the judge’s instructions as more confusing; found the task of assessing witness credibility and applying the law more difficult; reported a higher cognitive load; and found it harder to understand the charges, recall the facts, weigh the evidence and assess the case for the prosecution. Similarly, compared to the standard directions, jurors rated tendency evidence directions as more difficult to understand, and perceived that they increased the cognitive load. However, jurors also rated the charges as easier to understand when they received the tendency evidence directions in a joint trial.

4.5.1 Research aim

In Part 4.2, we assessed the influence of the type of trial on jury reasoning and verdict by providing all juries with the same standard jury directions, to avoid confounding the type of jury direction with the type of trial. In real-world trials, context evidence or tendency evidence is typically accompanied by specific jury directions providing guidance on the appropriate use of that evidence. The analyses presented in this section aimed to assess whether specific jury directions on context evidence and tendency evidence influenced jury reasoning. The hypothesis tested, based on judicial expectations, was that verdicts based on impermissible reasoning would decline when juries were given the relationship evidence and tendency evidence directions.

To test this hypothesis, we conducted separate analyses for (a) relationship evidence trials, (b) tendency evidence trials and (c) joint trials. First, we assessed the influence of the jury direction on context evidence by comparing the relationship evidence trial with standard jury directions to the same trial with a context evidence direction.³⁸⁹ Next, we analysed the influence of the presence or absence of a tendency evidence direction in a separate trial, by comparing the tendency evidence trial with standard jury directions against the same trial with a tendency evidence direction.³⁹⁰ In these trials, Timothy was the complainant and Simon and Justin were the tendency evidence witnesses. Third, we used a series of parallel analyses to examine a similar question in a joint trial with three complainants, by comparing a joint trial with standard jury directions and the same trial with a tendency evidence direction.³⁹¹ In these trials, Timothy, Simon and Justin were all complainants. The dependent measures presented in each set of analyses included:

- the criminal intent of the defendant
- the factual culpability of the defendant
- verdicts
- self-reported cognitive effort to understand the charges
- self-reported cognitive effort to understand the judge's instructions
- quantitative results of errors about jury directions in deliberations.

³⁸⁹ Trial 2 versus Trial 3.

³⁹⁰ Trial 5 versus Trial 6.

³⁹¹ Trial 7 versus Trial 10.

4.5.2 What is the influence of context evidence directions?³⁹²

We used responses from a total of 21 juries and 238 mock jurors to conduct these analyses. In all, nine juries ($n = 103$ jurors) viewed the trial without the context evidence direction and 12 juries ($n = 135$ jurors) viewed the trial with the context evidence direction.

We have reported results for measures of the defendant's criminal intent, factual culpability and verdict, for the complainant with moderately strong evidence (Timothy). We also looked at impressions of the judicial directions, perceived cognitive effort and qualitative analysis of jury deliberations.

Criminal intent of the defendant

The jury direction on context evidence had no significant effect on the perceived criminal intent of the defendant (context evidence directions absent: $M = 4.61$; context evidence directions present: $M = 4.97$).³⁹³

Factual culpability of the defendant

Two-level regression analyses revealed that mock jurors' *Child Sexual Abuse Knowledge* predicted the factual culpability of the defendant for the non-penetrative offence but not the penetrative offence.³⁹⁴ The more mock jurors knew about child sexual abuse, the more likely they were to find the defendant factually culpable for the non-penetrative offence.

Furthermore, the perceived convincingness of the complainant predicted factual culpability, unrelated to the offence type³⁹⁵; that is, the more mock jurors perceived the complainant to be convincing, the more likely they were to find the defendant factually culpable on both counts.

The effect on factual culpability of the mock jurors' *Child Sexual Abuse Knowledge* disappeared at the jury level.³⁹⁶ However, the type of judicial direction was a significant predictor. Juries that received the context evidence direction were more likely to find the defendant factually culpable for the non-penetrative offence ($M = 5.51$) and the penetrative

³⁹² Trials 2 versus Trial 3.

³⁹³ $\beta = 0.362$, $SE = 0.225$, $Z = 1.423$, $p = .155$, $ICC = .056$; context evidence directions absent: $Std Dev = 1.49$; context evidence directions present: $M = 1.57$.

³⁹⁴ Non-penetrative offence: $\beta = 0.024$, $SE = 0.012$, $Z = 1.998$, $p = .046$; penetrative offence: $\beta = 0.023$, $SE = 0.014$, $Z = 1.586$, $p = .113$.

³⁹⁵ Non-penetrative offence: $\beta = 0.480$, $SE = 0.091$, $Z = 5.248$, $p < .001$; penetrative offence: $\beta = 0.518$, $SE = 0.093$, $Z = 5.566$, $p < .001$.

³⁹⁶ Non-penetrative offence: $\beta = -0.015$, $SE = 0.066$, $Z = -0.227$, $p = .821$; penetrative offence: $\beta = -0.019$, $SE = 0.068$, $Z = 0.282$, $p = .778$.

offence ($M = 5.47$) than juries who did not (non-penetrative offence: $M = 4.91$; penetrative offence: $M = 4.75$).³⁹⁷

The intra-class correlation for factual culpability was 0.114 for the non-penetrative offence and 1.115 for the penetrative offence, indicating that the jury groups were responsible for 11 per cent of the variance.

Verdict

Table 16 shows jury and mock juror verdicts for each of the two counts. Chi-square analyses indicated that juries were more likely to convict the defendant on the penetrative but not the non-penetrative offence when they were exposed to context evidence directions.³⁹⁸ These analyses revealed a higher proportion of acquittals when juries received the standard jury directions compared to standard plus context evidence directions.

Table 16. Jury and mock juror verdicts for the moderate strength case in the relationship evidence trial, by type of jury directions (per cent)

	Count 1: Masturbation of complainant			Count 2: Digital-anal penetration		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
Standard directions	58.3	8.3	33.3	58.3	0.0	41.7
Standard + context evidence directions	22.2	33.3	44.4	22.2	33.3	44.4
(b) Juror verdict						
Standard directions	77.8	22.2	–	80.7	19.3	–
Standard + context evidence directions	31.1	68.9	–	30.1	69.9	–

Note. Context evidence directions absent: 12 juries, $n = 135$ jurors; context evidence directions present: nine juries, $n = 103$ jurors.

We conducted two-level regression analysis on mock juror verdicts. There was no evidence that mock juror or mean jury *Child Sexual Abuse Knowledge* had an effect on the verdict for

³⁹⁷ Non-penetrative offence: $\beta = 0.628$, $SE = 0.305$, $Z = 2.059$, $p = .040$; context evidence directions absent: $Std Dev = 1.56$; context evidence directions present: $Std Dev = 1.61$. Penetrative offence: $\beta = 0.745$, $SE = 0.309$, $Z = 2.410$, $p = .016$; context evidence directions absent: $Std Dev = 1.58$; context evidence directions present: $Std Dev = 1.62$.

³⁹⁸ Non-penetrative offence: $\chi^2 = 2.09$, $p = .149$, $\Phi = .315$; penetrative offence: $\chi^2 = 4.67$, $p = .031$, $\Phi = .471$.

either count shown in Table 16.³⁹⁹ By contrast, the extent to which the complainant was convincing did predict the verdict⁴⁰⁰: the odds of conviction were 3.0 times greater when mock jurors perceived the complainant as more convincing. Furthermore, the presence of the context evidence direction increased the conviction rate on both counts.⁴⁰¹

When considering jury (instead of juror) verdicts, the proportion of hung juries increased in the presence of the context evidence direction. Even though individual mock juror conviction rates increased for both counts in the trial that included a context evidence direction, the jury verdicts did not reflect this influence.

Jury impressions of the judge's instructions

In general, on most measures of mock juror perceptions of the judge's directions we found no differences between the two trials. There was no difference in ratings of the helpfulness of the directions in understanding the case⁴⁰²; how easy it was to understand the directions⁴⁰³; whether the directions changed mock jurors' minds about their verdicts⁴⁰⁴; or the extent to which the judge's instructions were fair to the defendant.⁴⁰⁵ By contrast, juries that received the context evidence direction rated the judge's instructions as significantly more confusing than did the juries that received the standard jury directions.⁴⁰⁶ Figure 11 presents mean mock jurors' perceptions of the judge's instructions by trial.

³⁹⁹ Juror level: Non-penetrative offence: $\beta = 0.006$, $SE = 0.033$, $Z = 0.188$, $p = .851$, Odds Ratio = 1.006, 95% CI [0.943; 1.073]; penetrative offence: $\beta = 0.006$, $SE = 0.032$, $Z = 0.184$, $p = .854$, Odds Ratio = 1.060, 95% CI [0.945; 1.071]. Jury level: Non-penetrative offence: $\beta = 0.040$, $SE = 0.323$, $Z = 0.124$, $p = .902$; penetrative offence: $\beta = 0.131$, $SE = 0.334$, $Z = 0.393$, $p = .694$.

⁴⁰⁰ Non-penetrative offence: $\beta = 1.045$, $SE = 0.199$, $Z = 5.263$, $p < .001$, Odds Ratio = 2.843, 95% CI [1.927; 4.196]; penetrative offence: $\beta = 1.178$, $SE = 0.289$, $Z = 4.075$, $p < .001$, Odds Ratio = 3.248, 95% CI [1.843; 5.725].

⁴⁰¹ Non-penetrative offence: $\beta = 4.602$, $SE = 1.613$, $Z = 2.853$, $p = .004$; penetrative offence: $\beta = 5.538$, $SE = 1.904$, $Z = 2.908$, $p = .004$.

⁴⁰² $\beta = -0.007$, $SE = 0.176$, $Z = -0.037$, $p = .970$, $ICC = .043$.

⁴⁰³ $\beta = -0.094$, $SE = 0.187$, $Z = -0.502$, $p = .616$, $ICC = .034$.

⁴⁰⁴ $\beta = 0.212$, $SE = 0.310$, $Z = 0.684$, $p = .494$, $ICC = .065$.

⁴⁰⁵ $\beta = 0.172$, $SE = 0.240$, $Z = 0.717$, $p = .473$, $ICC = .023$.

⁴⁰⁶ $\beta = 0.562$, $SE = 0.196$, $Z = 2.859$, $p = .004$, $ICC = .044$.

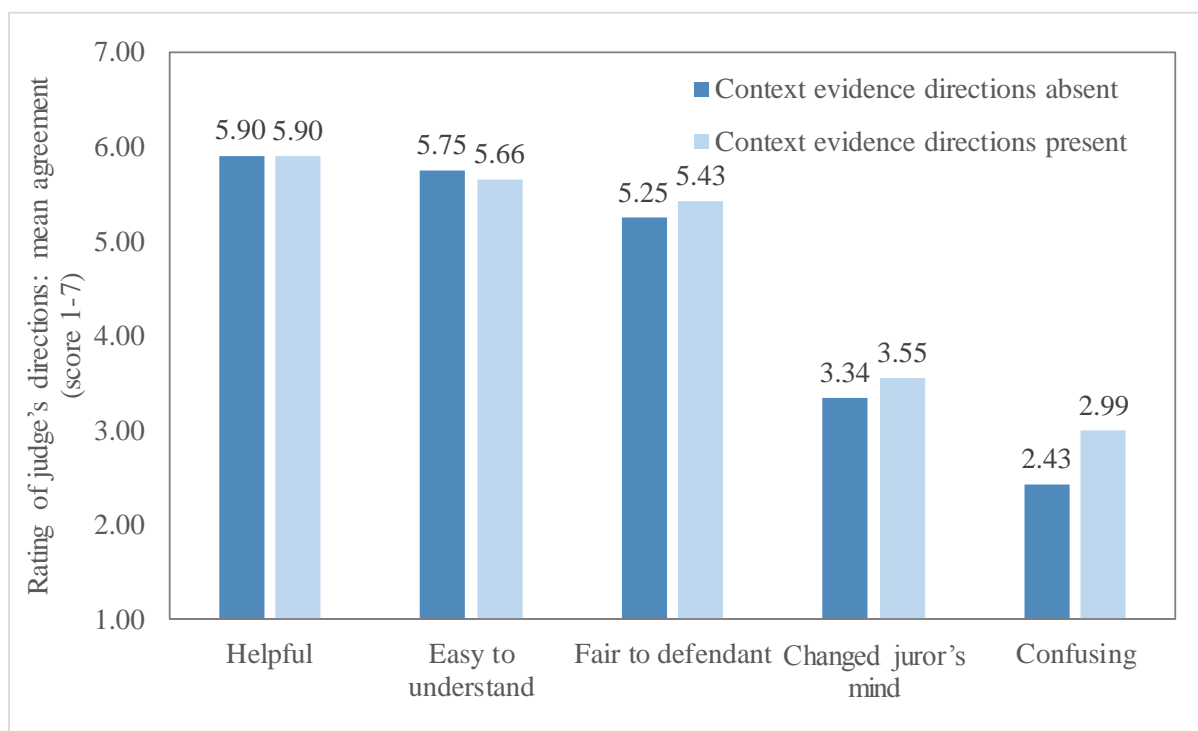


Figure 11. Mock jurors' perceptions of the judicial instructions, with and without context evidence directions

Self-reported cognitive effort with context direction

We found that the type of direction (standard versus context evidence) had no effect on mock jurors' overall reported cognitive effort.⁴⁰⁷ When we examined specific questions about mental effort expended, mock jurors who received the context evidence direction reported that they expended significantly more cognitive effort on two tasks compared to those jurors who did not receive this judicial direction. Those tasks were: assessing the credibility of witnesses (context evidence direction present: $M = 5.33$; absent: $M = 4.93$)⁴⁰⁸ and applying the law to the facts (context evidence direction present: $M = 5.44$; absent: $M = 5.01$).⁴⁰⁹ There were no differences on any other questions that addressed perceived cognitive effort.

Quantitative analysis of errors in applying the context evidence direction

A legally trained research assistant recorded the number of times that juries were correct or incorrect in their understanding of context evidence, and in applying the context evidence. The same proportion of juries used the evidence correctly and incorrectly, independent of the

⁴⁰⁷ $\beta = 0.297$, $SE = 0.167$, $Z = 1.770$, $p = .076$, $ICC = .016$.

⁴⁰⁸ $\beta = 0.401$, $SE = 0.206$, $Z = 1.944$, $p = .052$, $ICC = .029$; context evidence direction present: $Std Dev = 1.46$, absent: $Std Dev = 1.69$.

⁴⁰⁹ $\beta = 0.431$, $SE = 0.224$, $Z = 1.924$, $p = .054$, $ICC = .032$; context evidence direction present: $Std Dev = 1.46$, absent: $Std Dev = 1.66$.

presence of this direction. Furthermore, there was no difference in the proportion of deliberation time juries devoted to a discussion of the judge's directions, whether or not they had the guidance on context evidence.

Despite the non-significant increase in the conviction rate in Trial 3 compared to Trial 2, juries in Trial 3 made very little reference to the judge's direction regarding the use of the relationship evidence in their deliberations. None of the juries who referred to the judge's direction made any link between it and their final verdicts.

In four juries (26⁴¹⁰, 50⁴¹¹, 56⁴¹² and 57⁴¹³), there was no mention of the judge's direction about how to use the relationship evidence. Two juries (25⁴¹⁴ and 49⁴¹⁵) referred only once to the judge's direction about the relationship evidence, with little elaboration. For example, Jury 25 merely said:

JUROR 11: The equipment room was just a background story.

JUROR 8: To give it context.

(Most agreed).

JUROR 1: It was just to establish a pattern of behaviour, that he might have been grooming this young boy for something else.

By contrast, three juries (47⁴¹⁶, 48⁴¹⁷ and 55⁴¹⁸) discussed their understanding of the judge's direction, although they did not link their interpretations to their final verdicts. In particular, Jury 47 used the judge's direction to remind jurors not to engage in tendency reasoning:

JUROR 6: Well, I believe that he groomed -----

JUROR 12: That's not the count -----

JUROR 2: That's not what he's being charged on.

JUROR 6: He took advantage of the situation.

JUROR 2: That's not what he's being charged on.

JUROR 6: I know that.

⁴¹⁰ Trial 3, separate trial with relationship evidence, and standard and context evidence jury directions.

⁴¹¹ Ibid.

⁴¹² Id.

⁴¹³ Id.

⁴¹⁴ Id.

⁴¹⁵ Id.

⁴¹⁶ Id.

⁴¹⁷ Id.

⁴¹⁸ Id.

JUROR 2: But that's what we have got to go on.

JUROR 6: I know. He groomed the guy, had access to the property; took the kid in there that night and -----

JUROR 1: The judge told you not to -----

JUROR 6: Not to what?

JUROR 1: Not to pay attention to what transpired before. All you are interested in is the actual two acts that occurred in that -----

JUROR 6: Right.

Similarly, Jury 55⁴¹⁹ specifically referred to the judge's direction when it identified that it could not use the relationship evidence to establish a tendency on the part of the accused:

JUROR 3: The modelling one is another one. That is not actually considered in Count 1.

(Most jurors voiced disagreement).

JUROR 5: The judge said not to have that -----

JUROR 10: Don't let that influence.

JUROR 5: ----- our assessment/deliberation on this because that is the other acts and it's been mentioned that it is not really to impact on our decision, really.

[...]

JUROR 5: No, the modelling stuff happened before and we are not -----

JUROR 2: Yeah, and that's to -----

JUROR 5: ----- no, we are not to look at that at all.

JUROR 2: Oh.

JUROR 8: She actually said that is contextual -----

JUROR 2: Okay.

JUROR 8: To make you realise that this didn't happen completely out of the blue -----

JUROR 2: Okay.

⁴¹⁹ Id.

4.5.3 What is the influence of tendency evidence directions in a separate trial?⁴²⁰

This study included a total of 17 juries and 197 mock jurors. In all, nine juries ($n = 112$ jurors) received the tendency evidence direction and eight juries ($n = 85$ jurors) received standard jury directions only.

We report results of outcomes for measures of the criminal intent of the defendant, factual culpability and verdict for the complainant with moderately strong evidence (Timothy), and in addition, impressions of the judicial directions, perceived cognitive effort, and qualitative analysis of jury deliberations.

Criminal intent of the defendant

The presence of standard directions versus tendency evidence jury directions made no difference in the perceived criminal intent of the defendant.

Factual culpability of the defendant

Two-level regression analysis revealed that mock juror' and juries' *Child Sexual Abuse Knowledge* had no effect on their perception of the defendant's factual culpability.⁴²¹ The extent to which the complainant was perceived as convincing increased ratings of factual culpability⁴²², although the tendency evidence direction did not significantly increase ratings of the defendant's factual culpability for the two counts.⁴²³ The mean perceived factual culpability of the defendant was $M = 5.92$ and $M = 5.77$ for the non-penetrative offence, and $M = 5.58$ and $M = 5.64$ for the penetrative offence, when tendency evidence directions were absent and present, respectively.⁴²⁴ Intra-class correlations revealed that jury groups had a strong effect on findings of factual culpability (non-penetrative offence: $ICC = .243$; penetrative offence: $ICC = .310$).

⁴²⁰ Trial 5 versus Trial 6.

⁴²¹ Non-penetrative offence: Juror level: $\beta = 0.003$, $SE = 0.013$, $Z = 0.230$, $p = .818$; jury level: $\beta = 0.010$, $SE = 0.114$, $Z = 0.087$, $p = .931$. Penetrative offence: Juror level: $\beta = 0.000$, $SE = 0.013$, $Z = -0.028$, $p = .977$; jury level: $\beta = 0.089$, $SE = 0.148$, $Z = 0.602$, $p = .547$.

⁴²² Non-penetrative offence: $\beta = 0.332$, $SE = 0.066$, $Z = 5.043$, $p < .001$; penetrative offence: $\beta = 0.326$, $SE = 0.069$, $Z = 4.750$, $p < .001$.

⁴²³ Non-penetrative offence: $\beta = -0.119$, $SE = 0.384$, $Z = -0.308$, $p = .758$; penetrative offence: $\beta = 0.287$, $SE = 0.542$, $Z = 0.529$, $p = .596$.

⁴²⁴ Non-penetrative offence: Tendency evidence directions absent: $Std Dev = 1.36$, present: $Std Dev = 1.46$; penetrative offence: tendency evidence directions absent: $Std Dev = 1.55$, present: $Std Dev = 1.63$.

Verdict

Table 17 shows jury and mock juror verdicts recorded in response to the type of jury directions presented by the judge. When tendency evidence direction was provided, the proportion of hung juries increased and the conviction rate declined, but chi-square analyses showed no significant differences in verdicts between jury groups that received the standard versus the tendency evidence direction.⁴²⁵

Table 17. Jury and mock juror verdicts for the moderate strength case with tendency evidence, by type of jury directions (per cent)

	Count 1: Masturbation of complainant		Count 2: Digital-anal penetration	
	Not Guilty	Hung jury	Not Guilty	Hung jury
(a) Jury verdict				
Standard directions				
Standard + tendency evidence directions	25.0	62.5	12.5	37.5
	22.2	55.6	22.2	22.2
(b) Juror verdict				
Standard directions				
Standard + tendency evidence directions	28.6	71.4	–	33.6
	24.5	75.5	–	23.6

Note. Tendency evidence directions absent: eight juries, $n = 85$ jurors; tendency evidence directions present: nine juries, $n = 112$ jurors.

Two-level regression analysis revealed that mock jurors' *Child Sexual Abuse Knowledge* predicted verdicts for the penetrative offence but not the non-penetrative offence.⁴²⁶ The odds of conviction for the penetrative offence were 1.1 times greater when mock jurors had more *Child Sexual Abuse Knowledge*. However, mean jury *Child Sexual Abuse Knowledge* did not predict verdict.⁴²⁷ By contrast, perceived convincingness of the complainant predicted verdicts on both counts⁴²⁸; mock jurors who perceived the complainant to be more convincing were

⁴²⁵ $p > .05$.

⁴²⁶ Non-penetrative offence: $\beta = 0.056$, $SE = 0.035$, $Z = 1.603$, $p = .109$, Odds Ratio = 1.058, 95% CI [0.988; 1.133]; penetrative offence: $\beta = 0.073$, $SE = 0.036$, $Z = 2.026$, $p = .043$, Odds Ratio = 1.076, 95% CI [1.002; 1.155].

⁴²⁷ Non-penetrative offence: $\beta = -0.664$, $SE = 0.657$, $Z = -1.011$, $p = .312$; penetrative offence: $\beta = -0.832$, $SE = 0.823$, $Z = -1.011$, $p = .312$.

⁴²⁸ Non-penetrative offence: $\beta = 0.696$, $SE = 0.156$, $Z = 4.461$, $p < .001$, Odds Ratio = 2.007, 95% CI [1.478; 2.725]; penetrative offence: $\beta = 0.765$, $SE = 0.187$, $Z = 4.095$, $p < .001$, Odds Ratio = 2.148, 95% CI [1.490; 3.098].

twice as likely to convict. However, the tendency evidence direction had no impact on mock juror verdicts compared to standard judicial directions.⁴²⁹

Jury impressions of the judge’s instructions

The type of judicial direction provided had no effect on the perceived helpfulness of the judicial instructions.⁴³⁰ Furthermore, the tendency evidence direction had no influence on the perceived fairness of the judicial direction to the defendant.⁴³¹

In contrast, mock jurors were more likely to agree that the judge changed their mind about the verdict when a tendency evidence direction was included in the judicial summing-up than when it was not.⁴³² Figure 12 shows mock juror perceptions of the judicial direction.

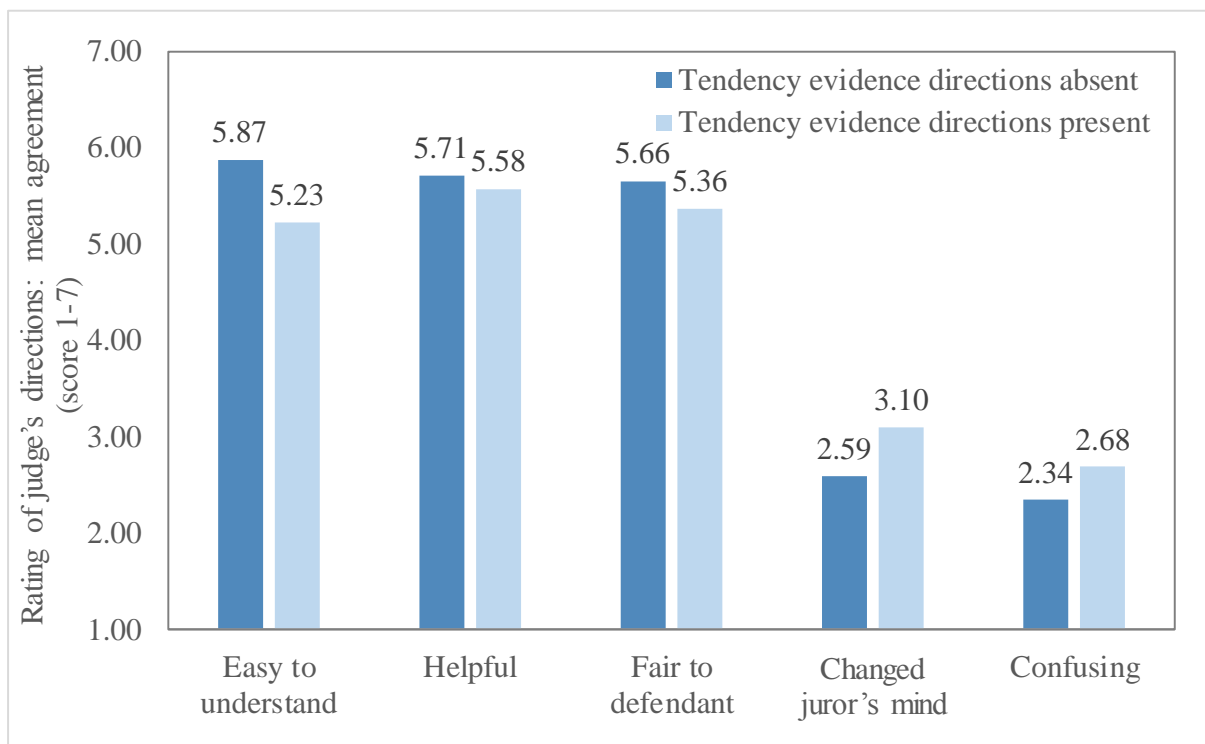


Figure 12. Mock jurors’ perceptions of the judge’s instructions with and without tendency evidence directions

⁴²⁹ Non-penetrative offence: $\beta = -2.412$, $SE = 2.464$, $Z = -0.979$, $p = .328$; penetrative offence: $\beta = -2.836$, $SE = 3.151$, $Z = -0.757$, $p = .449$.

⁴³⁰ $\beta = -0.385$, $SE = 0.218$, $Z = -1.762$, $p = .078$, $ICC = .076$.

⁴³¹ $\beta = -0.298$, $SE = 0.240$, $Z = -1.240$, $p = .215$, $ICC = .019$.

⁴³² $\beta = 0.514$, $SE = 0.243$, $Z = 2.115$, $p = .034$, $ICC = .022$.

Self-reported cognitive effort with tendency evidence direction

Juries perceived that they required increased cognitive effort when they received tendency evidence directions than when they received standard judicial directions.⁴³³ Specifically, mock jurors reported expending higher cognitive effort to understand the charges against the defendant (tendency evidence direction present: $M = 5.07$, absent: $M = 4.31$)⁴³⁴; remember the case facts (tendency evidence direction present: $M = 5.23$, absent: $M = 4.74$)⁴³⁵; weigh the evidence (tendency evidence direction present: $M = 5.56$, absent: $M = 5.17$)⁴³⁶; and evaluate the prosecution case (tendency evidence direction present: $M = 5.46$, absent: $M = 4.15$).⁴³⁷

There was no difference in mock jurors' assessments of the cognitive effort required to assess the credibility of the witnesses or to understand the judicial instructions.

4.5.4 What is the influence of directions on tendency evidence in a joint trial?⁴³⁸

These analyses involved a total of 16 juries and 176 mock jurors. In all, eight juries ($n = 93$ jurors) viewed trials in which the judge provided standard directions plus a tendency evidence direction, and eight other juries ($n = 83$ jurors) viewed trials in which only standard jury directions were provided.

Criminal intent of the defendant

Our analyses showed no difference in the perceived criminal intent of defendant when only standard directions were provided compared to those trials where the jury received standard plus tendency evidence directions (tendency evidence directions present: $M = 6.28$, absent: $M = 6.05$).⁴³⁹

⁴³³ $\beta = 0.379$, $SE = 0.176$, $Z = 2.147$, $p = .032$, $ICC = .019$.

⁴³⁴ $\beta = 0.766$, $SE = 0.240$, $Z = 3.199$, $p = .001$, $ICC = .038$; tendency evidence direction present: $Std Dev = 1.80$, absent: $Std Dev = 2.16$.

⁴³⁵ $\beta = 0.486$, $SE = 0.239$, $Z = 2.033$, $p = .042$, $ICC = .023$; tendency evidence direction present: $Std Dev = 1.52$, absent: $Std Dev = 1.81$.

⁴³⁶ $\beta = 0.392$, $SE = 0.197$, $Z = 1.985$, $p = .047$, $ICC = .016$; tendency evidence direction present: $Std Dev = 1.45$, absent: $Std Dev = 1.73$.

⁴³⁷ $\beta = 0.306$, $SE = 0.155$, $Z = 1.980$, $p = .048$, $ICC = .010$; tendency evidence direction present: $Std Dev = 1.57$, absent: $Std Dev = 1.64$.

⁴³⁸ Trial 7 versus Trial 10.

⁴³⁹ $\beta = -0.225$, $SE = 0.265$, $Z = -0.849$, $p = .396$, $ICC = .076$; tendency evidence direction present: $Std Dev = 1.60$, absent: $Std Dev = 1.07$.

Factual culpability of the defendant

We assessed the jurors' perception of factual culpability of the defendant for each of the six counts. Variables that might enable a prediction of this perception were mock jurors' *Child Sexual Abuse Knowledge*, the perceived convincingness of the complainants, and the type of jury direction (standard directions, versus standard plus tendency evidence directions). Results are presented separately for each complainant.

Factual culpability for the weak claim

Mock jurors' *Child Sexual Abuse Knowledge* and perceptions of the complainant's convincingness increased the perception of the defendant's factual culpability for the counts in the weak case ($ICC = .037$).⁴⁴⁰ Juries' *Child Sexual Abuse Knowledge* was not significant, indicating that the effect of mock jurors' *Child Sexual Abuse Knowledge* disappeared at the jury level.⁴⁴¹ There were no differences in perceived factual culpability between juries that received standard directions and those that received standard plus tendency evidence directions (tendency evidence directions present: $M = 5.44$, absent: $M = 5.40$).⁴⁴²

Factual culpability for the moderately strong claims

Two-level regression analyses revealed that mock juror perceptions of the complainant's convincingness was the sole predictor that impacted the perceived culpability of the defendant: The extent to which the complainant was perceived as convincing increased factual culpability ratings for both the non-penetrative and the penetrative offence.⁴⁴³

Mock juror and mean jury *Child Sexual Abuse Knowledge* did not predict the factual culpability of the defendant.⁴⁴⁴ Similarly, the tendency evidence direction had no impact on factual culpability, unrelated to the type of count, that is, the difference in perceived factual culpability between the trial with and without tendency evidence instructions was negligible (non-

⁴⁴⁰ Knowledge about child sexual abuse: $\beta = 0.025$, $SE = 0.011$, $Z = 2.283$, $p = .022$; Convincingness: $\beta = 0.456$, $SE = 0.069$, $Z = 6.656$, $p < .001$.

⁴⁴¹ $\beta = -0.035$, $SE = 0.078$, $Z = -0.450$, $p = .653$.

⁴⁴² $\beta = -0.010$, $SE = 0.279$, $Z = -0.035$, $p = .972$; tendency evidence direction present: $Std Dev = 1.67$, absent: $Std Dev = 1.69$.

⁴⁴³ Non-penetrative offence: $\beta = 0.485$, $SE = 0.120$, $Z = 4.046$, $p < .001$; penetrative offence: $\beta = 0.474$, $SE = 0.129$, $Z = 3.688$, $p < .001$.

⁴⁴⁴ Juror level: Non-penetrative offence: $\beta = 0.012$, $SE = 0.009$, $Z = 1.267$, $p = .205$; penetrative offence: $\beta = 0.015$, $SE = 0.008$, $Z = 1.838$, $p = .066$. Jury level: Non-penetrative offence: $\beta = 0.036$, $SE = 0.054$, $Z = 0.676$, $p = .499$; penetrative offence: $\beta = 0.073$, $SE = 0.047$, $Z = 1.540$, $p = .124$.

penetrative offence: $M = 6.10$ and $M = 6.34$, each; penetrative offence: $M = 6.08$ and $M = 5.98$, each) ⁴⁴⁵

Factual culpability ratings for the penetrative offences were more dependent on jury groups ($ICC = .165$), that is, more variability in factual culpability was attributable to the influence of jury groups than was the case for factual culpability ratings for the non-penetrative offence ($ICC = .078$).

Factual culpability for the strong claims

Analyses revealed stronger variability due to jury groups for the single penetrative offence ($ICC = .189$) than for the two non-penetrative offences (masturbation of defendant: $ICC = .056$, masturbation of complainant: $ICC = .055$).

Mock jurors' *Child Sexual Abuse Knowledge* had a different impact on perception of factual culpability for the penetrative versus non-penetrative offences.⁴⁴⁶ While mock jurors' *Child Sexual Abuse Knowledge* had no impact in the non-penetrative offences, it was more likely to increase the finding of factual culpability for the penetrative offence. This effect remained significant at the jury level, such that juries who knew more about child sexual abuse were more likely to perceive the defendant factually culpable for the penetrative offence.⁴⁴⁷

Furthermore, the more mock jurors perceived Justin to be convincing, the more likely they were to rate the defendant factually culpable.⁴⁴⁸ Again, tendency evidence directions did not increase or decrease findings that the defendant was factually culpable in relation to any of the three counts (masturbation of defendant: $M = 6.15$ and $M = 6.36$; masturbation of complainant: $M = 6.10$ and $M = 6.42$; oral-penile penetration: $M = 6.08$ and $M = 5.99$, when tendency evidence directions were present and absent, respectively).⁴⁴⁹

⁴⁴⁵ Non-penetrative offence: $\beta = -0.200$, $SE = 0.225$, $Z = -0.888$, $p = .375$; tendency evidence direction present: $Std Dev = 1.45$, absent: $Std Dev = 1.40$. Penetrative offence: $\beta = 0.098$, $SE = 0.292$, $Z = 0.334$, $p = .738$; tendency evidence direction present: $Std Dev = 1.40$, absent: $Std Dev = 1.34$.

⁴⁴⁶ Masturbation of defendant: $\beta = 0.014$, $SE = 0.009$, $Z = 1.489$, $p = .136$; masturbation of complainant: $\beta = 0.014$, $SE = 0.009$, $Z = 1.519$, $p = .129$; oral-penile penetration: $\beta = 0.020$, $SE = 0.009$, $Z = 2.341$, $p = .019$.

⁴⁴⁷ Masturbation of defendant: $\beta = 0.043$, $SE = 0.046$, $Z = 0.931$, $p = .352$; masturbation of complainant: $\beta = 0.011$, $SE = 0.047$, $Z = 0.228$, $p = .819$; oral-penile penetration: $\beta = 0.090$, $SE = 0.040$, $Z = 2.221$, $p = .026$.

⁴⁴⁸ Masturbation of defendant: $\beta = 0.379$, $SE = 0.108$, $Z = 3.513$, $p < .001$; masturbation of complainant: $\beta = 0.393$, $SE = 0.115$, $Z = 3.423$, $p = .001$; oral-penile penetration: $\beta = 0.414$, $SE = 0.102$, $Z = 4.049$, $p < .001$.

⁴⁴⁹ Masturbation of defendant: $\beta = -0.183$, $SE = 0.195$, $Z = -0.936$, $p = .349$; tendency evidence direction present: $Std Dev = 1.29$, absent: $Std Dev = 0.88$. Masturbation of complainant: $\beta = -0.249$, $SE = 0.186$, $Z = -1.340$, $p = .1800$; tendency evidence direction present: $Std Dev = 1.41$, absent: $Std Dev = 0.78$. Oral-penile penetration: $\beta = 0.055$, $SE = 0.307$, $Z = 0.179$, $p = .858$; tendency evidence direction present: $Std Dev = 1.37$, absent: $Std Dev = 1.37$.

Verdict

Table 18 presents jury and mock juror verdicts by type of jury direction given. Chi-square analyses indicate that in the joint trial, the tendency evidence direction versus the standard direction had no significant influence on jury verdicts.⁴⁵⁰ Inspecting the verdicts showed that the conviction rate was higher for counts pertaining to the moderate and strong cases than for those pertaining to the weak case. For the counts pertaining to the moderately strong and the strong cases, the conviction rate was higher for non-penetrative than penetrative offences. The conviction rate for the counts pertaining to the weak case increased in the joint trial when a tendency evidence direction was given.

Table 18. Jury and mock juror verdicts for non-penetrative and penetrative offences in joint trials, with and without tendency evidence directions (per cent)

	Standard directions		Standard + tendency evidence directions			
	Not Guilty	Hung jury	Not Guilty	Hung jury		
(a) Jury verdict						
Weak: Masturbation of complainant	75.0	12.5	12.5	62.5	37.5	0.0
Moderate: Masturbation of complainant	0.0	100.0	0.0	25.0	75.0	0.0
Moderate: Digital-anal penetration	25.0	75.0	0.0	25.0	75.0	0.0
Strong: Masturbation of defendant	0.0	100.0	0.0	12.5	87.5	0.0
Strong: Masturbation of complainant	0.0	100.0	0.0	0.0	100.0	0.0
Strong: Oral-penile penetration	25.0	75.0	0.0	25.0	75.0	0.0
(b) Juror verdict						
Weak: Masturbation of complainant	67.9	32.1	–	46.7	53.3	–
Moderate: Masturbation of complainant	6.2	93.8	–	11.8	88.2	–
Moderate: Digital-anal penetration	16.0	84.0	–	12.0	88.0	–
Strong: Masturbation of defendant	1.2	98.8	–	11.8	88.2	–

⁴⁵⁰ $p > .05$.

Strong: Masturbation of complainant	1.2	98.8	–	3.3	96.7	–
Strong: Oral-penile penetration	17.3	82.7	–	12.9	87.1	–

Note. Tendency evidence directions absent: eight juries, $n = 83$ jurors; tendency evidence directions present: eight juries, $n = 93$ jurors.

Jury impressions of the judge's directions

The type of jury directions (standard directions, versus standard plus tendency evidence directions) had no impact on mock jurors' impressions as to whether the judge's instructions: helped them to understand the case (tendency evidence directions present: $M = 5.76$, absent: $M = 5.84$)⁴⁵¹, changed their minds about their verdicts (tendency evidence directions present: $M = 3.40$, absent: $M = 3.33$)⁴⁵², were confusing (tendency evidence directions present: $M = 2.96$, absent: $M = 2.59$)⁴⁵³ or were fair to the defendant (tendency evidence directions present: $M = 5.75$, absent: $M = 5.46$).⁴⁵⁴

Mock jurors rated the standard jury directions as easier to understand ($M = 5.78$) than the standard plus tendency evidence directions ($M = 5.45$)⁴⁵⁵, although there were no differences in reports from the mock jurors in either type of trial (with and without the tendency evidence direction) as to whether:

- (a) the testimony of the complainant with the weak case influenced them in relation to the counts in the moderate and the strong cases
- (b) the testimony of the complainant with the moderate case influenced them in relation to the counts in the weak and the strong cases
- (c) the testimony of the complainant with the strong case influenced them in relation to the counts in the weak and the moderate cases.⁴⁵⁶

The mean scores varied between $M = 4.11$ and $M = 4.85$ for the trial with tendency evidence directions, and between $M = 4.25$ and $M = 4.93$ for the trial without the tendency evidence

⁴⁵¹ $\beta = -0.082$, $SE = 0.205$, $Z = -0.401$, $p = .688$, $ICC = .016$; tendency evidence direction present: $Std Dev = 1.42$, absent: $Std Dev = 1.19$.

⁴⁵² $\beta = 0.069$, $SE = 0.328$, $Z = 0.212$, $p = .832$, $ICC = .016$; tendency evidence direction present: $Std Dev = 2.17$, absent: $Std Dev = 2.00$.

⁴⁵³ $\beta = 0.366$, $SE = 0.300$, $Z = 1.221$, $p = .222$, $ICC = .031$; tendency evidence direction present: $Std Dev = 1.90$, absent: $Std Dev = 1.72$.

⁴⁵⁴ $\beta = 0.291$, $SE = 0.265$, $Z = 1.100$, $p = .271$, $ICC = .028$; tendency evidence direction present: $Std Dev = 1.55$, absent: $Std Dev = 1.67$.

⁴⁵⁵ $\beta = -0.332$, $SE = 0.146$, $Z = -2.276$, $p = .023$, $ICC = .014$; tendency evidence direction present: $Std Dev = 1.60$, absent: $Std Dev = 1.22$.

⁴⁵⁶ $p > .05$.

directions. This indicates neither strong agreement nor disagreement by mock jurors as to whether the testimony of one complainant influenced their decision in relation to the counts relating to another complainant.

Self-reported cognitive effort when tendency evidence direction is present

Our analyses revealed no differences in the overall reported cognitive effort mock jurors expended when they had received standard judicial directions versus those that received standard plus tendency evidence directions.⁴⁵⁷ When considering the perceived cognitive effort measures separately, two significant differences between the trials emerged. When they were given the tendency evidence direction, mock jurors reported requiring more cognitive effort to understand the judge's instructions ($M = 5.34$)⁴⁵⁸ than when no tendency evidence direction was given ($M = 4.85$), but that they required less cognitive effort to understand the charges against the defendant when they were given the tendency evidence direction ($M = 4.80$) compared to when they only received standard directions ($M = 5.41$).⁴⁵⁹

Quantitative analysis of errors in applying the tendency evidence direction

To further examine the impact of the tendency evidence directions, we compared deliberations in juries that deliberated with the tendency direction in a separate and a joint trial (Trials 6 and 7) and juries that deliberated without the tendency direction in a separate and a joint trial (Trials 5 and 10).

Results revealed that juries with the tendency evidence direction were more likely to correctly state the law on tendency evidence in their deliberations compared to juries that only received the standard jury directions.⁴⁶⁰ In addition, juries that did not receive a tendency direction were more likely to discuss the meaning of 'beyond reasonable doubt' than juries that did receive the direction.⁴⁶¹ This may indicate more confusion or disagreement about how to weigh and use the evidence among the juries who did not receive guidance on its use. However, there was no difference between these groups of juries when it came to applying the law on tendency evidence. The juries that received tendency evidence directions did not apply the tendency

⁴⁵⁷ $\beta = -0.122$, $SE = 0.182$, $Z = -0.672$, $p = .502$, $ICC = .014$.

⁴⁵⁸ $\beta = 0.492$, $SE = 0.262$, $Z = 1.877$, $p = .060$, $ICC = .021$; tendency evidence direction present: $Std Dev = 1.56$, absent: $Std Dev = 1.55$.

⁴⁵⁹ $\beta = -0.615$, $SE = 0.250$, $Z = -2.454$, $p = .014$, $ICC = .032$; tendency evidence direction present: $Std Dev = 1.86$, absent: $Std Dev = 1.66$.

⁴⁶⁰ Pearson's $R = -.356$, $p = .04$

⁴⁶¹ Kendall's Tau-b = .490, $p = .04$

evidence any more correctly than those that received the standard direction.⁴⁶² Similarly, the juries that received the tendency evidence direction did not make any fewer mistakes in applying the tendency evidence than juries that did not receive this guidance.⁴⁶³

⁴⁶² $p > .05$

⁴⁶³ $p > .05$

4.6 The influence of question trails on jury reasoning and decision making

In this section, we examine the influence of a question trail on jury reasoning and decision making, to discern whether this benefited the juries in their deliberations. Overall, use of a question trail appears to increase the efficiency of jury decision making. The main finding is that question trails reduced the overall duration of deliberation in the relationship evidence trial, where deliberations persisted far longer in the absence of a question trail. Mock jurors who used a question trail reported requiring significantly less cognitive effort to reach a unanimous verdict than those who deliberated without this aid. The question trail had no influence on mock jurors' memory for the case facts. Separate analyses in relationship evidence trials showed that with the aid of a question trail, juries rated the defendant as significantly less factually culpable, and accordingly, the conviction rate for both penetrative and non-penetrative offences declined. Content analysis of deliberations in those trials revealed that with a question trail, juries devoted a significantly greater proportion of deliberation time to discussing the counts and the judge's instructions. With a question trail, the mock jurors perceived that they required less cognitive effort to evaluate the defence case, resulting in a verdict shift from hung juries to acquittals. By comparison, in a joint trial, the presence of a question trail had no significant influence on the defendant's factual culpability or conviction rates, regardless of the evidence strength or offence type. However, 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.

4.6.1 Research aim

In this project, juries received question trails in two types of trials: a separate trial with relationship evidence, and a joint trial. The analyses presented in this section aimed to determine whether the presence of a fact-based question trail assisted juries in their reasoning and decision making.

We conducted these analyses in three stages. First we compared all juries that received a question trail with comparable sets of juries that did not, irrespective of whether the question trail was provided in a separate or a joint trial. This allowed us to discern whether the structured decision-making aid benefited the juries in their deliberations in general. Second, we examined the use of a question trail in a relationship evidence trial, looking only at relationship evidence

trials with and without the question trail. Finally, we examined the use of a question trail in a joint trial, looking only at joint trials with and without the question trail. The dependent measures of interest are listed below with each set of analyses.

4.6.2 Overall effects of question trails

The analysis presented in this section aimed to determine whether the presence of a fact-based question trail assisted juries in their reasoning and decision making in general, across both separate and joint trials.

In this section, comparisons across four trial groups involved a total of 35 juries⁴⁶⁴ ($n = 403$). The measures of interest were the mock jurors' factual recall, their perception of the defendant's factual culpability, their overall perceived cognitive effort and the duration of deliberation.

Table 19 displays the mean scores in the two trial types for: the accuracy with which jurors remembered the case facts about the focal complainant; the perceived cognitive effort and deliberation time expended; and the perception of the defendant's factual culpability for the counts relating to the moderately strong complainant. The results are broken down by type of trial and the presence or absence of a question trail.

Table 19. Mean accuracy of remembered case facts, perceived cognitive effort, deliberation time, and perceived factual culpability, by trial type and presence of a question trail

	Relationship evidence trial		Joint trial	
	No QT	QT	No QT	QT
Accuracy of remembered case facts (score 1-6)	4.48	4.42	3.30	3.70
Cognitive effort expended (score 1-7)	5.40	5.12	5.24	5.12
Deliberation time (minutes)	71.3	45.8	36.5	40.1
Factual culpability				
Non-penetrative offence (score 1-7)	5.51	4.73	6.10	6.06
Penetrative offence (score 1-7)	5.47	4.59	6.08	6.00

⁴⁶⁴ Trials 3, 4, 7 and 8.

Memory for case facts

Two-level regression analysis revealed that the type of trial had a significant effect on mock jurors' memory for case facts.⁴⁶⁵ Mock jurors remembered more case facts correctly in the relationship evidence trial than in the joint trial, and were more likely to confuse the case facts in the joint trial. Table 19 illustrates the influence of trial type and the presence of a question trail on mock jurors' memory for case facts.

The presence of a question trail had no effect on memory for case facts in either trial⁴⁶⁶; jury groups were responsible for 10 per cent of the variance in accurate memory of the case facts ($ICC = .101$). However, trial type did have a significant effect on memory of the case facts, such that jurors in the joint trial were less likely to remember the case facts correctly compared to jurors in the separate trial.⁴⁶⁷

Self-reported cognitive effort when using question trails

A two-level regression analysis examined the presence of a question trail and the trial type as predictors of mock jurors' overall perceived cognitive effort. We followed this with a series of analyses on the cognitive effort expended on each jury task.

Neither the presence of a question trail nor the trial type had any effect on the overall reported cognitive effort involved.⁴⁶⁸ However, when separately considering perceived cognitive effort for each task, the results revealed that the only jury task involving statistical differences was the effort expended to reach a unanimous verdict.⁴⁶⁹ Mock jurors reported that significantly less mental effort was required to reach a unanimous verdict when they deliberated with the assistance of the question trail compared to when they deliberated without it. Thus, the use of a question trail helped juries reach a unanimous verdict. These results are presented in Figure 13a and 13b.

⁴⁶⁵ $\beta = -0.942$, $SE = 0.152$, $Z = -6.199$, $p < .001$.

⁴⁶⁶ $\beta = 0.162$, $SE = 0.153$, $Z = 1.058$, $p = .290$.

⁴⁶⁷ $\beta = -0.942$, $SE = 0.152$, $Z = -6.199$, $p < .001$.

⁴⁶⁸ QT versus no QT: $\beta = -0.016$, $SE = 0.137$, $Z = -0.120$, $p = .905$; trial type: $\beta = 0.124$, $SE = 0.138$, $Z = 0.898$, $p = .369$; $ICC = .031$.

⁴⁶⁹ QT versus no QT: $\beta = -0.384$, $SE = 0.151$, $Z = -2.541$, $p = .011$; trial type: $\beta = -0.122$, $SE = 0.154$, $Z = -0.796$, $p = .426$; $ICC = .025$.

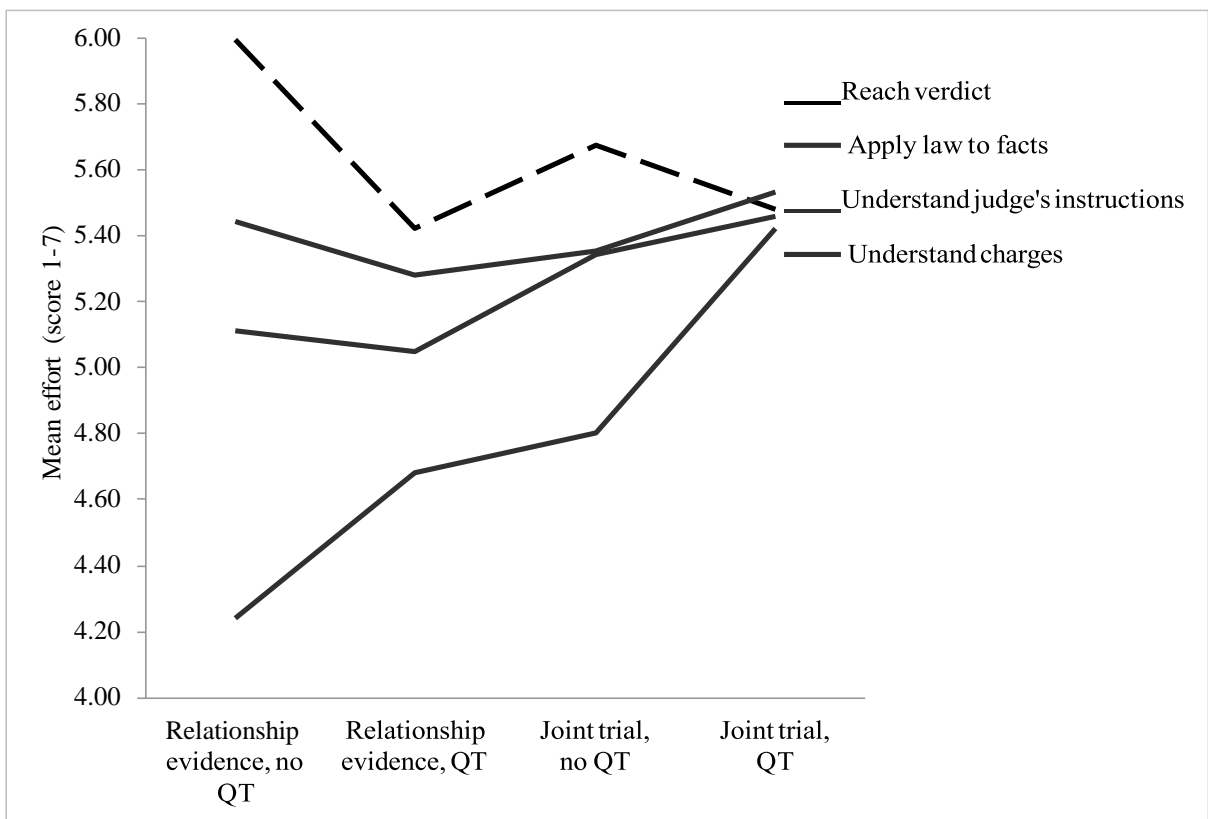
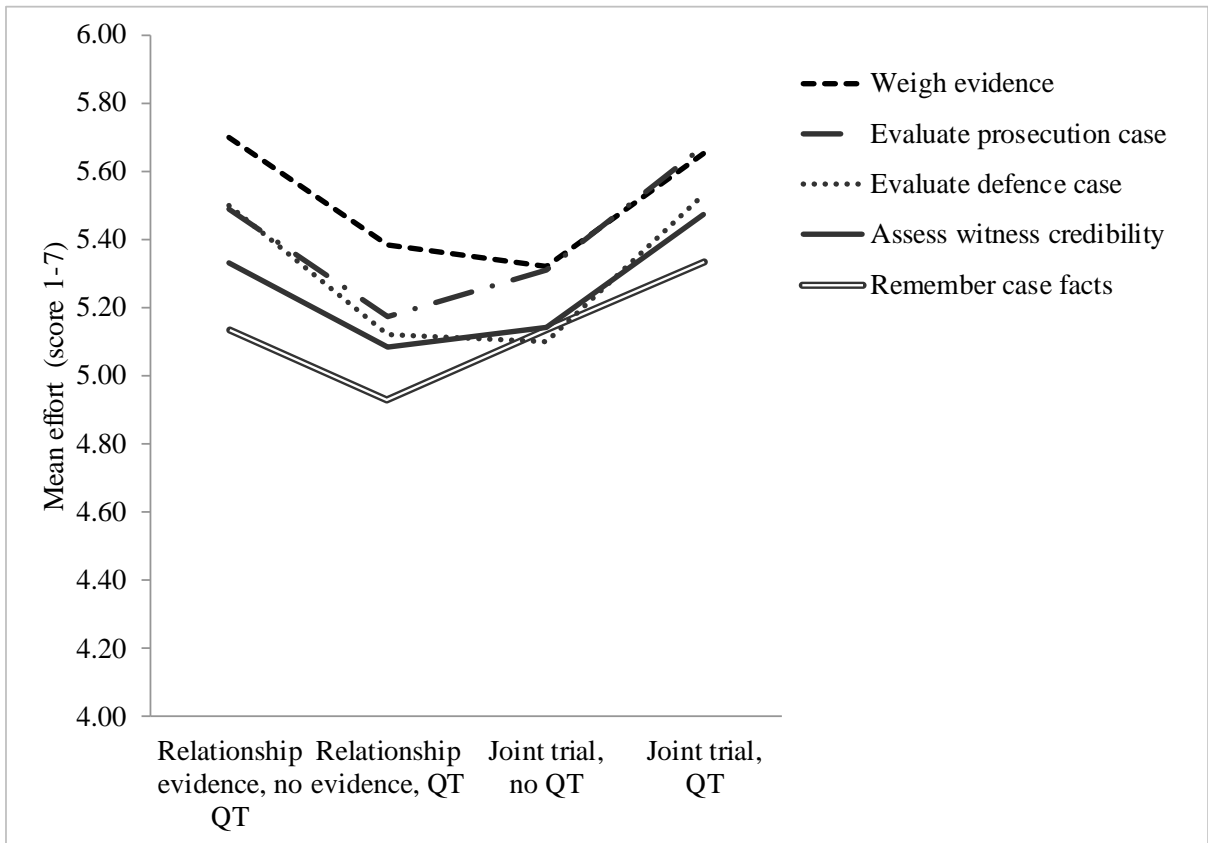


Figure 13a and 13b. Reported cognitive effort in relationship evidence trials and joint trials, with and without a question trail (QT)

Deliberation time and content

We conducted a two-way analysis to test differences in deliberation time between the trials; that is, the influence on deliberation time of the type of trial (separate trial with relationship evidence versus joint trial) and the presence of question trail (present versus absent). In our results, the type of trial had the greatest effect, in that juries deliberated for significantly longer in the relationship evidence trial than the joint trial, as shown in Table 19.⁴⁷⁰

Furthermore, there was a significant interaction between trial type and the presence of the question trail.⁴⁷¹ Whereas the presence of a question trail had no effect on deliberation time in the joint trial, in the relationship evidence trial, mock jurors deliberated for a significantly shorter period with the assistance of a question trail than without it.

Content analysis of the jury deliberations revealed that juries given a question trail in a trial with relationship evidence spent a greater proportion of their deliberation time discussing the counts⁴⁷² and the judge's directions on the law⁴⁷³ than juries that deliberated without the question trail. However, there was no significant difference between these groups in the frequency with which they mentioned the particular elements of the counts.⁴⁷⁴ No similar significant effects emerged in the joint trial.

In general, the presence of the question trail appeared to foster more efficient deliberation, irrespective of the final jury verdict.

Factual culpability of the defendant

We tested perceptions of factual culpability for the moderately strong case, as the two counts for this complainant were common to both the separate and the joint trials. A two-level regression analysis tested the impact of *Child Sexual Abuse Knowledge* (at the juror and jury levels), perceived convincingness (at the juror level), question trail presence (at the jury level) and trial type (at the jury level) on the perceived factual culpability of the defendant for each count.

Intra-class correlation revealed that about one-quarter of the variance in findings of factual culpability was attributable to the jury groups (non-penetrative offence: $ICC = .253$; penetrative

⁴⁷⁰ Trial type: $F(1, 35) = 11.26, p < .001, \eta_p^2 = .266$; Question trail: $F(1, 35) = 3.29, p = .079, \eta_p^2 = .096$.

⁴⁷¹ $F(1, 35) = 5.84, p = .022, \eta_p^2 = .158$.

⁴⁷² Kendall's Tau-b .5, $p = .011$.

⁴⁷³ Kendall's Tau-b .53, $p = 0.007$.

⁴⁷⁴ $p < .05$.

offence: $ICC = .270$). Mock jurors' *Child Sexual Abuse Knowledge* predicted their perception of the defendant's factual culpability, unrelated to the type of offence.⁴⁷⁵ The more mock jurors knew about child sexual abuse, the more likely they were to find the defendant factually culpable for the two counts. This effect disappeared at the jury level.⁴⁷⁶ Mock jurors' perceptions of the complainants' convincingness predicted findings of factual culpability⁴⁷⁷; jurors who perceived the complainant to be more convincing were more likely to perceive the defendant as factually culpable for the two counts. The presence of a question trail did not affect perceptions of the defendant's factual culpability⁴⁷⁸, whereas the trial type did – for both offences.⁴⁷⁹ As shown in Table 19, the defendant's factual culpability was rated significantly higher in the joint trial than the relationship evidence trial, unrelated to the seriousness of the offence. (Further analysis of the influence of a question trail on perceptions of the defendant's factual culpability within each type of trial is presented below.)

4.6.3 What is the influence of a question trail in a relationship evidence trial?

This section presents further analyses comparing the presence or absence of a question trail in a trial where relationship evidence is present. In the relationship evidence trial, the defendant was charged with two counts: a non-penetrative offence and a penetrative offence, both against Timothy. A total of 10 juries that viewed this trial received a question trail ($n = 107$ jurors) while nine juries did not ($n = 103$ jurors).

Factual culpability of the defendant

Two-level regression analysis revealed that mock jurors' *Child Sexual Abuse Knowledge*⁴⁸⁰ and perceptions of the complainant's convincingness⁴⁸¹ increased ratings of the defendant's

⁴⁷⁵ Non-penetrative offence: $\beta = 0.024$, $SE = 0.008$, $Z = 3.113$, $p = .002$; penetrative offence: $\beta = 0.023$, $SE = 0.008$, $Z = 2.967$, $p = .003$.

⁴⁷⁶ Non-penetrative offence: $\beta = -0.014$, $SE = 0.042$, $Z = -0.338$, $p = .736$; penetrative offence: $\beta = -0.001$, $SE = 0.045$, $Z = -0.027$, $p = .978$.

⁴⁷⁷ Non-penetrative offence: $\beta = 0.452$, $SE = 0.081$, $Z = 5.569$, $p < .001$; penetrative offence: $\beta = 0.427$, $SE = 0.082$, $Z = 5.236$, $p < .001$.

⁴⁷⁸ Non-penetrative offence: $\beta = -0.394$, $SE = 0.244$, $Z = -1.615$, $p = .106$; penetrative offence: $\beta = -0.469$, $SE = 0.248$, $Z = -1.893$, $p = .058$.

⁴⁷⁹ Non-penetrative offence: $\beta = 0.952$, $SE = 0.235$, $Z = 4.055$, $p < .001$; penetrative offence: $\beta = 1.005$, $SE = 0.239$, $Z = 4.203$, $p < .001$.

⁴⁸⁰ Non-penetrative offence: $\beta = 0.028$, $SE = 0.011$, $Z = 2.516$, $p = .012$; $\beta = 0.037$, $SE = 0.012$, $Z = 3.095$, $p = .002$.

⁴⁸¹ Non-penetrative offence: $\beta = 0.419$, $SE = 0.108$, $Z = 3.888$, $p < .001$; $\beta = 0.401$, $SE = 0.105$, $Z = 3.815$, $p < .001$.

factual culpability for both the non-penetrative and the penetrative offence. However, mean jury *Child Sexual Abuse Knowledge* had no effect on findings of factual culpability.⁴⁸²

The question trail did affect perceptions of the defendant's factual culpability; mock jurors who used a question trail perceived the defendant to be less factually culpable than those who did not, as shown in Table 19.⁴⁸³ The intra-class correlation for factual culpability was 20 per cent (non-penetrative offence: $ICC = .200$; penetrative offence: $ICC = .206$).

Self-reported cognitive effort with a question trail

The presence of the question trail did not impact mock jurors' overall perceived cognitive effort (overall mean scores), as shown in Table 19.⁴⁸⁴ Analyses of the effort required by specific jury tasks revealed significant differences due to the presence of the question trail. First, mock jurors who used a question trail reported that they exerted less cognitive effort to evaluate the defence case ($M = 5.17$) than those who did not use a question trail ($M = 5.49$).⁴⁸⁵ Second, mock jurors who used a question trail reported that they exerted less cognitive effort to reach a unanimous decision ($M = 5.42$) than those who did not use a question trail ($M = 5.99$).⁴⁸⁶ Overall, mock jurors who used a question trail reported that it decreased the mental effort required to perform some aspects of their decision making.

Verdict

Table 20 presents the outcomes of jury and mock juror verdicts reached with and without the aid of a question trail, in a separate trial with relationship evidence. The question trail helped juries reach consensus and a unanimous verdict. A comparison of the two groups showed a reduction in hung juries (shifting towards acquittal) with the aid of a question trail, from 44 per cent without the question trail to 20–30 per cent with the question trail. Thus, the conviction rate dropped in juries deliberating with the assistance of a question trail: one-third of juries found the defendant guilty on both counts in the absence of the question trail, compared to 10 per cent on the non-penetrative offence and zero per cent on the penetrative offence when deliberating with the question trail. Inspection of individual mock juror verdicts supported these findings and revealed greater disparities, as shown in Table 20.

⁴⁸² Non-penetrative offence: $\beta = -0.035$, $SE = 0.070$, $Z = -0.508$, $p = .612$; $\beta = -0.037$, $SE = 0.077$, $Z = -0.486$, $p = .627$.

⁴⁸³ Non-penetrative offence: $\beta = -0.747$, $SE = 0.348$, $Z = -2.150$, $p = .032$; $\beta = -0.865$, $SE = 0.352$, $Z = -2.460$, $p = .014$.

⁴⁸⁴ $\beta = -0.272$, $SE = 0.168$, $Z = -1.617$, $p = .106$, $ICC = .014$.

⁴⁸⁵ $\beta = -0.383$, $SE = 0.182$, $Z = -2.100$, $p = .036$, $ICC = .017$. No QT: $Std Dev = 1.61$; QT: $Std Dev = 1.56$.

⁴⁸⁶ $\beta = -0.569$, $SE = 0.201$, $Z = -2.836$, $p = .005$, $ICC = .041$. No QT: $Std Dev = 1.32$; QT: $Std Dev = 1.65$.

Table 20. Jury and mock juror verdicts in the relationship evidence trial, with and without a question trail (per cent)

	Count 1: Masturbation of complainant			Count 2: Digital-anal penetration		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
No question trail	22.2	33.3	44.4	22.2	33.3	44.4
Question trail	70.0	10.0	20.0	70.0	0.0	30.0
(b) Juror verdict						
No question trail	31.1	68.9	–	30.1	69.9	–
Question trail	76.2	23.8	–	80.8	19.2	–

Note. No question trail: nine juries, $n = 103$ jurors; with question trail: ten juries, $n = 107$ jurors.

Two-level logistic regression analysis revealed that mock jurors' *Child Sexual Abuse Knowledge* did not predict verdicts at either the juror⁴⁸⁷ or the jury⁴⁸⁸ level. The extent to which jurors perceived the complainant to be convincing increased the probability of a conviction.⁴⁸⁹ When mock jurors perceived the complainant to be more convincing, the odds of conviction were 2.0 times greater for the non-penetrative offence, and 2.5 times greater for the penetrative offence.

In the presence of the question trail, the likelihood of a conviction decreased for both counts.⁴⁹⁰

4.6.4 What is the influence of a question trail in a joint trial?⁴⁹¹

We examined the influence of a question trail on jury reasoning and decision making in a joint trial by comparing responses of juries that were given a question trail with those that were not.⁴⁹² A total of eight juries received a question trail to assist their deliberations, while eight juries did not. All 16 juries received the same evidence and judicial directions. The sole

⁴⁸⁷ Non-penetrative offence: $\beta = -0.010$, $SE = 0.038$, $Z = -0.263$, $p = .792$, Odds Ratio = 0.990, 95% CI [0.920; 1.066]; penetrative offence: $\beta = 0.024$, $SE = 0.032$, $Z = 0.745$, $p = .456$, Odds Ratio = 1.024, 95% CI [0.961; 1.092].

⁴⁸⁸ Non-penetrative offence: $\beta = -0.013$, $SE = 0.300$, $Z = -0.044$, $p = .965$; penetrative offence: $\beta = -0.043$, $SE = 0.341$, $Z = -0.128$, $p = .898$.

⁴⁸⁹ Non-penetrative offence: $\beta = 0.656$, $SE = 0.205$, $Z = 3.202$, $p = .001$, Odds Ratio = 1.927, 95% CI [1.290; 2.879]; penetrative offence: $\beta = 0.928$, $SE = 0.250$, $Z = 3.717$, $p < .001$, Odds Ratio = 2.529, 95% CI [1.551; 4.126].

⁴⁹⁰ Non-penetrative offence: $\beta = -4.033$, $SE = 1.492$, $Z = -2.704$, $p = .007$; penetrative offence: $\beta = -5.692$, $SE = 2.062$, $Z = -2.761$, $p = .006$.

⁴⁹¹ Trial 7 versus Trial 8.

⁴⁹² Trials 7 versus Trial 8.

differentiating factor was the presence or absence of a question trail, although juries provided with a question trail were not instructed that they were required to use it. The measures of interest were the factual culpability of the defendant, and conviction rates and self-reported cognitive effort.

In joint trials, the presence of the question trail had no significant influence on how mock jurors rated the defendant's factual culpability, nor did it influence their verdicts.⁴⁹³

In the joint trial, mock jurors' overall self-reported cognitive effort did not differ between juries who deliberated with and without the assistance of a question trail.⁴⁹⁴ Examination of the perceived cognitive effort of jury tasks separately revealed that the presence of a question trail did have an effect in relation to jurors' understanding of the charges against the defendant.⁴⁹⁵ Mock 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. There was no difference in mental effort for the other jury tasks in a joint trial.

⁴⁹³ $p > .05$.

⁴⁹⁴ $\beta = 0.264$, $SE = 0.201$, $Z = 1.315$, $p = .189$, $ICC = .056$.

⁴⁹⁵ $\beta = 0.628$, $SE = 0.230$, $Z = 2.728$, $p = .006$, $ICC = .038$.

4.7 Self-reported cognitive effort by type of trial

In this study the four types of trial varied in complexity. The cognitive load imposed on jurors varied by trial type; it was greater in the tendency evidence and joint trials than in the basic separate and relationship evidence trials. Following their deliberations, all mock jurors responded to a series of questions about the extent of effort they had expended in performing different aspects of their duties when coming to a verdict. To gain further insight into mock jury reasoning and decision making, in this section, we present mock jurors' own perceptions of their tasks, by trial type. Mock jurors reported requiring more effort to understand the charges as more inculpatory evidence was admitted. As might be expected, mock jurors also perceived that recalling the case facts was significantly more demanding in the tendency evidence trials than in the basic separate trials, and that understanding jury instructions was more difficult in the joint trial than in the basic separate trial. Surprisingly, mock jurors perceived that they required significantly more cognitive effort in the separate trial with relationship evidence and the tendency evidence trial than in a basic separate trial, while they perceived the joint and basic separate trials as of equivalent difficulty. The same pattern held for the tasks of assessing witness credibility, weighing the evidence, and evaluating the case for the prosecution and defence. Jurors rated the process of reaching a unanimous verdict as significantly more difficult in the relationship evidence trial than in the basic separate trial. Jurors rated the deliberation process in the relationship evidence and joint trials as more useful in understanding the case compared to the same process in basic separate and tendency evidence trials. Finally, mock jurors reported that the deliberation process significantly increased their confidence in the verdict reached.

4.7.1 Research aim

The complexity of the four types of trials varied. Accordingly, the different types of trials imposed a different cognitive load on the jurors. In particular, the cognitive load was greater in the tendency evidence and joint trials, which involved more witnesses and counts than the basic separate and relationship evidence trials. To gain more insight into the mock jurors' reasoning and decision making, we explored the influence of trial type on mock juror perceptions of their tasks in terms of (a) the cognitive effort required for each task they performed, (b) the helpfulness of the group deliberation and (c) their confidence in the ultimate verdict.

Our analyses involved 35 juries and 412 mock jurors drawn from all four types of trials. Specifically, nine juries viewed the separate trial ($n = 105$ jurors)⁴⁹⁶, nine juries viewed the relationship evidence trial ($n = 103$ jurors)⁴⁹⁷, nine juries viewed the tendency evidence trial ($n = 111$ jurors)⁴⁹⁸ and eight juries viewed the joint trial ($n = 93$ jurors).⁴⁹⁹

The measures described in this section were self-reports gathered from mock jurors after their deliberations concluded, in response to a series of questions about the extent of effort required to remember the case facts; assess the credibility of witnesses; weigh the evidence; evaluate the prosecution and defence cases; understand the charges against the defendant; understand the judge's directions; and reach a unanimous verdict. In addition, mock jurors were asked whether the deliberation with other jurors helped them understand the case facts, changed their views of the defendant's guilt or led to disagreements about the verdict. Finally, mock jurors reported their confidence in the verdicts reached.

We conducted two types of multi-level regression analysis. First, we compared the basic separate trial to the three other types of trials (separate trials with relationship evidence, separate trials with tendency evidence, and joint trials). Next, we compared the joint trial to the separate trials with and without relationship evidence, and to the separate trials with and without tendency evidence (referred to as the additional analyses).

4.7.2 Self-reported cognitive effort

Multi-level regression analyses revealed that trial type had a significant effect on the overall perceived cognitive effort when taking jury groups into consideration ($ICC = .037$).⁵⁰⁰ Compared to the basic separate trial ($M = 4.83$), the mean reported cognitive effort was significantly higher in relationship evidence trials ($M = 5.40$) and tendency evidence trials ($M = 5.31$). Notably, the extent of perceived cognitive effort expended did not differ significantly between the basic separate ($M = 4.83$) and the joint trials ($M = 5.24$).⁵⁰¹

⁴⁹⁶ Trial 1, basic separate trial with standard jury directions.

⁴⁹⁷ Trial 3, relationship evidence trial with standard and relationship evidence jury directions.

⁴⁹⁸ Trial 6, tendency evidence trial with standard and tendency evidence jury directions.

⁴⁹⁹ Trial 7, joint trial with six prosecution witnesses, and standard and tendency evidence jury directions.

⁵⁰⁰ Separate versus relationship evidence trial: $\beta = 0.562$, $SE = 0.202$, $Z = 2.778$, $p = .005$; separate versus tendency evidence trial: $\beta = 0.472$, $SE = 0.170$, $Z = 2.770$, $p = .006$; separate versus joint trial: $\beta = 0.406$, $SE = 0.219$, $Z = 1.885$, $p = .064$.

⁵⁰¹ Separate trial: $Std Dev = 1.32$; relationship evidence trial: $Std Dev = 1.23$; tendency evidence trial: $Std Dev = 1.32$; joint trial $Std Dev = 1.25$.

Separate trial = Joint trial

Separate trial < Relationship evidence trial = Tendency evidence trial

When we examined specific questions about mental effort expended, the pattern of differences between the experimental groups remained the same for the perceived effort required to assess credibility of the witnesses⁵⁰², weigh the evidence⁵⁰³, and evaluate the prosecution⁵⁰⁴ and defence⁵⁰⁵ cases.

Separate trial = Joint trial

Separate trial < Relationship evidence trial = Tendency evidence trial

Figure 14a shows the perceived cognitive effort required, in each type of trial, to remember the case facts, assess witness credibility, weigh the evidence, and evaluate the prosecution and the defence cases. Figure 14b shows the effort required to understand the charges against the defendant, understand the judge's instructions, apply the law to the facts and reach a verdict.

⁵⁰² Separate versus relationship evidence trial: $\beta = 0.530$, $SE = 0.185$, $Z = 2.865$, $p = .004$; separate versus tendency evidence trial: $\beta = 0.579$, $SE = 0.169$, $Z = 3.422$, $p = .001$; separate versus joint trial: $\beta = 0.342$, $SE = 0.184$, $Z = 1.855$, $p = .064$; $ICC = .024$.

⁵⁰³ Separate versus relationship evidence trial: $\beta = 0.524$, $SE = 0.216$, $Z = 2.426$, $p = .015$; separate versus tendency evidence trial: $\beta = 0.386$, $SE = 0.162$, $Z = 2.384$, $p = .017$; separate versus joint trial: $\beta = 0.151$, $SE = 0.247$, $Z = 0.614$, $p = .539$; $ICC = .030$.

⁵⁰⁴ Separate versus relationship evidence trial: $\beta = 0.523$, $SE = 0.230$, $Z = 2.277$, $p = .023$; separate versus tendency evidence trial: $\beta = 0.499$, $SE = 0.165$, $Z = 3.028$, $p = .002$; separate versus joint trial: $\beta = 0.352$, $SE = 0.234$, $Z = 1.503$, $p = .133$; $ICC = .022$.

⁵⁰⁵ Separate versus relationship evidence trial: $\beta = 0.655$, $SE = 0.214$, $Z = 3.068$, $p = .002$; separate versus tendency evidence trial: $\beta = 0.351$, $SE = 0.204$, $Z = 1.719$, $p = .086$; separate versus joint trial: $\beta = 0.251$, $SE = 0.254$, $Z = 0.989$, $p = .323$; $ICC = .034$.

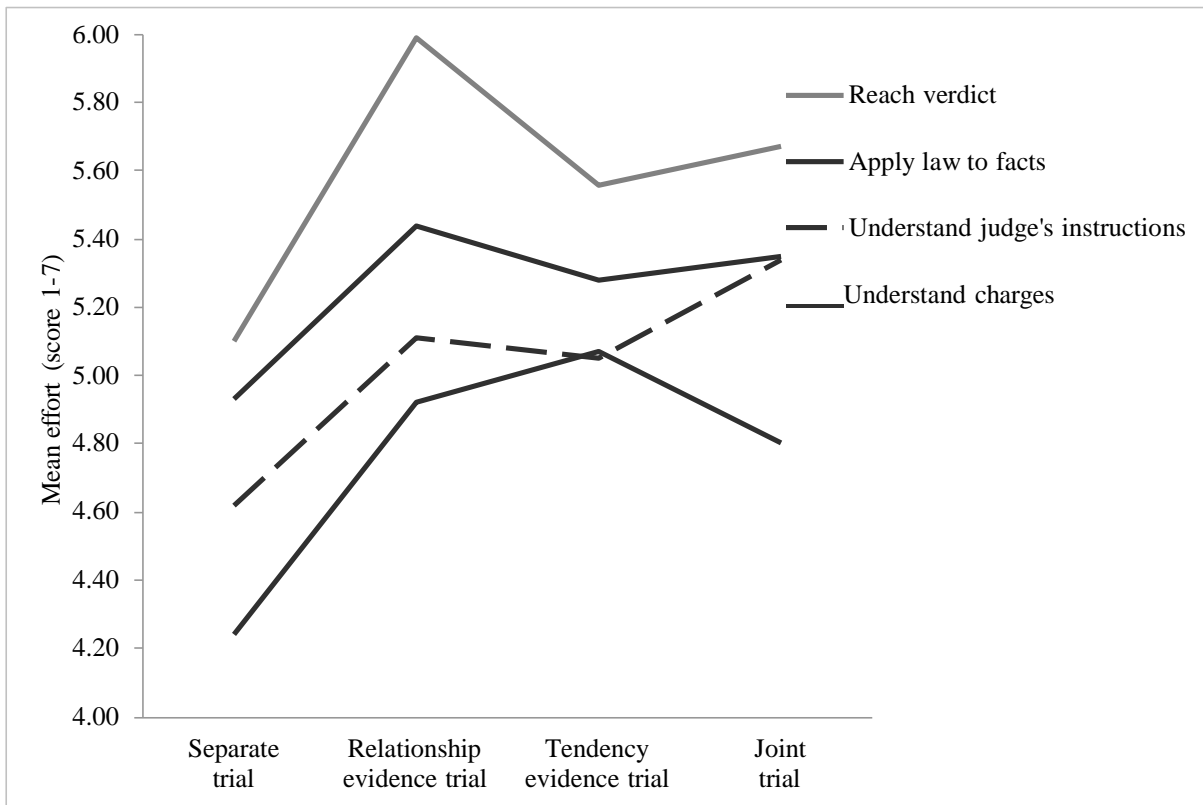
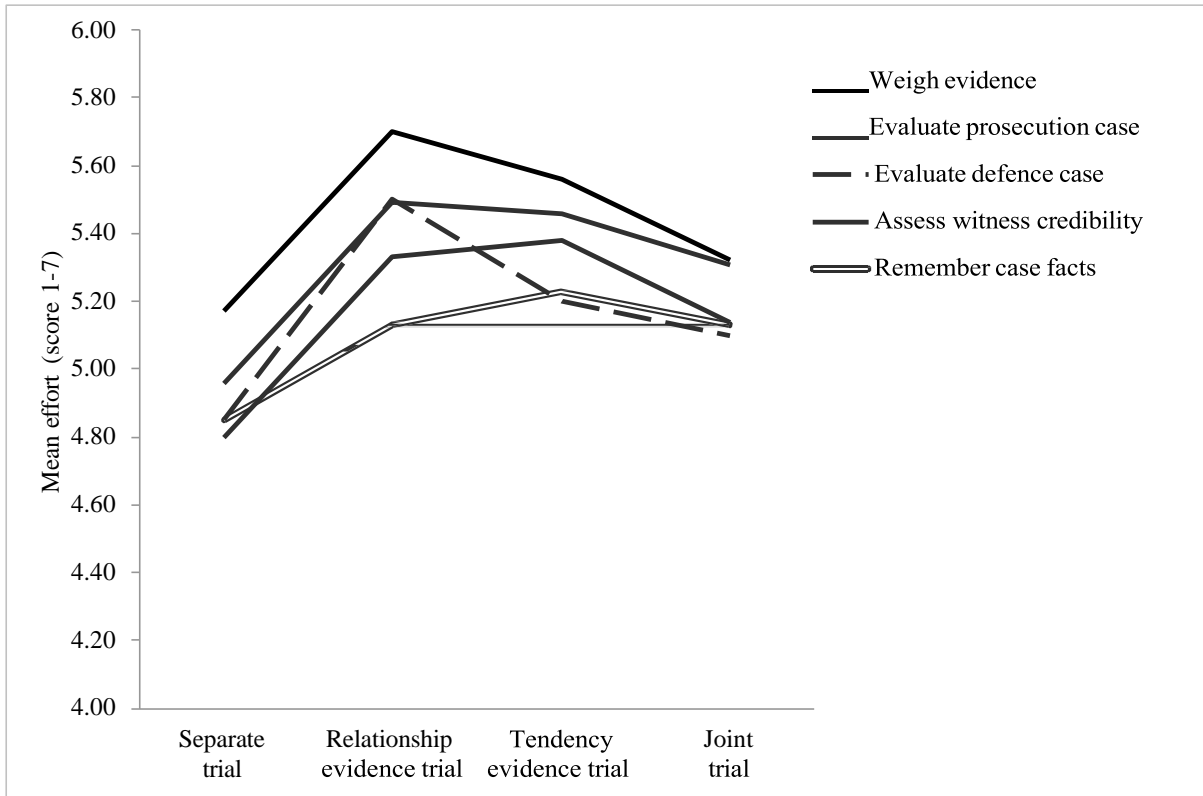


Figure 14a and 14b. Self-reported cognitive effort, by type of trial

Furthermore, compared to the most straightforward separate trial, mock jurors reported requiring significantly more cognitive effort to:

- (i) remember the case facts in the tendency evidence trial⁵⁰⁶ (separate trial < tendency evidence trial)
- (ii) understand the charges against the defendant in the relationship evidence, tendency evidence and joint trials⁵⁰⁷ (separate trial < relationship evidence trial = tendency evidence trial = joint trial)
- (iii) understand the judge's instructions in the joint trial⁵⁰⁸ (separate trial < joint trial)
- (iv) reach a unanimous verdict in the relationship evidence trial⁵⁰⁹ (separate trial < relationship evidence trial).

Additional analyses were conducted to assess mock jurors' self-reported cognitive effort by type of trial and verdict. Results of these analyses revealed that juries who convicted the defendant for at least one count reported that this decision required significantly more cognitive effort ($M = 5.33$) than that reported by juries who voted to acquit ($M = 4.90$).⁵¹⁰ This pattern held for every type of trial, and no differences emerged between the trial types.⁵¹¹ As is shown in Figure 15, the perceived cognitive effort by mock jurors on juries who convicted exceeded that reported by mock jurors on juries who acquitted for every type of trial.

⁵⁰⁶ Separate versus relationship evidence trial: $\beta = 0.278$, $SE = 0.240$, $Z = 1.158$, $p = .247$; separate versus tendency evidence trial: $\beta = 0.380$, $SE = 0.167$, $Z = 2.274$, $p = .023$; separate versus joint trial: $\beta = 0.283$, $SE = 0.203$, $Z = 1.392$, $p = .164$; $ICC = .011$.

⁵⁰⁷ Separate versus relationship evidence trial: $\beta = 0.683$, $SE = 0.278$, $Z = 2.453$, $p = .014$; separate versus tendency evidence trial: $\beta = 0.834$, $SE = 0.232$, $Z = 3.589$, $p < .001$; separate versus joint trial: $\beta = 0.558$, $SE = 0.282$, $Z = 1.977$, $p = .048$; $ICC = .027$.

⁵⁰⁸ Separate versus relationship evidence trial: $\beta = 0.488$, $SE = 0.312$, $Z = 1.563$, $p = .118$; separate versus tendency evidence trial: $\beta = 0.426$, $SE = 0.297$, $Z = 1.395$, $p = .163$; separate versus joint trial: $\beta = 0.570$, $SE = 0.321$, $Z = 1.775$, $p = .076$; $ICC = .034$.

⁵⁰⁹ Separate versus relationship evidence trial: $\beta = 0.889$, $SE = 0.313$, $Z = 2.838$, $p = .005$; separate versus tendency evidence trial: $\beta = 0.452$, $SE = 0.324$, $Z = 3.589$, $p < .001$; separate versus joint trial: $\beta = 0.558$, $SE = 0.282$, $Z = 1.977$, $p = .048$; $ICC = .088$.

⁵¹⁰ $\beta = 0.322$, $SE = 0.161$, $Z = 2.003$, $p = .045$. $ICC = .042$. Guilty: $Std Dev = 1.32$; Not guilty: $Std Dev = 1.28$.

⁵¹¹ Separate v. relationship evidence trial: $\beta = 0.209$, $SE = 0.273$, $Z = 0.766$, $p = .443$; Separate v. tendency evidence trial: $\beta = 0.302$, $SE = 0.206$, $Z = 1.467$, $p = .142$; Separate v. joint trial: $\beta = 0.213$, $SE = 0.234$, $Z = 0.909$, $p = .363$.

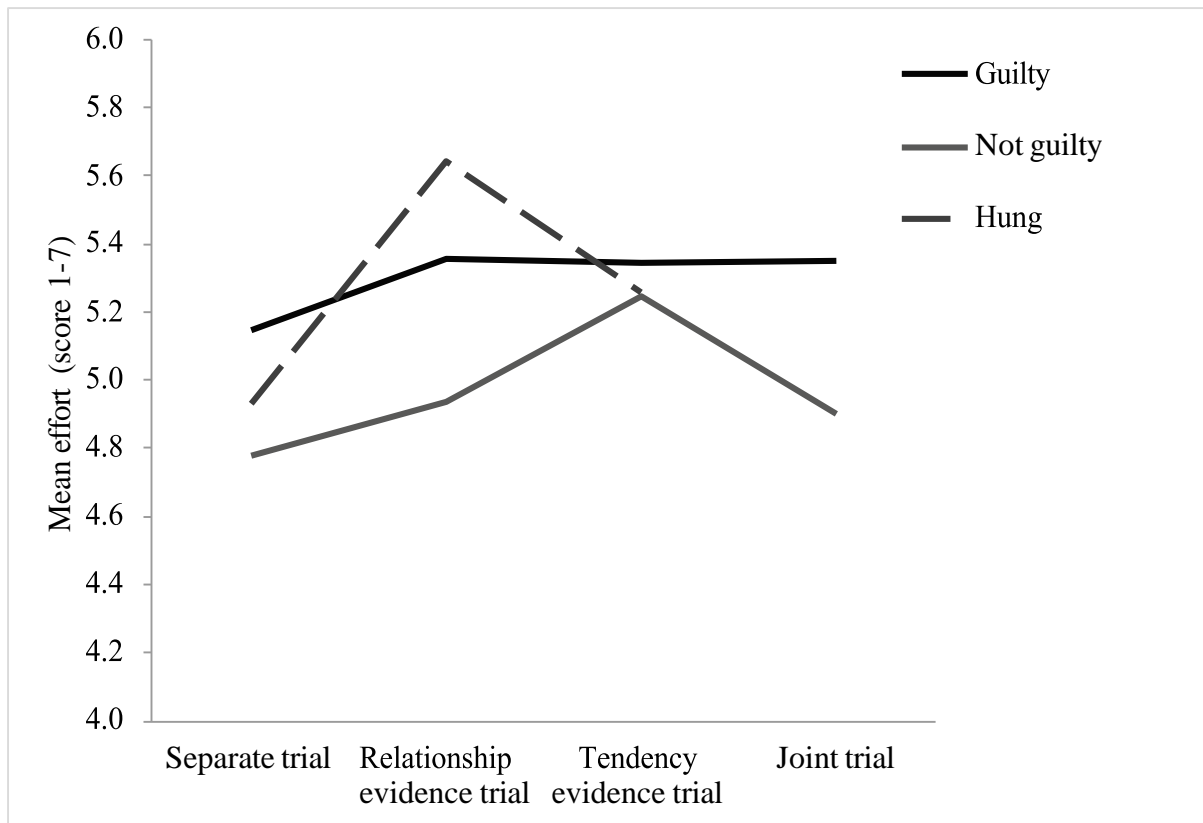


Figure 15. Self-reported cognitive effort by jury verdict and type of trial.

In addition, Figure 15 revealed that among hung juries, the self-reported cognitive effort was influenced by the type of trial, and was rated highest in the trial with relationship evidence.

Perceptions of the deliberation

After the deliberation, mock jurors were asked to what extent deliberation had helped them to understand the case, or led them to change their minds about the verdict, or led to disagreements with other mock jurors. Multi-variate regression analyses indicated that mock jurors were more likely to agree that the deliberation helped them understand the case in the relationship evidence trial ($M = 5.60$) and joint trial ($M = 5.56$) than in the basic separate trial ($M = 5.05$).⁵¹² There were no differences in how jurors perceived the deliberation process in the tendency evidence ($M = 5.09$) versus the basic separate trials ($M = 5.05$).⁵¹³ Additional analyses revealed that mock jurors felt deliberation helped them to understand the case more in the joint trials

⁵¹² Separate versus relationship evidence trial: $\beta = 0.546$, $SE = 0.197$, $Z = 2.771$, $p = .006$; separate versus tendency evidence trial: $\beta = 0.024$, $SE = 0.211$, $Z = 0.112$, $p = .911$; separate versus joint trial: $\beta = 0.499$, $SE = 0.230$, $Z = 2.173$, $p = .030$; $ICC = .053$.

⁵¹³ Separate trial: $Std Dev = 1.27$; relationship evidence trial: $Std Dev = 1.28$; tendency evidence trial: $Std Dev = 1.60$; joint trial: $Std Dev = 1.36$.

than the tendency evidence trials⁵¹⁴: separate = tendency evidence < relationship evidence = joint trial. Trial type had no effect on the extent to which deliberation changed mock jurors' verdicts⁵¹⁵ or led to disagreements with other mock jurors in deliberation.⁵¹⁶

Confidence in verdict

After deliberating to reach a verdict, mock jurors were asked how confident they were in their verdicts. Multi-level regression analyses revealed that almost one-quarter of the variance in confidence was attributable to jury groups ($ICC = .239$), indicating that deliberation had a strong effect on jurors' confidence in the decision.

When taking jury groups into account, mock jurors who deliberated in a relationship evidence trial ($M = 5.65$) and in a joint trial ($M = 5.87$) reported significantly more confidence in their verdicts than mock jurors in the basic separate trial ($M = 4.85$).⁵¹⁷ However, mock jurors in the tendency evidence trial ($M = 5.46$) were as confident in their verdicts as mock jurors in the basic separate trial ($M = 4.85$). Similarly, mock jurors in the joint trial were as confident in their verdicts as mock jurors in the relationship evidence and tendency evidence trials.⁵¹⁸

⁵¹⁴ Joint versus relationship evidence trial: $\beta = 0.048$, $SE = 0.196$, $Z = 0.242$, $p = .808$; joint versus tendency evidence trial: $\beta = -0.475$, $SE = 0.209$, $Z = -2.272$, $p = .023$.

⁵¹⁵ $p > .05$; $ICC = .047$.

⁵¹⁶ $p > .05$; $ICC = .115$.

⁵¹⁷ Separate versus relationship evidence trial: $\beta = 0.802$, $SE = 0.298$, $Z = 2.697$, $p = .007$; separate versus tendency evidence trial: $\beta = 0.609$, $SE = 0.442$, $Z = 1.376$, $p = .169$; separate versus joint trial: $\beta = 1.017$, $SE = 0.324$, $Z = 3.143$, $p = .002$. Separate trial: $Std Dev = 1.58$; relationship evidence trial: $Std Dev = 1.34$; tendency evidence trial: $Std Dev = 1.65$; joint trial: $Std Dev = 1.30$.

⁵¹⁸ $p > .05$.

4.8 Fairness of the trial

A primary concern when considering the use of separate versus joint trials in child sexual abuse cases is the fairness of the trial. In this section, we present a series of analyses that assessed juries' perceptions of the fairness of the trial, broken down by type of trial. These analyses are based on mock jurors' post-trial responses to a range of questions about the fairness of the trial; their expectations that they would be informed of any prior offending by the defendant; and the threshold they applied in interpreting the standard 'beyond reasonable doubt'. In addition, we drew on some coding of the content of the jury deliberations.

The main outcome of these analyses was a series of convergent findings that mock jurors rated the joint trial as more fair to the defendant than the basic separate trial. As we expected, mock jurors inferred more criminal intent on the part of the defendant as more inculpatory evidence was admitted in the different types of trials, and intent was rated as equivalent in the tendency evidence trials and joint trials. These ratings showed that mock jurors made a logical analysis of the inculpatory evidence presented in the different types of trials.

However, other results were unexpected. We found that mock jurors viewed the basic separate trial as significantly less fair to the defendant than trials that included more inculpatory evidence. In addition, mock jurors rated the defendant as significantly less convincing in the separate trial with relationship evidence than in the joint trial with tendency evidence. Similarly, the mock jurors perceived the judge's instructions as significantly less fair to the defendant in the basic separate trial than in the joint trial. A fourth unexpected difference – in the opposite direction from what many judges and practitioners anticipate – was mock jurors' own interpretations of the meaning of the threshold 'beyond reasonable doubt' as applied to convict the defendant. Mock jurors in the basic separate trial reported applying an average threshold that was less stringent and significantly lower (85.2 per cent) than the average threshold applied in the joint trial (92.1 per cent).

Finally, with respect to information about the defendant's prior offending, a substantial proportion (three-fifths) of the mock jurors expected that they would have been informed at trial of any prior incidents, charges or convictions for child sexual abuse involving the defendant. Significantly more mock jurors who attended a separate trial believed that if other charges had been made against the defendant, they would have been informed. In the course of jury deliberation, content analyses revealed that mock jurors rarely expressed concern about prior allegations against the defendant, and there was no significant relationship

between the trial type and mock jurors' comments stating their expectation that they would or would not have been informed of prior allegations of sexual misconduct.

4.8.1 Research aim

The results in this section were based on analyses conducted with the same 35 juries and 412 mock jurors as in Part 4.7, comparing mock juror perceptions in the basic separate trials with those in the relationship evidence trials, tendency evidence trials and joint trials.

Figure 16 displays the overall mean scores on these measures, by type of trial.

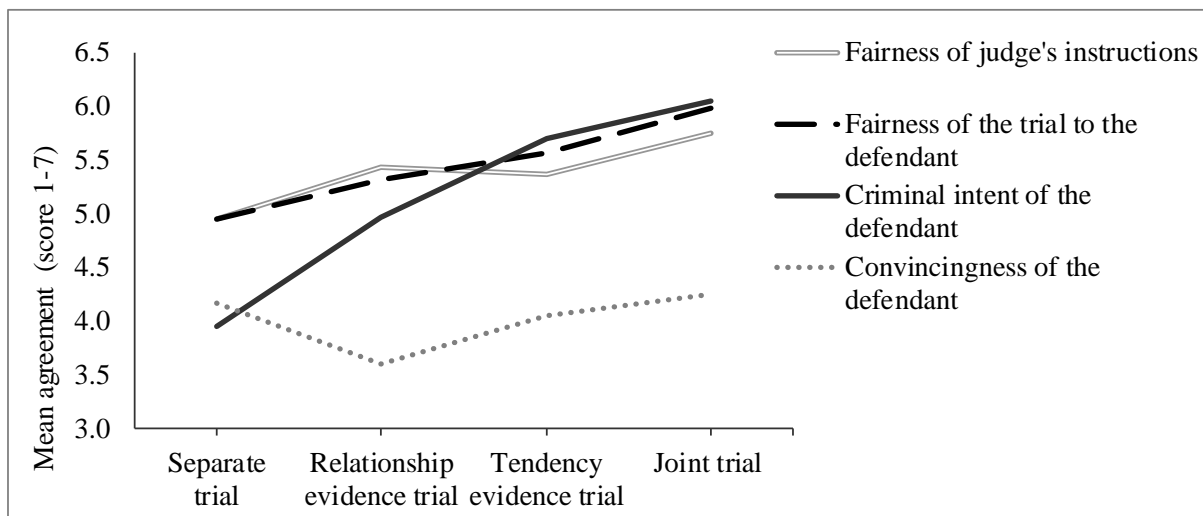


Figure 16. Mock jurors' perception of the convincingness of the defendant, criminal intent of the defendant, fairness of the judge's instructions and overall fairness of the trial, by type of trial.

4.8.2 Fairness to defendant by trial type⁵¹⁹

The convincingness of the defendant

Mock jurors reported the extent to which they perceived the defendant to be convincing, and we compared their responses across different types of trials. Multi-level analyses testing the impact of trial type on the perceived convincingness of the defendant revealed that the defendant was equally convincing in the basic separate trial ($M = 4.16$), the tendency evidence ($M = 4.04$) and joint trial ($M = 4.25$).⁵²⁰ The admission of inculpatory evidence from two

⁵¹⁹ Trials 1, 3, 6 and 7.

⁵²⁰ Separate versus tendency evidence trial: $\beta = -0.120$, $SE = 0.236$, $Z = -0.508$, $p = .612$; separate versus joint trial: $\beta = 0.094$, $SE = 0.217$, $Z = 0.432$, $p = .666$. Separate trial: $Std Dev = 1.47$; tendency evidence trial: $Std Dev = 1.82$; joint trial: $Std Dev = 1.80$.

additional witnesses, whether or not they were complainants, did not diminish mock jurors' ratings of how convincing they found the defendant to be.

Notably, mock jurors who heard the relationship evidence trial ($M = 3.60$) perceived the defendant to be significantly less convincing than their counterparts in the separate trial ($M = 4.16$).⁵²¹ Similarly, mock jurors in the relationship evidence trial perceived the defendant to be significantly less convincing than those in the joint trial.⁵²²

The intra-class correlation for convincingness of the defendant was .047, indicating that deliberations accounted for approximately 5 per cent of the variance in these perceptions.

Criminal intent of the defendant

Multi-level regression analysis tested the effect of the trial type on the extent to which mock jurors inferred the criminal intent of the defendant. Results revealed that as more inculpatory evidence was added to the trial, perceptions of the defendant's criminal intent increased.⁵²³

Compared to the basic separate trial ($M = 3.94$) mock jurors perceived significantly more criminal intent in the defendant's actions in the relationship evidence trial ($M = 4.97$), the tendency evidence trial ($M = 5.70$) and the joint trial ($M = 6.05$).⁵²⁴ The effect sizes increased as more inculpatory evidence and counts were presented to the juries. Additional analyses revealed that the perceived criminal intent of the defendant was higher in the joint trial than in the relationship evidence trial, but was undifferentiated in the joint and tendency evidence trials, where the evidence presented was the same.⁵²⁵ More than one-quarter of the variance in the perceived criminal intent of the defendant was attributable to the jury groups ($ICC = .281$), indicating that jury deliberations influenced these perceptions.

Perceptions of fairness

Overall, the jury groups were responsible for 7 per cent of the extent to which judicial directions were considered fair to the defendant. Analyses revealed no differences in the mean perceived

⁵²¹ $\beta = -0.557$, $SE = 0.282$, $Z = -1.975$, $p = .048$. Separate trial: *Std Dev* = 1.47; relationship evidence trial: *Std Dev* = 1.72.

⁵²² Joint versus relationship evidence trial: $\beta = -0.650$, $SE = 0.293$, $Z = -2.220$, $p = .026$; joint versus tendency evidence trial: $\beta = -0.214$, $SE = 0.249$, $Z = -0.858$, $p = .391$.

⁵²³ Separate versus relationship evidence trial: $\beta = 1.051$, $SE = 0.268$, $Z = 3.390$, $p < .001$; separate versus tendency evidence trial: $\beta = 1.755$, $SE = 0.274$, $Z = 6.415$, $p < .001$; separate versus joint trial: $\beta = 2.114$, $SE = 0.263$, $Z = 8.034$, $p < .001$.

⁵²⁴ Separate trial: *Std Dev* = 1.48; relationship evidence trial: *Std Dev* = 1.57; tendency evidence trial: *Std Dev* = 1.52; joint trial: *Std Dev* = 1.60.

⁵²⁵ Joint versus relationship evidence trial: $\beta = -1.063$, $SE = 0.314$, $Z = -3.390$, $p = .001$; joint versus tendency evidence trial: $\beta = -0.359$, $SE = 0.319$, $Z = -1.125$, $p = .261$.

fairness in the relationship evidence trial ($M = 5.43$) and tendency evidence trial ($M = 5.36$) compared to the basic separate trial ($M = 4.95$).⁵²⁶ However, mock jurors perceived the judge's directions to be significantly more fair to the defendant in the joint trial ($M = 5.75$) than in the basic separate trial ($M = 4.95$).⁵²⁷

Mock jurors were asked how fair the trial was to the defendant, and we compared their responses across different types of trials. After taking into account the jury groups ($ICC = .109$), analyses revealed that mock jurors who were assigned to the tendency evidence trial ($M = 5.57$) and joint trial ($M = 5.98$) perceived these trials as significantly more fair to the defendant than jurors who were assigned to the basic separate trial ($M = 4.95$).⁵²⁸ The perceived fairness of the trial to the defendant was equivalent in the relationship evidence trial ($M = 5.32$) and separate trial ($M = 4.95$). Additional analyses revealed greater perceived fairness to the defendant in the joint trial than in the relationship evidence trial.⁵²⁹

In sum, the effect size – that is, the difference between the basic separate trials and trials with additional evidence for the prosecution – increased as more inculpatory evidence for the prosecution was admitted. Mock jurors perceived the trials as fairer in relationship evidence trials, tendency evidence trials or joint trials with tendency evidence from multiple complainants than in the basic separate trial.

Because perceptions of fairness increased in trials with additional inculpatory evidence, it is possible that the mock jurors perceived the basic separate trial as less fair due to the perceived gaps in the evidence in that trial – a deliberate design feature of this study. The sole supporting evidence in the basic separate trial came from a former acquaintance of the defendant who confirmed the complainant's (Timothy's) description of a bedroom in which he alleged he was sexually abused by the defendant. Other than that, there was no independent evidence to support Timothy's account. In addition, there were inconsistencies in Timothy's evidence, such as the name of the movie he watched on the night he alleged he was abused, and the defendant

⁵²⁶ Separate versus relationship evidence trial: $\beta = 0.474$, $SE = 0.262$, $Z = 1.810$, $p = .070$; separate versus tendency evidence trial: $\beta = 0.405$, $SE = 0.229$, $Z = 1.770$, $p = .077$; separate versus joint trial: $\beta = 0.809$, $SE = 0.239$, $Z = 3.387$, $p = .001$.

⁵²⁷ Separate trial: $Std Dev = 1.55$; relationship evidence trial: $Std Dev = 1.61$; tendency evidence trial: $Std Dev = 1.55$; joint trial: $Std Dev = 1.55$.

⁵²⁸ Separate versus relationship evidence trial: $\beta = 0.367$, $SE = 0.250$, $Z = 1.469$, $p = .142$; separate versus tendency evidence trial: $\beta = 0.607$, $SE = 0.303$, $Z = 2.005$, $p = .045$; separate versus joint trial: $\beta = 1.026$, $SE = 0.220$, $Z = 4.656$, $p < .001$. Separate trial: $Std Dev = 1.78$; relationship evidence trial: $Std Dev = 1.59$; tendency evidence trial: $Std Dev = 1.44$; joint trial: $Std Dev = 1.29$.

⁵²⁹ Joint versus relationship evidence trial: $\beta = -0.660$, $SE = 0.225$, $Z = -2.929$, $p = .003$; joint versus tendency evidence trial: $\beta = -0.420$, $SE = 0.282$, $Z = -1.485$, $p = .137$.

denied parts of Timothy's evidence, such as Timothy's allegation that he had run away from the defendant's house after being abused.

Mock jurors may have queried why Timothy's allegations would be prosecuted in the absence of supporting evidence, such as evidence from Timothy's mother – who he said he had told about the abuse at the time, but who did not believe her son.

For example, in Jury 33⁵³⁰, the mock jurors had several questions about the conduct of the trial, including why Timothy's mother had not been called as a witness. Although one juror suggested the defence had neglected to call Mrs Lyons as a witness, another juror recognised that the prosecution had a responsibility to do so but said that "it wouldn't suit their interests". The following excerpt of their deliberations reflected these concerns:

JUROR 2: I have a lot of questions. Where was the mother and why wasn't she used as a witness?

JUROR 8: Mmm.

JUROR 2: Why did it take him, like -----

JUROR 13: I would say the strain on the relationship. That is a good question, though.

JUROR 2: Yeah.

JUROR 3: Why isn't the mother a witness?

JUROR 2: Yeah.

JUROR 6: The inability, they were explaining that, because of the delay and the facts that they didn't have time to prepare a defence that way; would explain why they couldn't call on his mother as a defence option.

JUROR 3: And the Crown, why didn't they call them?

JUROR 2: Yeah.

JUROR 6: Because it wouldn't suit their interests.

JUROR 9: But the Crown has to. They have got to -----

Later Juror 9 again noted the importance of the mother's evidence:

JUROR 9: Because the crucial piece of evidence, which would test – tie everything up together is the mother.

⁵³⁰ Trial 1, basic separate trial.

Juror 33 appears to have perceived that a case based on word-against-word evidence is inherently unfair to the defendant because it raises too much doubt:

JUROR 9: Based on the current evidence, which is basically his word against his word, Timothy's word against his word, there's not enough evidence to convict beyond reasonable doubt.

JUROR 8: Both have faults in their presentation.

JUROR 9: That's right, and therefore there is doubt.

In addition, some mock jurors were disturbed by the fact there was no other evidence about the defendant's history of offending so that the jury was unable to get 'a clear picture' of the context in which the alleged abuse occurred. This feature of the trial may have also contributed to perceptions of unfairness to the defendant:

JUROR 13: ... it was just one witness and then the other witness was just sort of like that, but there wasn't enough evidence on the barrister or the prosecution – even though the prosecution did have two witnesses. At the end, it's still – didn't give us a clear picture as to what was happening. So we need more. If maybe some – and also, from what I believe and what I have read about paedophiles and whatever, people that assault boys and stuff, it is normally a habit -----

JUROR 2: Mmm.

JUROR 13: -----and there is a history. So there isn't just a one-off incident. Something triggers it and it could come back time and time again. For him to be the only boy – uhm, he could have gotten other people that played soccer with him and asked, "Has this happened to you?" No other boys were interviewed. So I mean, a one-off – I don't know, paedophiles normally have a habit of repeating -----

We conducted further analyses to assess the extent to which the perceived fairness of the trial to the defendant was influenced by the trial outcome, after controlling for trial type.⁵³¹ For the purpose of these analyses, convictions were coded as a jury verdict of 'guilty' on at least one count versus jury verdicts of 'not guilty'.

Irrespective of type of trial, when juries convicted the defendant (for at least one count) the trial was rated as significantly more fair ($M = 5.94$) than when the juries voted to acquit ($M =$

⁵³¹ Multilevel analyses were conducted without hung juries, as there was no hung jury in the joint trial.

5.01).⁵³² Overall, mock jurors perceived the joint trial as significantly more fair to the defendant than the basic separate trial.⁵³³ Figure 17 displays the perceived fairness of the trial to the defendant by jury verdict and type of trial.

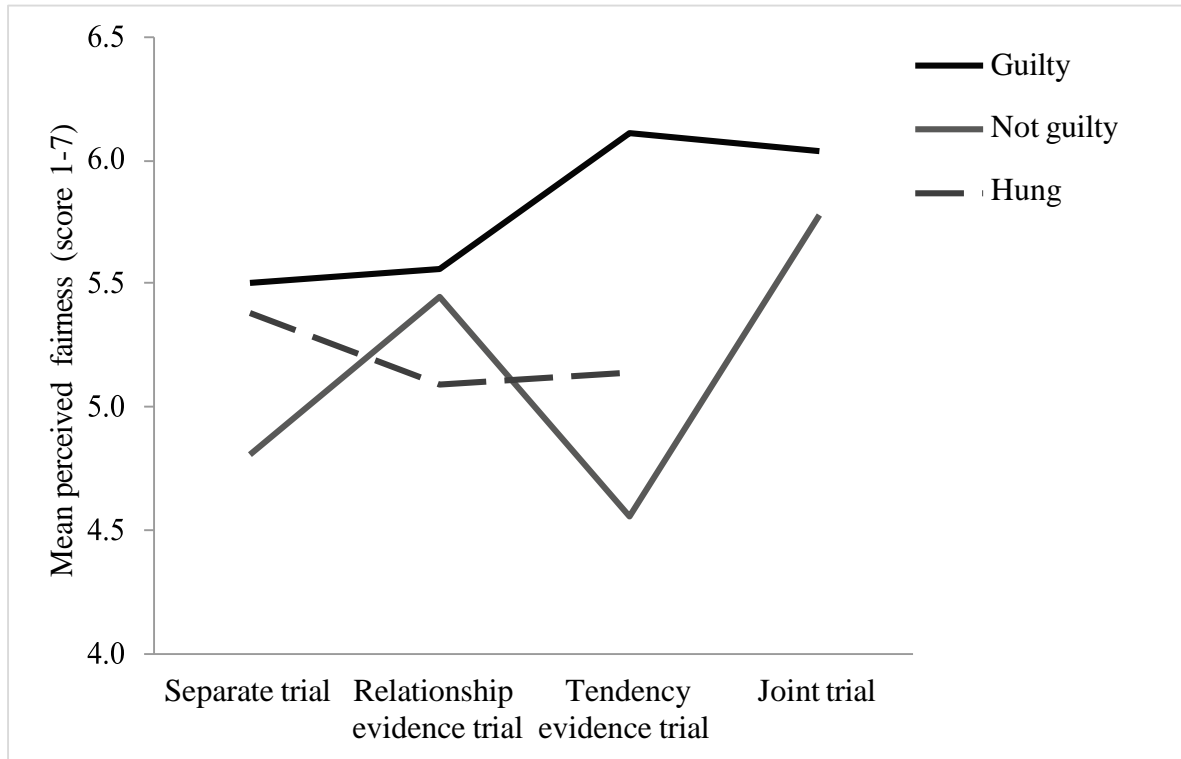


Figure 17. Perceived fairness of the trial to the defendant by type of trial

Expectations about information on prior offending

As shown in Table 21, analyses of mock jurors' post-trial expectations about the defendant's criminal history show that the majority expected they would be informed of prior charges against the defendant, of evidence of other sexual misconduct on other occasions, and of his prior convictions for child sexual abuse and other crimes. These comparisons were based on all 1,029 mock jurors and 90 juries.

⁵³² $\beta = 0.687$, $SE = 0.232$, $Z = 2.955$, $p = .003$. $ICC = .115$. Guilty: $Std Dev = 1.29$; Not guilty: $Std Dev = 1.76$.

⁵³³ Separate v. relationship evidence trial: $\beta = 0.287$, $SE = 0.258$, $Z = 1.110$, $p = .267$; Separate v. tendency evidence trial: $\beta = 0.365$, $SE = 0.331$, $Z = 1.103$, $p = .270$; Separate v. joint trial: $\beta = 0.647$, $SE = 0.263$, $Z = 2.464$, $p = .014$.

Table 21. Mock juror expectations of information they would receive at trial (per cent agreeing)

We would have been informed if:	Overall	Separate trial	Relationship evidence trial	Tendency evidence trial	Joint trial
1. ... other charges were made against the defendant	58.3	61.9	63.2	63.6	54.6
2. ... the defendant was sexually abusive on other occasions	60.2	59.0	63.4	58.7	58.5
3. ... the defendant had a prior conviction for child sexual abuse	59.7	62.9	62.7	56.6	57.8
4. ... the defendant had a prior conviction for any other crime.	45.9	48.6	45.8	41.3	47.6

Multi-level logistic regression analyses explored how mock jurors' education (at the juror level) and trial type (at the jury level) affected these responses, after taking the jury groups into account. In the first set of analyses, we used the separate trial as the baseline group; that is, we compared the separate versus relationship evidence trials, separate versus tendency evidence trials, and separate versus joint trials. Mock jurors' education had no impact on jurors' expectations of the information about the defendant they would receive at trial. This result held for all four questions listed in Table 21.⁵³⁴

The only question that reached statistical significance by type of trial was whether jurors would have been informed if other charges were made against the defendant. Mock jurors in the separate trial were more likely than those in the joint trial to agree that they would have been informed of other charges against the defendant.⁵³⁵ There were no differences between groups regarding expected information about prior sexual abuse by the defendant⁵³⁶, a prior child sexual abuse conviction⁵³⁷ or a conviction for any other crime.⁵³⁸

⁵³⁴ 1. $\beta = -0.049$, $SE = .126$, $Z = -0.338$, $p = .698$; 2. $\beta = 0.011$, $SE = .138$, $Z = 0.083$, $p = .934$; 3. $\beta = -0.040$, $SE = .135$, $Z = -0.297$, $p = .767$; 4. $\beta = -0.084$, $SE = .131$, $Z = -0.644$, $p = .520$.

⁵³⁵ Separate versus relationship evidence trial: $\beta = 0.095$, $SE = .167$, $Z = 0.568$, $p = .570$; separate versus tendency evidence trial: $\beta = -0.314$, $SE = .202$, $Z = -1.554$, $p = .120$; separate versus joint trial: $\beta = -0.306$, $SE = .153$, $Z = -1.995$, $p = .046$.

⁵³⁶ Separate versus relationship evidence trial: $\beta = 0.193$, $SE = .176$, $Z = 1.096$, $p = 0.273$; separate versus tendency evidence trial: $\beta = -0.020$, $SE = .210$, $Z = -0.094$, $p = .925$; separate versus joint trial: $\beta = -0.024$, $SE = .190$, $Z = -0.128$, $p = .898$.

⁵³⁷ Separate versus relationship evidence trial: $\beta = 0.005$, $SE = .162$, $Z = 0.030$, $p = .976$; separate versus tendency evidence trial: $\beta = -0.271$, $SE = .216$, $Z = -1.253$, $p = .210$; separate versus joint trial: $\beta = -0.201$, $SE = .182$, $Z = -1.105$, $p = .269$.

⁵³⁸ Separate versus relationship evidence trial: $\beta = -0.101$, $SE = .155$, $Z = -0.652$, $p = .515$; separate versus tendency evidence trial: $\beta = -0.328$, $SE = .221$, $Z = -1.484$, $p = .138$; separate versus joint trial: $\beta = -0.045$, $SE = .160$, $Z = -0.283$, $p = .777$.

In a second set of analyses, we used the joint trial as the baseline group and compared it against the other types of trials, after controlling for mock jurors' education. These analyses gave a more detailed response to the question as to whether jurors thought they would have been informed about other charges made against the defendant. Specifically, mock jurors exposed to the separate trial and also mock jurors exposed to the relationship evidence trial were more likely than those in the joint trial to expect that they would have been informed if other charges were made against the defendant.⁵³⁹ The difference between the tendency evidence trial and the joint trial was not significant. Furthermore, there were no differences between the groups for the remaining three questions in Table 21.⁵⁴⁰

Expectations about information of prior sexual abuse allegations

In light of the high rate of mock jurors expecting be informed of prior allegations of child sexual abuse by the defendant, we explored this topic further by assessing the content of jury deliberations. The purpose of this analysis was to determine whether erroneous expectations influenced juries' reasoning and decision making. All jury deliberations were coded for mentions of the topic.

Only one of the 90 juries – a jury that heard a tendency evidence case – raised the view that they would have been informed of any prior offences by the defendant. In Jury 82⁵⁴¹, Juror 8 commented that the jury would have been told about prior allegations against the defendant. The other jurors were dubious:

JUROR 8: Yes, I was going to add from my experience as a jury: if the accused have previous criminal records or anything like that, the Crown would definitely bring it up as to assist in the character of the person.

JUROR 13: I thought they weren't allowed to.

JUROR 12: When I was on a jury, they didn't ----

JUROR 13: Yeah.

JUROR 8: My one was. Like, they brought up issues and asked ----

JUROR 12: What was your case?

⁵³⁹ Joint versus separate trial: $\beta = 0.306$, $SE = .153$, $Z = 1.995$, $p = .046$; joint versus relationship evidence trial: $\beta = 0.401$, $SE = .157$, $Z = 2.548$, $p = .011$; Joint trial v. tendency evidence; $\beta = -0.008$, $SE = .195$, $Z = -0.043$, $p = .966$.

⁵⁴⁰ $p > .05$.

⁵⁴¹ Trial 5, separate trial with tendency evidence and standard jury directions.

JUROR 8: It was basically two men bashing each other up and, you know, they questioned the defence. Like, “in such and such year, were you involved in an incident?” They didn’t say “if you are guilty or innocent”, but, “Were you involved in an incident?” ----- things like that. So they will bring such things up. The fact that – that’s what I mean – compared to the Crown that I had before, I don’t think this Crown presented enough evidence or sufficiently convinced anyone that this evidence stands.

By comparison, five juries (5.6 per cent) mentioned in their deliberations at least once that they would *not* have been informed of any prior allegations against the defendant. For example, in Jury 51⁵⁴², a joint trial, jurors were of the opinion that they would *not* have been told about the accused’s previous criminal offences:

JUROR 5: Are we able – I’m just curious to know his past history – of any other criminal offences?

JUROR 12: No.

JUROR 5: We are not allowed to know that?

(Most jurors voiced agreement).

A statistical comparison of the frequency with which jurors expressed the belief, while deliberating, that they would or would not have been told about the defendant’s prior allegations yielded no significant differences between the different types of trials.⁵⁴³

⁵⁴² Trial 8, joint trial with tendency evidence jury directions and a question trail.

⁵⁴³ $p > .05$

4.8.3 Self-reported threshold for ‘beyond reasonable doubt’

After reaching a verdict, a post-trial questionnaire asked mock jurors what number between zero and 100 per cent represents ‘beyond reasonable doubt’. The overall average quantitative definition was 88.8 per cent.⁵⁴⁴ Multi-level regression analyses revealed that about 8 per cent of the variance depended on the jury groups.⁵⁴⁵ There were significant differences between trial types, showing that the threshold for ‘beyond reasonable doubt’ increased as more inculpatory evidence against the defendant was admitted at trial. Whereas the threshold was below 90 per cent in the basic separate and relationship evidence trials, the threshold exceeded 90 per cent in the joint trial.⁵⁴⁶ Compared to the separate trial (85.2 per cent), mock jurors’ definition of ‘beyond reasonable doubt’ was significantly more stringent when tendency evidence was admitted, whether in a separate trial (88.0 per cent) or a joint trial (92.1 per cent).⁵⁴⁷ Differences in the threshold in the basic separate trial (85.2 per cent) and relationship evidence trial (88.0 per cent) were not significant.

Additional multi-level analyses with the joint trial as the baseline group revealed a significant difference in the standard reported by mock jurors in the basic separate trial and the joint trial (85.2 per cent versus 92.1 per cent).⁵⁴⁸ The difference in the mean standard applied in the joint trial versus the relationship evidence trial was marginally significant, showing the tendency to apply a higher threshold in the trial with additional independent witnesses (92.1 per cent) than was applied in the trial with additional evidence from the complainant about relationship evidence (88.0 per cent).

These findings were unexpected, and contradicted concern among judges and practitioners that jurors would apply a lower threshold of proof in a joint trial than in a separate trial, due to the higher number of counts and witnesses in a joint trial.

⁵⁴⁴ *Std Dev* = 13.05.

⁵⁴⁵ *ICC* = .078.

⁵⁴⁶ Separate trial: *M* = 85.19, *Std Dev* = 15.24; relationship evidence trial: *M* = 87.99, *Std Dev* = 12.76; tendency evidence trial: *M* = 90.14, *Std Dev* = 12.10; joint trial: *M* = 92.11, *Std Dev* = 10.57.

⁵⁴⁷ Separate versus relationship evidence trial: $\beta = 2.746$, *SE* = 2.644, *Z* = 1.039, *p* = .299; separate versus tendency evidence trial: $\beta = 4.884$, *SE* = 2.026, *Z* = 2.410, *p* = .016; separate versus joint trial: $\beta = 6.765$, *SE* = 2.299, *Z* = 2.942, *p* = .003.

⁵⁴⁸ Joint versus separate trial: $\beta = -6.765$, *SE* = 2.299, *Z* = -2.942, *p* = .003; joint versus relationship evidence trial: $\beta = -4.019$, *SE* = 2.278, *Z* = -1.764, *p* = .078; joint versus tendency evidence trial: $\beta = -1.881$, *SE* = 1.501, *Z* = -1.253, *p* = .210.

Chapter 5: Discussion

Convergent measures from 90 mock juries and more than 1,000 jury-eligible citizens confirmed that they applied similar reasoning in joint and separate trials of institutional child sexual abuse, and that their verdicts were not based on impermissible reasoning. Overall, the risk of unfair prejudice to defendants in trials with tendency evidence was negligible. In fact, more instances of impermissible reasoning arose in basic separate trials and relationship evidence trials than in more complex trials that also included tendency evidence.

No joinder effect: Across all types of trials, the jurors' assessment of the credibility of the complainants predicted their perceptions of the culpability of the defendant, and the subsequent convictions. Conviction rates increased as more inculpatory tendency evidence was admitted, but these increases occurred in both the tendency evidence and joint trials. In the absence of significant differences between conviction rates in the tendency evidence trials and joint trials, there was no support for the hypothesised 'joinder effect'.

Convictions were not based on inter-case conflation of the evidence: The complexity of the trial evidence, not joinder, predicted the accuracy of mock jurors' factual recall. Intra-case errors were more common than inter-case errors. Because individual jurors' inaccuracies were promptly corrected within the jury, we found no evidence that persistent uncorrected errors had any causal effect on jury verdicts in trials involving tendency evidence.

Convictions were not based on accumulation prejudice: We examined whether juries were prone to convict based on the overall number of charges or overall number of witnesses for the prosecution. No juries in either the tendency evidence or joint trials reasoned that the defendant was guilty because of the number of allegations against him. 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. Similarly, there was no significant increase in conviction rates or ratings of the defendant's factual culpability when additional witnesses gave evidence for the prosecution, and the presence of additional witnesses did not increase the conviction rate for the count based on the weakest evidence. These findings controverted the accumulation prejudice hypothesis – that weak claims will not be distinguished due to inappropriate generalisation in the presence of multiple counts or witnesses in a joint trial. Rather, mock jurors' ratings of the defendant's factual culpability differed according to the evidence strength, and independently of the type of offence, showing jurors' capacity to differentiate the evidence of each complainant.

Convictions were not based on character prejudice: We examined whether jury decision making was based on reasoning that the defendant was guilty because he was a person of bad character. The defendant was rated equally convincing in the basic separate trial, the tendency evidence trial and the joint trial. The admission of inculpatory evidence from two additional prosecution witnesses did not reduce ratings of the defendant's convincingness, suggesting that mock jurors were not reasoning on the basis of character prejudice. While a few individual jurors' comments may have reflected character prejudice, these instances were isolated. No jury reached a verdict on the basis of emotional reactions to the severity of the allegations or out of a desire to punish the defendant.

The influence of jury directions on jury reasoning: The relationship evidence direction was of some assistance to juries and increased ratings of the factual culpability of the defendant. However, we found no difference in the ratings of perceived criminal intent or factual culpability of the defendant for each of the counts when standard directions versus a tendency direction were given, whether in the tendency evidence trial or the joint trial. Analyses of the content of jury deliberations showed that many juries either ignored or misunderstood the direction, and consequently failed to apply it or misapplied it.

The influence of question trails on jury reasoning: We identified three features associated with the use of a question trail in deliberations. First, deliberations in separate trials were less protracted. Second, juries spent a greater proportion of time discussing the specific counts and the judges' directions. Third, mock jurors who used a question trail reported requiring significantly less cognitive effort to reach a unanimous verdict. Together, these findings suggest that question trails facilitate jury consensus and efficiency.

This chapter is organised by reference to the joinder effect, to determine the extent to which a defendant's right to a fair trial may be compromised in joint trials. First, we discuss whether a joinder effect emerged in a joint trial of six counts involving three complainants, and, if so, whether this effect was due to any or all of the three types of impermissible reasoning identified in Chapter 2 – that is, inter-case conflation of the evidence, accumulation prejudice or character prejudice. Next, we discuss whether judicial directions and/or question trails reduce any impermissible reasoning by juries. Finally, drawing on the results of the multiple regression analyses and content analysis of jury deliberations, we make some general observations about juries' reasoning and decision making, and the probability of unfair prejudice to the defendant.

5.1 Was there a joinder effect?

Strictly speaking, the joinder effect describes elevated conviction rates for the focal counts in a joint trial compared to similar counts in separate trials. In trials that involve multiple complainants and a single defendant, the law assumes there will be some degree of unfair prejudice to the defendant because the jury has heard about the defendant's other criminal misconduct from the evidence of multiple complainants.

In order to examine the likelihood of a joinder effect (if any) in a joint trial, we compared conviction rates in separate versus joint trials for the two counts involving the focal complainant, Timothy, whose evidence was common to all trials. This entailed comparing verdicts for Counts 1 and 2 in the three types of separate trials (basic, relationship evidence and tendency evidence trials), with Counts 5 and 6 in the joint trial (together, the focal counts). We expected that the conviction rates for these focal counts would be highest in the joint trial, where cross-admissible tendency evidence from the three complainants supported the prosecution's case.

As expected, we found that conviction rates varied according to the strength of the inculpatory evidence presented by type of trial. Conviction rates increased as more inculpatory evidence was admitted for the prosecution in the form of tendency evidence. Since these increases in the conviction rate occurred in both the tendency evidence trial and the joint trial, these findings did not support the hypothesised joinder effect.

When we compared conviction rates in the tendency evidence trial and the joint trial, we found no significant differences. Although the conviction rates by juries and by individual jurors in the joint trial were higher on average than those in the tendency evidence trial, these increases were not statistically significant, and were not due to the type of trial; that is, they were not due to the joinder of counts in the joint trial. Thus, we identified no significant joinder effect.

The admission of relationship evidence did not increase convictions for either Count 1 or Count 2 compared to the conviction rates for those counts in the separate trial. The higher incidence of hung juries in the relationship evidence trial compared to the other three types of trials (33 per cent and 41.7 per cent, respectively, for Counts 1 and 2), appeared to be related to the source of that additional prosecutorial evidence, namely the complainant himself, rather than independent witnesses. By contrast, in the tendency evidence and joint trials, when independent witnesses supported the prosecution case, the hung jury rates were either zero or close to zero.

Aside from the absence of a significant increase in conviction rates for the counts involving Timothy in the joint trial compared to the tendency evidence trial, our examination of jury reasoning and decision making using numerous other measures did not show that verdicts were based on impermissible or unfairly prejudicial jury reasoning. This conclusion is based on quantitative and qualitative analyses of the content of the jury deliberations, and from systematic comparisons of trial groups using multiple post-trial measures. For example, our analysis of credibility ratings confirmed that juries were sensitive to the source of additional inculpatory prosecution evidence when assessing Timothy's credibility. These credibility ratings were higher when independent witnesses supported Timothy's evidence (in the tendency evidence trial) than when the additional inculpatory evidence came from Timothy, himself (in the relationship evidence trial). These findings suggest that increases in ratings of Timothy's credibility were attributable to systematic and permissible reasoning, based on the probative value of the admissible tendency evidence.

A further example of what appeared to be jury reasoning logically related to the evidence was the finding that in the absence of tendency evidence, juries were generally more cautious about convicting for the more serious penetrative offence. When tendency evidence was admitted, the conviction rates for both the non-penetrative and the penetrative offences were significantly higher than in the separate trials without that evidence. This suggests that juries appropriately and correctly used the tendency evidence to dispel their doubts about Timothy's allegations of a penetrative offence committed by the defendant. Nonetheless, deliberations in some joint trials revealed a persistent reluctance among juries to convict for the more serious penetrative offences. Our analyses of predictors of convictions for the penetrative offences demonstrated that mock jurors who were better informed about child sexual abuse and had more accurate *Child Sexual Abuse Knowledge* were more likely to convict than their counterparts who had misconceptions about child sexual abuse. These analyses further confirmed that conviction rates for these offences were not driven by the joinder of additional counts, but were attributable to other factors.

While our study yielded no evidence of a joinder effect *per se*, we did find significantly higher conviction rates for the focal counts in the tendency evidence and joint trials – that is, where additional tendency witnesses supported the prosecution's case. As discussed in Chapter 2, on its own a joinder effect in a joint trial does not demonstrate unfair prejudice to the defendant any more than an equivalent conviction rate does in a tendency evidence trial. An increase in conviction rates in trials with tendency evidence is uninformative about the reasoning that

motivated the increase, so it is not possible to conclude, based on conviction rates alone, that the admission of tendency evidence was unfairly prejudicial to the defendant.

Importantly, an increase in the conviction rates may be due to *permissible* reasoning that is logically related to the evidence, such as permissible tendency reasoning by juries. Multiple convergent findings showed that jury decision making in the tendency evidence trial was similar to that in the joint trial, indicating that the juries in both types of trials were not reasoning in a different or haphazard manner. Thus, the admission of the tendency evidence, whether in the context of a separate or a joint trial, did not lead to impermissible reasoning, despite the difficulty juries experienced in following the judge's directions about how to use the tendency evidence. For instance, the presence of multiple witnesses or complainants presenting inculpatory evidence in support of the prosecution's case enhanced Timothy's credibility and convincingness, and across all types of trials, the juries' credibility assessments of Timothy predicted the convictions on the two counts.

Notwithstanding the overall absence of a joinder effect *per se*, we examined the extent to which convictions in the joint trials were underpinned by juries' permissible reasoning (logically related to the evidence) or impermissible reasoning (factual conflation, accumulation prejudice, character prejudice).

5.2 Did juries convict on the basis of impermissible reasoning?

Inter-case conflation of the evidence

To test the judicial hypothesis that jurors confuse or conflate the evidence tendered in support of different counts in joint trials, we compared the accuracy of jurors' factual recall and their ratings of the defendant's factual culpability in four different experimental conditions: the basic separate trial, the relationship evidence trial, the tendency evidence trial and the joint trial.⁵⁴⁹

We assessed the accuracy of factual recall in two ways. We measured the accuracy of individual mock jurors using an objective post-trial multiple-choice questionnaire, and we gauged the accuracy of the jury as a group by objectively coding the transcribed jury deliberations to record all instances of factual errors. Analysis of the deliberation transcripts permitted an assessment of whether any factual inaccuracies formed the basis of a verdict, that is, whether mock jurors relied on conflation of the evidence when deciding the guilt or innocence of the defendant.

⁵⁴⁹ Trials 1, 3, 6 and 7.

Post-trial recall accuracy was greatest and undifferentiated for the separate and relationship evidence trials in which less inculpatory evidence had been admitted. When mock jurors in the tendency evidence and joint trials were exposed to two additional witnesses/complainants who experienced events similar to those experienced by the focal complainant (Timothy), they were significantly more likely to confuse the case facts than were mock jurors who viewed the separate trial without such additional evidence. The mock jurors' mean recall accuracy scores decreased as the complexity of the trial increased. In other words, the complexity of the trial evidence rather than joinder, *per se*, significantly predicted factual recall accuracy. This complexity derived from the substantially longer duration of trials containing tendency evidence (2,427 words and 45 minutes for the shortest trial without tendency evidence, versus 3,875 words and 110 minutes for the longest trial with tendency evidence), as well as the increased complexity as a result of including a tendency evidence direction.

These individual post-trial scores did not take into account the facts discussed in deliberations or the extent to which jury deliberation successfully corrected individual jurors' factual inaccuracies. The deliberations revealed that more jurors in trials with tendency evidence made factual errors than jurors in trials without this evidence, driven by the higher number of witnesses in those trials. However, these inaccuracies were promptly corrected. We found no support for the hypothesis that persistent uncorrected errors were a feature of jury decision making in trials with tendency evidence, and found no evidence that errors of this nature had any causal effect on jury verdicts. While factual inaccuracies by individual mock jurors predicted their individual conviction rates in all types of trials, the analyses of jury decisions provided a different picture.

The video-trial simulations presented an abbreviated trial experience, and mock jurors had relatively little time to absorb the evidence and discuss it before rendering a verdict. Compared to a real trial, where there is considerable repetition throughout the trial and more opportunity for juries to discuss the evidence and deliberate, our experimental simulations may have fostered the potential for more confusion than occurs in an actual trial. The mock juries that were required to render a verdict on six counts had an equivalent amount of deliberation time as those faced with two counts. This artefact of the trial simulation is likely to have contributed to the higher factual error rate observed in cases with more complex evidence. Nonetheless, in real child sexual abuse trials, whether separate or joint, a similar potential for confusion cannot be discounted, as a result of, for example, trial length, juror fatigue, juror disinterest, changing levels of concentration and trial complexity.

The addition of two other witnesses/complainants in the tendency evidence and joint trials was very similar to that pertaining to *both* the non-penetrative and penetrative offences alleged by Timothy, unlike the additional evidence that was presented by Timothy in the relationship evidence trial. This additional tendency evidence was probative of the defendant's culpability and significantly increased jurors' factual culpability ratings of the defendant, compared to their ratings in the separate trial, independent of offence type. These findings indicate that jurors made inferences logically related to the offences, particularly as the prosecution described the additional evidence as tendency evidence in the video trial, along with prosecutorial instructions about how the evidence could be used. In summary, the observed pattern of factual culpability ratings showed that juries relied more on systematic reasoning rather than evidentiary conflation. This evidence of jury reasoning in response to additional evidence of the defendant's other criminal misconduct controverted the hypothesis that juries in joint trials or in trials with complex tendency evidence engage in impermissible prejudicial reasoning because of inter-case conflation of the evidence.

Past research on jury decision making has shown that juries are quite adept at recalling trial evidence.⁵⁵⁰ Our study replicated this finding. Firstly, over one-third of the deliberations were completely free of factual errors. Where factual errors were observed, more were intra-case than inter-case errors, and the overall rate was low. The maximum number of errors in deliberations across all 90 juries was four, but the average was 0.98. Second, when a juror made a factual error, fellow jurors corrected it. The capacity of the jury to marshal the facts in evidence and to correct individual juror errors prevented the incidence of inter-case evidential conflation, and no verdict to convict was premised on an inter-case factual error.

In sum, any observed decrease in factual accuracy was the result of an increased cognitive load on the mock jurors' working memory due to the greater complexity of trials with tendency evidence, and was not due to joinder *per se*. Furthermore, this effect had no detrimental influence on jury reasoning and decision making.

Accumulation prejudice

To examine the extent to which an increase in conviction rates in tendency evidence trials could be due to impermissible reasoning, we examined whether juries were prone to convict based on the overall number of charges against the defendant and the overall number of witnesses

⁵⁵⁰ Vicky L Smith 'Impact of Pretrial Instruction on Jurors' Information Processing and Decision Making' (1991) 76 *Journal of Applied Psychology* 220.

called by the prosecution. In addition, we examined whether juries were capable of distinguishing between the evidence tendered in support of each count when reaching their verdicts in joint trials.

We further scrutinised the accumulation prejudice hypothesis by examining the dependent measures for the complainant whose evidence was common to four different trial conditions: the separate, relationship evidence, tendency evidence and joint trials.⁵⁵¹ These measures included the credibility and convincingness of Timothy as a witness. We expected ratings for these dependent measures to be lower for Timothy in the separate and relationship evidence trials compared to the tendency evidence and joint trials, since we anticipated that Timothy's credibility and convincingness as a witness would be enhanced when other witnesses/complainants gave similar accounts of sexual abuse by the defendant.

As expected, jurors' ratings for the perceived credibility of Timothy were equivalent in the separate and relationship evidence trials, and significantly higher when tendency evidence was admitted in either a tendency evidence trial or a joint trial. In other words, the evidence of independent witnesses/complainants who reported similar criminal conduct by the defendant enhanced the perception of Timothy's credibility, irrespective of whether the defendant was charged with counts pertaining to those individuals. By contrast, the addition of relationship evidence had no impact on Timothy's credibility scores, as that information was from the same source – that is, from the complainant himself. Similarly, ratings of Timothy's convincingness were significantly higher when tendency evidence was admitted, compared to the separate and relationship evidence trials where there were no such supporting witnesses.

Importantly, juries in tendency evidence trials that received specific directions on using the tendency evidence⁵⁵² were advised that they were permitted to use this evidence against the defendant if they were satisfied beyond reasonable doubt that the defendant had a sexual interest in Justin and/or Simon. If so satisfied, the juries were permitted to use that sexual interest when coming to a verdict in relation to the counts involving Timothy. This direction may have further contributed to the enhanced credibility and convincingness ratings for Timothy in trials that involved tendency evidence. In other words, permissible reasoning – not impermissible reasoning based on an accumulation of witnesses – accounted for the observed increases.

⁵⁵¹ Trials 1, 3, 6 and 7.

⁵⁵² Trials 6, 7, 8 and 9.

Multiple counts

Courts have hypothesised that a defendant will be prejudiced in joint trials because juries are prone to reasoning that the defendant is guilty simply because of the number of charges brought by the prosecution (as discussed in Chapter 2). In other words, juries are presumed to infer guilt from the impression created by the high numbers of counts, rather than by engaging in a systematic review of the evidence for each count in a joint trial.

To test whether juries were affected by the number of counts in the joint trial, we compared trials in which the prosecution presented the same evidence but the number of charges against the defendant varied. In the tendency evidence trial, two counts of sexual assault arose from the moderately strong evidence from Timothy, while two witnesses (Justin and Simon) gave evidence of their experiences of sexual abuse by the defendant. In the joint trial, the defendant was charged with six counts, with three complainants (Simon, Justin and Timothy) giving evidence of their experiences of sexual abuse. In this section we review the findings from measures other than verdict, to gain more insight into the nature of jury reasoning in these trials.

Although no verdicts were required regarding the sexual abuse alleged by Justin and Simon in the tendency evidence trial, we asked jurors to rate the factual culpability of the defendant for that alleged conduct, and compared the ratings of factual culpability in the tendency evidence and joint trials for all six alleged offences. When these offences were formal counts, factual culpability ratings for the defendant in relation to Justin's allegations were higher in the joint trial than in the tendency evidence trial, yielding a significant effect for the type of trial based on jurors' exposure to a higher total number of counts. Mock jurors were not more likely to perceive the defendant as being factually culpable of sexually assaulting Simon, whether his evidence was presented in a trial with two or six counts. This finding directly controverted the accumulation hypothesis that jurors will not distinguish weak claims due to inappropriate generalisation in the presence of multiple counts in a joint trial. In the joint trials of this study, juries devoted most deliberation time to evaluating the evidence of Simon, the complainant with the weakest claim.

Justin's evidence in the joint trial was also supported by two additional witnesses who did not appear in the tendency evidence trial. While it is possible that the higher factual culpability rating for the defendant in this trial was attributable to jurors' accumulation prejudice across multiple counts and witnesses, an alternative explanation is that jurors were influenced by

differences in the extent to which they were required to evaluate the evidence in each of these experimental conditions. One previous study demonstrated that mock juror verdicts and defendant character evaluations were affected to a greater extent by the number of charges the jurors actually evaluated, rather than the number of charges they were aware of.⁵⁵³

In a further contradiction to the accumulation of counts hypothesis, a comparison of jury deliberations in the tendency evidence trials with two counts versus the joint trial with six counts revealed no significant differences in the rates of impermissible reasoning in deliberations between the groups.

According to the accumulation prejudice hypothesis, in a joint trial one would expect factual culpability ratings for the defendant would not differ according to allegations of varying evidential strength; that is, that juries would rate the count based on the weakest evidence and the count based on stronger evidence similarly. As such, we compared factual culpability ratings for the six different counts in the joint trial.

These expectations were not confirmed. Factual culpability ratings differed according to evidence strength, independent of the type of offence. The lowest factual culpability rating was for Count 1 (weak case evidence) and the highest was for Count 3 (strong case evidence), although the ratings for the counts arising from the moderately strong and the strong case evidence were similar. These results did not confirm the accumulation prejudice hypothesis. Instead, they demonstrated that juries are able to evaluate the factual culpability of the defendant for the counts appropriately, according to their different evidential strength, and therefore are capable of distinguishing between the evidence of different complainants.

Multiple witnesses

Courts have hypothesised that juries are susceptible to the cumulative effects of multiple witnesses, which are expected to increase in joint trials. This formulation of the accumulation prejudice hypothesis holds that a defendant will be unfairly prejudiced because juries are prone to reasoning that the defendant is guilty simply because of the number of witnesses appearing for the prosecution. In other words, juries will infer guilt from the impression created by the

⁵⁵³ Irwin A Horowitz and Kenneth S Bordens, 'Prejudicial Joinder of Multiple Offenses: Relative Effects of Cognitive Processing and Criminal Schema' (1986) 7(4) *Basic and Applied Social Psychology*, 243.

number of witnesses, rather than engaging in a systematic review of the evidence for each count in the joint trial.

To test jury susceptibility to the cumulative effect of multiple witnesses, we examined conviction rates when either four or six witnesses appeared for the prosecution in a joint trial.⁵⁵⁴ The two additional witnesses were the friend and the mother of the complainant with the strong case evidence. The addition of these two prosecution witnesses did not significantly increase conviction rates or ratings of the defendant's factual culpability, thus providing no support for the accumulation hypotheses. Most importantly, the presence of these witnesses did not increase the conviction rate for the count with the weakest evidence. Quantitative and qualitative analyses confirmed that jurors and juries appropriately distinguished between the same types of offence alleged by different complainants, based on the strength of their evidence. Although the analysis of the frequency of factual errors made in deliberations revealed that juries who were exposed to more witnesses made more factual errors, as prior research has shown, this did not translate into a higher conviction rate. On the contrary, there was no difference in uncorrected or persistent errors across the trials. Together, these findings directly controvert the accumulation prejudice hypothesis by indicating that jurors and juries evaluated the evidence of multiple witnesses based on its probative value, not simply the number of witnesses.

Character prejudice

Character prejudice arises when the severity or number of allegations of criminal misconduct by the defendant is used to reason that the defendant is a person of bad character, and is therefore probably guilty of the current charges. Encompassed within this concept is the hypothesis that juries will be less concerned about convicting the defendant if they have knowledge of prior criminal misconduct, because they believe the defendant deserves punishment for the prior misconduct, charged or uncharged.

As discussed in Chapter 2, researchers typically test for the influence of character prejudice by analysing juror or jury evaluations of the defendant's character across trial types. In the present study, we tested mock jurors' ratings of the defendant's sexual interest in boys; the perceived convincingness of the defendant; the inferred criminal intent of the defendant; the factual culpability of the defendant; and the blameworthiness of the complainants. If juries were reasoning based on character prejudice, we would expect a decrease in the convincingness of

⁵⁵⁴ Trial 7 versus Trial 9.

the defendant in trials with tendency evidence, and a decrease in the reported cognitive effort based on heuristic reasoning associated with negative impressions of the defendant, rather than systematic weighing of the evidence that pertained to each count. Conversely, character prejudice should produce undifferentiated ratings of the defendant's criminal intent and factual culpability in the joint and tendency evidence trials, driven by negative impressions of the defendant. Below, we look at two examples from these lists of measures, convincingness and factual culpability.

We compared mock jurors' ratings of the extent to which they perceived the defendant to be convincing across different types of trials. The results revealed that jurors rated the defendant equally convincing in the basic separate trial, the tendency evidence trial and the joint trial. The admission of inculpatory evidence about four other acts of sexual abuse from two additional independent witnesses, irrespective of whether they were witnesses or complainants, did not diminish the ratings of how convincing the defendant was. This suggests that jurors were not engaging in impermissible reasoning on the basis of character prejudice. If they had, these ratings would have differed significantly between the separate trial and the trials with tendency evidence, where juries were exposed to evidence of the defendant's other acts of sexual abuse.

Comparing the factual culpability ratings for the conduct alleged by each of the three witnesses or complainants gave a further indication that juries did not reason globally and return parallel ratings about the defendant on measures of culpability in trials with tendency evidence, but made appropriate distinctions between the evidence and counts.

While useful in building a picture of jury reasoning and examining sources of potential prejudicial reasoning, one weakness of these measures in assessing the presence or absence of character prejudice is that any adverse inference drawn about the defendant could be validly formed *after* a systematic consideration of the probative value of the evidence. In other words, juries and jurors could have appropriately reasoned that the defendant had a tendency to commit sexual abuse as a result of finding that he had a sexual interest in boys, as directed by the trial judge. Hence, this conclusion would not necessarily be unfairly prejudicial to the defendant.

Moreover, while these findings suggest the absence of character prejudice, because those measures were retrospective and taken after the conclusion of the jury deliberations, we cannot definitively rule out the possibility that character prejudice played a causal role in the reasoning of some juries. An assessment of jury deliberations themselves can provide insight into whether

juries' discussion of the evidence was motivated by character prejudice in interpreting the evidence, and if so, the stage at which it was raised in the deliberation process.

To examine the potential effects of character prejudice reasoning within the juries, we thematically analysed jury deliberations in trials where juries were exposed to tendency evidence, to identify whether character prejudice influenced their decision making.

A quantitative and qualitative content analysis of jury deliberations revealed that no juries in either the tendency evidence or joint trials used the tendency evidence to conclude that the defendant was guilty because of the number of allegations of prior misconduct he was facing. Furthermore, there was no evidence that emotional reactions to the severity of the allegations – such as a sense of horror regarding the allegations, or a desire to punish the defendant – drove the verdicts. To the contrary, we found evidence that juries were more reluctant to convict the defendant for the counts pertaining to the most serious allegations of sexual intercourse than for indecency.

Summary

In sum, the low frequency and isolated examples of reasoning that involved inter-case conflation of the evidence, accumulation prejudice, or character prejudice suggests that the likelihood of impermissible reasoning, whether in joint or separate trials, is exceedingly low. This low probability suggests a negligible risk of unfair prejudice to the defendant in joint trials or trials in which tendency evidence is admitted.

5.3 Legal safeguards against prejudice

As discussed in Chapter 2, the law attempts to curb, via judicial directions, the perceived unfairly prejudicial effect of joint trials and evidence of a defendant's other misconduct. We examined the extent to which judicial directions and/or fact-based question trails controlled impermissible reasoning by juries.

The influence of directions on jury reasoning

Mock jurors who received the context evidence direction reported that it was harder to assess the credibility of the witnesses and to apply the law than those who did not. In the relationship evidence trial, the directions increased mock jurors' perceptions of the defendant's culpability, as reflected in increased ratings of the factual culpability of the defendant. Overall, the context evidence direction did help juries overcome their reluctance to convict for the penetrative

offence, although conviction rates for the non-penetrative offence were unaffected by the direction.

Juries may avoid impermissible propensity reasoning when given clear judicial directions about the use of cross-admissible tendency evidence in a joint trial.⁵⁵⁵ These directions, if followed, allow the jury to engage in permissible tendency or propensity reasoning in a joint trial. To consider the validity of this viewpoint, we examined the influence of a tendency evidence direction in a separate trial with one complainant plus two tendency evidence witnesses, and in a more complex joint trial with three complainants whose evidence was cross-admissible as tendency evidence. The purpose of these analyses was to determine whether the tendency evidence direction was effective in reducing the extent to which juries would engage in impermissible reasoning. We compared tendency evidence trials with and without a tendency direction⁵⁵⁶ and joint trials with and without a tendency direction⁵⁵⁷, focusing on the counts involving Timothy that were common to both trials (Counts 1 and 2 in the tendency evidence trials; Counts 5 and 6 in the joint trials) and examining jury perceptions of the factual culpability and criminal intent of the defendant.

We found no differences in the ratings of perceived criminal intent or of factual culpability for each of the counts when standard directions were given – compared to when a tendency direction was given – in either the tendency evidence trial or joint trial.⁵⁵⁸ These results were consistent with findings from our deliberation analysis, which found that many juries appeared to either ignore or misunderstand the tendency direction and, consequently, failed to apply it or misapplied it. Specifically, when we analysed the jury deliberations to test whether or not juries that received the tendency evidence direction were more likely to get the law on tendency evidence correct and apply the direction correctly, we found that juries that received the tendency evidence direction were more likely to say correct things about tendency evidence but still made the same number of mistakes when using the tendency evidence as did juries that deliberated without this direction.

Nonetheless, juries reported that they required increased cognitive effort when given a tendency evidence direction compared to standard judicial directions. Specifically, mock jurors perceived that more cognitive effort was required to understand the charges against the

⁵⁵⁵ See Part 2.4.⁵⁵⁶ Trial 5 versus Trial 6.

⁵⁵⁶ Trial 5 versus Trial 6.

⁵⁵⁷ Trial 7 versus Trial 10.

⁵⁵⁸ Trials 5 versus 6 and 7 versus 10.

defendant, remember the case facts, weigh the evidence and evaluate the prosecution case. By contrast, in the joint trials, there were no differences in the overall reported cognitive effort expended by mock jurors who received standard judicial directions versus tendency evidence directions.

When the perceived cognitive effort measures were considered individually in the joint trial, two significant differences emerged. In the joint trial with the tendency evidence direction, jurors reported requiring more cognitive effort to understand the judge's directions compared to those in the joint trial who did not receive that direction. Overall, it appeared that the tendency evidence direction was not only difficult to understand, it was difficult to apply – no doubt because the direction, based on accepted legal practice, was not written in plain English and was comprised of dense, legal language that, anecdotally, appears to pose comprehension problems for lawyers as well.

We also found that juries not given a tendency direction in the joint trials were more likely to discuss the meaning of 'beyond reasonable doubt' than juries that received this direction, suggesting that they experienced more confusion or disagreement about how to weigh the evidence. Receiving the tendency direction may have helped clarify deliberations in this respect.

Overall, we found that although mock jurors perceived that increased cognitive effort was required when they had received a tendency evidence direction, this increased effort was not sufficient for juries to apply the instructions as directed. This outcome may favour the defence case rather than the prosecution case. If juries had applied the tendency evidence direction as instructed by the trial judge, we would have seen *increased* ratings of factual culpability and criminal intent in the trials with the tendency evidence direction, because there would be a greater likelihood of more jurors finding that the defendant had a sexual interest in one or more of the complainants, and they would have used that finding of sexual interest in their assessment of the defendant's culpability and criminal intent. Instead, the dense legal language of the tendency evidence direction hampered the prosecution case, which was based on proving that the defendant had a tendency to have a sexual interest in young boys under the age of 12; to engage in sexual activities with young boys under the age of 12; and to use his position of authority as a soccer coach to gain access to young boys under the age of 12 so that he could engage in sexual activity with them. As a result, there were few instances of explicit permissible tendency reasoning in accordance with the tendency direction.

Making specific recommendations about possible modifications to current jury directions was beyond the scope of this study. To develop more effective jury directions on appropriate uses of relationship evidence and tendency evidence, further research is required to empirically compare juries' understanding and application of modified instructions in the context of a simulated trial.

The influence of question trails on jury reasoning

The use of question trails has not previously been tested in the joint trial context, but existing research suggested that they could potentially help juries assess the evidence and reach a verdict.

The question trail helped juries in the relationship evidence trial reach a verdict more rapidly, on average 25 minutes faster than in the absence of the question trail. A second feature of the question trails, apparent from the content analysis of the deliberations, was that the question trail increased the proportion of time that juries devoted to discussing the specific counts.

Juries needed more time to reach a verdict in the separate trial with relationship evidence despite the fact that it involved less complex evidence, fewer witnesses and fewer counts than the joint trial (which included three times as many counts and more witnesses, as well as complainants who presented similar evidence that had to be distinguished). The relationship evidence about uncharged acts may have confused the juries and raised more ambiguities and issues about the credibility of the complainant, requiring more time to reach a decision than the convergent evidence in the more complex joint trial. The decrease in the conviction rate in the relationship evidence trials, together with the finding that less mental effort was required to reach a unanimous verdict when a question trail was used, suggests that the question trail helped juries resolve their doubts about the use of that evidence by voting to acquit.

One essential difference between the question trails and the other judicial directions was that the question trails defined the charged conduct at a less abstract level. Whereas the jury direction referred to the charged conduct as an act of indecency or as sexual intercourse – without specifying the incidents or events that matched those abstractions – the question trails referred to the specific conduct in issue. This more concrete presentation also reduces the potential for intra-case conflation of the evidence, as the jury is not left to resolve which of the acts described in the evidence is the charged conduct referred to in the more abstract definitions.

The use of question trails to assist juries appears promising. We recommend further research to examine the benefits to juries of a structured decision aid such as a fact-based question trail.

5.4 General conclusions about prejudice in joint trials

Although it was expected that more complex trials with tendency evidence would result in unfair prejudice to the defendant, we found more evidence of impermissible reasoning in the basic separate trial and in the relationship evidence trial than in the more complex tendency evidence and joint trials. In the separate trials, for example, juries were more likely to believe that there was an onus on the defendant to prove his innocence. This finding is a crucial outcome of the study.

Overall, the results of this study showed that unfair prejudice to a defendant in the form of impermissible reasoning – as a consequence of joinder of counts or the admission of tendency evidence – was unlikely. Given the low probability, there is negligible risk to the defendant of a conviction based on reasoning logically unrelated to the evidence.

Chapter 6: Conclusion

This study examined the risk of unfair prejudice to a defendant who is the subject of allegations by multiple complainants whose claims are tried either jointly or separately before a jury.

Our study involved a more rigorous test of joinder effects and unfair prejudice to the defendant than in prior research, because the similarities in the allegations against the defendant in this study exceeded degrees of similarity previously tested.

The project revealed a number of major outcomes:

- Collectively, mock juries were capable of distinguishing between the counts, and of basing their verdicts on the evidence that pertained to each count whether it was presented in a separate or a joint trial.
- The perceived credibility of the complainant predicted the culpability of the defendant.
- A complainant's credibility was enhanced when supported by evidence from an independent source.
- The same benefit to a complainant's credibility was obtained by admitting tendency evidence in a separate and a joint trial.
- Overall, jury reasoning and verdicts were logically related to the probative value of the evidence.
- The results provided little indication of mock juries being susceptible to a joinder effect, and, even if there was such an effect, there was no evidence that decisions to convict were the result of impermissible propensity reasoning.
- Given that the verdicts were not based on impermissible reasoning, there was no evidence of unfair prejudice to the defendant.

The Royal Commission identified that the perceived risk of unfair prejudice to the defendant in joint trials for child sexual abuse cases was a significant issue that required further research. The Commission supported the current project to generate empirical evidence that would inform this issue. Specifically, this study examined the risk of unfair prejudice to a defendant who is the subject of allegations by multiple complainants, when those claims are tried either jointly or separately before a jury.

6.1 Strengths of the study

The present study provided a more rigorous test of joinder effects and of unfair prejudice to the defendant than previous research, because the similarities of the allegations against the defendant in this study exceeded the degrees of similarity tested in previous studies of joinder. Juries in joint trials were required to distinguish evidence pertaining to six similar offences alleged by three different complainants.

To inform this question, we conducted a robust controlled experimental study using a large representative sample of jury-eligible citizens who were randomly allocated to one of 90 mock juries, and then deliberated to a verdict after viewing a realistic simulated video trial. Commentators have observed that over-reliance on verdict alone as a dependent measure is a weakness of much previous jury research⁵⁵⁹, as it is uninformative about the comprehension and reasoning processes used by individual jurors and juries.⁵⁶⁰ In this study, we supplemented conviction rates with a range of other measures to gain insight into jury reasoning. We scrutinised jury reasoning in a separate or a joint trial and analysed it using quantitative and qualitative methods, as well as sophisticated multi-level analyses that provided insight into the factors that influenced jury decision making and verdicts. This componential research design laid bare patterns and trends in jury decisions, and the specific impact of joinder was distinguished from the probative value of the evidence, the influence of the number of counts, and the nature and number of the witnesses.

We specifically looked for instances of verdicts driven by inter-case conflation of the evidence, reasoning by accumulation prejudice or character prejudice. Across four different types of trials, no convictions were made on those bases, and few mock juror comments reflected emotionally motivated, superficial or impressionistic considerations. Overall, jury reasoning and verdicts were logically related to the probative nature of the admitted evidence.

⁵⁵⁹ Mark Findlay and P Duff ‘*Jury Management in New South Wales*’ (1994) Australian Institute of Judicial Administration.

⁵⁶⁰ Ede and Goodman-Delahunty, above n 16, 118.

6.2 Major outcomes

The project produced a number of major outcomes:

- Collectively, mock juries were capable of distinguishing between the counts, and of basing their verdicts on the evidence that pertained to each count whether it was presented in a separate or a joint trial.
- The perceived credibility of the complainant predicted the culpability of the defendant.
- A complainant's credibility was enhanced when supported by evidence from an independent source.
- The same benefit to a complainant's credibility was obtained by admitting tendency evidence in a separate and a joint trial.
- Overall, jury reasoning and verdicts were logically related to the probative value of the evidence.
- The results provided little indication of mock juries being susceptible to a joinder effect, and, even if there was such an effect, there was no evidence that decisions to convict were the result of impermissible propensity reasoning.
- Given that the verdicts were not based on impermissible reasoning, there was no evidence of unfair prejudice to the defendant.

6.3 Limitations of the study

As with any research, this study has a number of limitations which we must acknowledge.

In Chapter 5, we noted that the experimental trial simulation method required that the juries attend a trial, deliberate to a verdict and complete a post-trial questionnaire, all in under four hours. 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.

A further difference between the experimental procedure in this study and that of a real jury is that the mock jurors were not instructed to elect a jury foreperson before commencing their deliberations. Unlike the practice in some international jurisdictions, in New South Wales juries elect a foreperson at the outset of the trial, and the jury is allowed to discuss the evidence on a daily basis as the trial proceeds. By the time the trial concludes and the jury deliberates to a verdict, the jury members are well acquainted. In this study, since none of the mock jurors had the opportunity to become acquainted before deliberations commenced, and because the time available for deliberation was limited, the verdict form was given to a juror selected on a random basis, and the juries commenced deliberation without taking time to elect a foreperson. Discussion leaders emerged spontaneously and rapidly, and none of the deliberations were adversely affected by this procedure.

In this study, unlike in real juries, the participating mock jurors were forewarned that their decisions would be subjected to some scrutiny. Although they were unaware of the specific purpose of the study and the research hypotheses, it is possible that they adjusted their reasoning and behaviours because they knew their verdicts were going to be carefully analysed. It is unknown whether this awareness caused some mock jurors to suppress any indication that their verdicts were based on impermissible reasons.

When a research hypothesis is not confirmed, as in the case of the joinder effect, there is a risk of erroneously over-interpreting the null results. In other words, failure to confirm a research hypotheses does not establish the absence of the phenomenon tested. To avoid misinterpreting the results in this study by erroneously underestimating the extent to which impermissible reasoning motivated the mock jury decisions, we adopted a research strategy that (a) used multiple convergent measures in the statistical analyses to provide confirmation of the findings and reduce reliance on any single observation or measure; (b) applied different methodologies to examine the same question from different perspective – both quantitative and qualitative methods; and (c) analysed the data at the level of the individual mock jurors as well as the jury groups.

Some researchers place more reliance on individual mock jurors' responses as indicative of how real jurors would reason, and others give more emphasis to the contribution of the group process and collective decision making in deliberation. Accordingly, this study took both individual and group measures into account. One limitation of individual mock jurors' post-deliberation assessments is that they are retrospective, provided after the fact. Even though we

gathered these measures immediately following the jury deliberations, they were not as immediate and direct as analyses of the actual deliberation transcripts. Results of one prior study on joinder showed that it was during the discussion of the charges in the deliberation phase of the joint trial that mock jurors developed a ‘criminal schema’ or negative impression of the defendant.⁵⁶¹ This finding underscored the importance of including the group deliberation process when assessing the extent of any prejudicial reasoning. However, analyses of the group deliberations – while they provided a more direct and immediate indication of the juries’ reasoning and decision processes than individual mock jurors’ post-deliberation responses – do not necessarily capture the views of all individual mock jurors, as some mock jurors may speak more than others in the group setting, and some may speak more frankly and openly than others.

Where mock jurors’ indirect self-reported measures were used – for example to gather information about the perceived cognitive load of the jury tasks – these needed to be distinguished from more direct measures of cognitive load. Similarly, a report from mock jurors that they applied a certain threshold of proof is not necessarily the same as that threshold actually being applied.

In this project, despite the comparison of four different types of trial, all the trial versions tested pertained to the same core case facts and witnesses. While this was necessary to control the influence of extraneous factors, it may also have limited the extent to which the results of the study will generalise to other cases.⁵⁶² Thus, we recommend further research using a different set of stimulus materials to test whether the results can be replicated.

6.4 Implications for the criminal justice system

Extensive prior research has shown that acquittal rates in child sexual abuse cases that depend on word-against-word evidence of a single complainant are very high. A critical outcome of this study was the key role of the perceived credibility of the complainants as a significant predictor of the culpability of the defendant for the offences which they alleged he had committed. A complainant’s credibility was enhanced by the presence of independent witnesses rather than by additional evidence from the same complainant, or by the joinder of the counts of multiple complainants.

⁵⁶¹ Tanford, Penrod and Collins, above n 41.

⁵⁶² Gary L Wells and Paul D Windschitl, ‘Stimulus Sampling and Social Psychological Experimentation’ (1997) 25 *Personality and Social Psychology Bulletin*, 1115–1125.

In this study, we found 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.

Appendices

Appendix A: Comparison of NSW juror and mock juror demographic characteristics (per cent)

Juror Characteristics	Royal Commission Study Sample	NSW Dept of Justice ¹	Jury Samples in Prior Studies						
			NSW jurors ²	NSW jurors ³ (CSA trials)	NSW jurors ⁴	Vic, NSW and Qld jurors ⁵	NSW, SA and Vic empanelled jurors ⁶	NSW, SA and Vic non-empanelled jurors ⁷	Victorian civil jurors ⁸
Jurisdiction	NSW jury eligible citizens	NSW jurors	NSW jurors ²	NSW jurors ³ (CSA trials)	NSW jurors ⁴	Vic, NSW and Qld jurors ⁵	NSW, SA and Vic empanelled jurors ⁶	NSW, SA and Vic non-empanelled jurors ⁷	Victorian civil jurors ⁸
Sample	1029 mock jurors	9989 jurors	1225 jurors (112 juries)	277 jurors (25 juries)	78 jurors (20 trials)	296 jurors (55 juries)	628 jurors (71 juries)	1048 jurors	398 jurors
Gender (percent) ⁹									
Male	43.6	53	50.8	54.2	35.9	51.7	33.0	48.1	48.2
Female	56.4	47	49.2	45.8	64.1	48.3	35.7	51.9	51.8
Age									

¹ Royal Commission advice from NSW Department of Justice re empanelled jurors 14 March 2015 to 14 March 2016.

² Lily Trimboli, 'Juror Understanding of Judicial Instructions in Criminal Trials' (2008) 119 *Contemporary Issues in Crime and Justice*.

³ Judy Cashmore and Lily Trimboli, 'Child Sexual Assault Trials: A Survey of Juror Perceptions' (2006) 102 *Contemporary Issues in Crime and Justice*.

⁴ Jill Hunter, *Jurors' Notions of Justice: An Empirical Study of Motivations to Investigate & Obedience to Judicial Directions* (UNSW Jury Study, 2014).

⁵ Ian Freckleton, Jane Goodman-Delahunty, Jacqueline Horan and Blake McKimmie, *Expert Evidence and Criminal Jury Trials* (Oxford University Press, 2016).

⁶ Jane Goodman-Delahunty et al, 'Practices, Policies and Procedures that Influence Juror Satisfaction in Australia' (2008) *Australian Institute of Criminology*.

⁷ Ibid. Non-empanelled jurors are jurors who attended court for jury duty but were not empanelled to serve on a jury.

⁸ Jacqueline Horan and David Tait, 'Do Juries Adequately Represent the Community? A Case Study of Civil Juries in Victoria' (2008) 16(3) *Journal of Judicial Administration* 179.

⁹ Missing data excluded.

18-29	25.3	20				18.6	13.9		
30-39	17.1	16				15.9	15.6		
40-49	16.1	19				26.4	22.5		
50-59	23.0	22				21.0	21.8		
60+	18.5	23				12.8	14.2		
Education									
High school or less	15.2		21.2	23.8		20.6	39.0	33.6	
Certificate or diploma	24.1		37.3	35.3			29.5	29.4	
Bachelor's degree or higher	47.1		41.3	40.9	48.0	48.3	26.1	25.4	32.0
Occupation									
Employed	70.1	71	83.2	77.9		86.8	78.9	71.2	
Unemployed	6.5		3.7	4.4			1.6	1.0	
Retired	8.5		10.0	9.9			5.9	9.1	
Student	11.2		3.2	5.1			1.4	1.5	
Home duties	3.8			2.6			6.2	5.3	
Parental Status									
Parent	54.0			58.2					
Non-parent	44.1			41.8					

This table summarises comparable categories from a number of studies. Where categories were not comparable, they were not included in this table. Accordingly, not all totals in every section sum to 100%.

Appendix B: Ethics approval from Charles Sturt University



OFFICE OF
ACADEMIC GOVERNANCE

Panorama Mail Bag 381
Panorama Avenue
Bathurst NSW 2795
Australia

Tel: (61 2) 6333 4165
Fax: (61 2) 6333 4184

1 June 2015

Professor Jane Goodman-Delahunty
PO Box 168
MANLY NSW 1655

Dear Professor Goodman-Delahunty,

Thank you for the additional information forwarded in response to a request from the Human Research Ethics Committee (HREC).

The CSU HREC reviews projects in accordance with the National Health and Medical Research Council's *National Statement on Ethical Conduct in Research Involving Humans*.

I am pleased to advise that your project entitled "**Jury decisions in child sexual abuse cases**" meets the requirements of the *National Statement*, and ethical approval for this research is granted for a twelve-month period from 1 June 2015.

The protocol number issued with respect to this project is 2015/088. Please be sure to quote this number when responding to any request made by the Committee.

Please note the following conditions of approval:

- all Consent Forms and Information Sheets are to be printed on Charles Sturt University letterhead. Students should liaise with their Supervisor to arrange to have these documents printed;
- you must notify the Committee immediately in writing should your research differ in any way from that proposed. Forms are available at: http://www.csu.edu.au/research/ethics_safety/human/hrec_forms (please copy and paste the address into your browser)
- you must notify the Committee immediately if any serious and/or unexpected adverse events or outcomes occur associated with your research, that might affect the participants and therefore ethical acceptability of the project. An Adverse Incident form is available from the website: as above;
- amendments to the research design must be reviewed and approved by the Human Research Ethics Committee before commencement. Forms are available at the website above;
- if an extension of the approval period is required, a request must be submitted to the Human Research Ethics Committee. Forms are available at the website above;

Approval_after_further_information.doc

Last updated: March 2015
Next review: March 2016

www.csu.edu.au

CRICOS Provider Numbers for Charles Sturt University are 02005F (NSW), 01647D (VIC) and 02090B (WA). CRICOS 03 079 008 551

- you are required to complete a Progress Report form, which can be downloaded as above, by 15 April 2016 if your research has not been completed by that date;
- you are required to submit a final report, the form is available from the website above.

YOU ARE REMINDED THAT AN APPROVAL LETTER FROM THE CSU HREC CONSTITUTES ETHICAL APPROVAL ONLY.

If your research involves the use of radiation, biological materials, chemicals or animals a separate approval is required from the appropriate University Committee.

The Committee wishes you well in your research and please do not hesitate to contact the Executive Officer on telephone (02) 6338 4628 or email ethics@csu.edu.au if you have any enquiries.

Yours sincerely



Julie Hicks
Executive Officer
Human Research Ethics Committee
Direct Telephone: (02) 6338 4628
Email: ethics@csu.edu.au
Cc:

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007)

Approval after further information.doc

Last updated: March 2015
Next review: March 2016

Appendix C: Pre-trial questionnaire in the online juror study

Mock Jury Registration, 2015

Welcome and thank you for agreeing to participate in our survey. Please complete this task on a desktop or notebook computer, not a mobile device (iPad, iPhone, etc). If you have opened this survey on a mobile device, please close the survey and use the same link to access it on a desktop or notebook computer.

You are invited to take part in this research because we are interested in your opinions of evidence in child sexual assault cases. **Outcomes will assist in improving the legal system and protecting children from sexual abuse.**

If you have ever suffered sexual abuse as a child or as an adult, or are otherwise vulnerable to emotional distress, we advise you not to participate in the study because of the possibility that you may suffer emotional distress.

Participation is voluntary, anonymous and confidential, and you have the right to withdraw from the research at any time.

This survey is being conducted by Anomaly on behalf of the University of New South Wales and Charles Sturt University. If you have any queries about this particular survey, please email research@anom.com.au.

What will happen to information about me?

By clicking the 'I agree' button you consent to use by the research team of information you provide. No personal information collected in the questionnaire will be stored. The University of New South Wales and Charles Sturt University will keep this information for 5 years.

What if I want to withdraw from the study?

You can withdraw at any time before you submit your responses. Once you have submitted them, your responses cannot be withdrawn because they are anonymous and we will not be able to tell which are yours.

What should I do if I have questions about my involvement in the study?

<p>To contact the researcher about any aspect of this study:</p> <p>Prof Jane Goodman-Delahunty Charles Sturt University</p> <p>Tel: (02) 9752 9017</p> <p>Fax: (02) 9934 4830</p> <p>Email: jdelahunty@csu.edu.au</p>	<p>NOTE: The Human Research Ethics Committee at Charles Sturt University has approved this study. If you have any complaints or reservations about the ethical conduct of this project, you may contact the Committee to ask about project 2015/088.</p> <p>The Executive Officer Human Research Ethics Committee Charles Sturt University Tel: (02) 6338 4628 Fax: (02) 6338 4194 Email: ethics@csu.edu.au</p> <p>Any issues you raise will be treated in confidence, investigated fully, and you will be informed of the outcome.</p>
--	---

Consent Form

I have read the above information;

I understand the purposes, study task and risks of the research;

I have had an opportunity to ask questions and I am satisfied with the answers I have received;

I understand that personal details about me gathered in the course of this study are confidential and no identifying information will be used or published;

I freely agree to participate in this study;

I understand that I am free to withdraw my participation in the research at any time and that if I do I will not be subjected to any penalty or discriminatory treatment;

<p>I agree,</p> <p>start questionnaire</p>
--

Section 1 Demographic

Please complete the following.

S1. What is your date of birth?

DAY	OPEN TEXT
MONTH	OPEN TEXT
YEAR	OPEN TEXT

S2. Are you... [*SINGLE RESPONSE*]

Male	1
Female	2
Other	3

S3. Which of the following best describes you/ your current household?

[*SINGLE RESPONSE*]

Single with no children	1
Single parent	2
Young couple with no children	3
Young couple with children	4
Mid-life family (families with teenage children)	5
Mid-life household (family with independent children)	6
Mid-life couple with no children	7
Empty nester (parents who no longer live with their children)	8

S4. Professional or Employment Status [*SINGLE RESPONSE*]

Employed for wages (Full Time or Part Time)	
Self-Employed	
Out of work and looking for work	

Out of work and not currently looking for work	
A homemaker	
A student	
Military	
Retired	
Unable to work	

S5. What is the highest level of education you have completed? [*SINGLE RESPONSE*]

Primary School	
High School (HSC)	
Some university (not completed)	
Trade/technical/vocational training	
Bachelor's Degree	
Master's Degree	
Doctoral Degree	
Professional Degree (MD, JD, etc)	
Other (please specify)	

Section 2

Indicate your agreement with the following statements.

	Strongly agree						Strongly disagree
A suspect who runs from police, probably committed a crime.	1	2	3	4	5	6	7
A defendant should be found guilty if 11 out of 12 jurors vote guilty.	1	2	3	4	5	6	7
Too often jurors hesitate to convict someone who is guilty out of pure	1	2	3	4	5	6	7
In most cases where the accused presents a strong defense, it is only because of a good lawyer.	1	2	3	4	5	6	7
Out of every 100 people brought to trial, at least 75 are guilty of the crime	1	2	3	4	5	6	7
For serious crimes like murder, a defendant should be found guilty so long as there is a 90% chance that	1	2	3	4	5	6	7
Defense lawyers don't really care about guilt or innocence, they are just in business to make money.	1	2	3	4	5	6	7

Generally, the police make an arrest only when they are sure who	1	2	3	4	5	6	7
Many accident claims filed against insurance companies are phony.	1	2	3	4	5	6	7
The defendant is often a victim of his or her own bad reputation.	1	2	3	4	5	6	7
Extenuating circumstances should not be considered; if a person commits a crime, that person should	1	2	3	4	5	6	7
If the defendant committed a victimless crime like gambling or possession of marijuana, that person	1	2	3	4	5	6	7
Defense lawyers are too willing to defend individuals they know are	1	2	3	4	5	6	7
Police routinely lie to protect other police officers.	1	2	3	4	5	6	7
Once a criminal, always a criminal.	1	2	3	4	5	6	7
Lawyers will do whatever it takes, even lie, to win a case.	1	2	3	4	5	6	7
Criminals should be caught and convicted by any means necessary.	1	2	3	4	5	6	7
A prior record of conviction is the best indicator of a person's guilt in the	1	2	3	4	5	6	7
Wealthy individuals are almost never convicted of their crimes.	1	2	3	4	5	6	7
A defendant who is a member of a gang, is definitely guilty of the crime.	1	2	3	4	5	6	7
Ethnic minorities use the 'race issue' only when they are guilty.	1	2	3	4	5	6	7
When it is the suspect's word against the police officer's, I believe the	1	2	3	4	5	6	7
Men are more likely to be guilty of crimes than women.	1	2	3	4	5	6	7
The large number of Australian Aboriginals and Torres Strait Islanders currently in prison is an example of the innate criminality of	1	2	3	4	5	6	7
A black person on trial with a predominantly white jury will always	1	2	3	4	5	6	7
Ethnic minority suspects are likely to be guilty, more often than not.	1	2	3	4	5	6	7
If a witness refuses to take a lie detector test, that person is hiding something.	1	2	3	4	5	6	7
Defendants who change their story are almost always guilty.	1	2	3	4	5	6	7

Famous people are often considered to be above the law.	1	2	3	4	5	6	7
Every crime can be solved with forensic	1	2	3	4	5	6	7
If no forensic evidence is recovered then the defendant is probably	1	2	3	4	5	6	7
Every criminal leaves trace evidence at	1	2	3	4	5	6	7
Forensic evidence alone is enough to	1	2	3	4	5	6	7
No forensic evidence means investigators did not look hard enough.	1	2	3	4	5	6	7
Forensics always identifies the guilty	1	2	3	4	5	6	7
If there is no forensic evidence, the jury should not convict.	1	2	3	4	5	6	7
Forensics always provides a conclusive	1	2	3	4	5	6	7
Police should not charge someone without forensic evidence.	1	2	3	4	5	6	7
Science is always the most reliable way to identify perpetrators.	1	2	3	4	5	6	7
A medical examination almost always shows whether or not a child	1	2	3	4	5	6	7
A child who has been sexually abused will tell someone soon	1	2	3	4	5	6	7
Child victims of sexual abuse respond in a similar way to the abuse.	1	2	3	4	5	6	7
Children sometimes make false claims of sexual abuse to get back at an adult.	1	2	3	4	5	6	7
A child victim of sexual abuse will avoid his or her abuser.	1	2	3	4	5	6	7
A sexually abused child typically cries out for help and tries to escape.	1	2	3	4	5	6	7
Repeating questions such as: "What happened? What else happened?" leads children to make false abuse claims.	1	2	3	4	5	6	7
Children are easily coached to make false claims of sexual abuse.	1	2	3	4	5	6	7
Children who are sexually abused display strong emotional reactions	1	2	3	4	5	6	7
Children sometimes make up stories about having been sexually abused when	1	2	3	4	5	6	7
Children are sometimes led by an adult to report they have been sexually abused	1	2	3	4	5	6	7
Children aged 7-8 years are easily manipulated to give false reports of	1	2	3	4	5	6	7
Sexual assault is the worst offence that can be committed against a child.	1	2	3	4	5	6	7
Child sexual offenders deserve life	1	2	3	4	5	6	7

Appendix D: Post-trial questionnaire in the online juror study⁵⁷²

Are you satisfied beyond reasonable doubt that Mark Booth, the accused	
Count 1 – masturbated Simon Rutter 's penis between 1 March and 1 September 1993?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 2 – forced Justin McCutcheon to touch and masturbate the accused's penis between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 3 – masturbated Justin McCutcheon 's penis between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 4 – inserted his penis into Justin McCutcheon 's mouth between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 5 – masturbated Timothy Lyons ' penis between 1 and 31 December 1997?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 6 – inserted his finger into Timothy Lyons 's anus between 1 and 31 December 1997?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>

Select the number that best represents your answer.							
	Not at all						Very much
How confident are you about your verdicts?	1	2	3	4	5	6	7
Are you satisfied that the evidence proved beyond reasonable doubt that the accused had a sexual interest in boys?	1	2	3	4	5	6	7
Did Simon Rutter 's testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Justin McCutcheon ?	1	2	3	4	5	6	7

⁵⁷² This is the questionnaire for Condition 6. Questionnaires for the other conditions differed regarding what complainants were referred to in the questions.

influence your verdicts in relation to the counts involving Timothy Lyons ?	1	2	3	4	5	6	7
Did Justin McCutcheon's testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Simon Rutter ?	1	2	3	4	5	6	7
influence your verdicts in relation to the charges involving Timothy Lyons ?	1	2	3	4	5	6	7
Did Timothy Lyons' testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Simon Rutter ?	1	2	3	4	5	6	7
influence your verdicts in relation to the charges involving Justin McCutcheon ?	1	2	3	4	5	6	7

The accused is presumed innocent unless the prosecution has proved his guilt beyond reasonable doubt.

What number between 0% and 100% represents “beyond reasonable doubt”.

_____ %

Please fill in the blanks.

(1) What was the main reason for your verdict?

(2) What other factors went into your decision?

Select the number that best represents your mental effort to							
	Very low mental effort					Very high mental effort	
understand the charges against Mark Booth.	1	2	3	4	5	6	7
remember the facts.	1	2	3	4	5	6	7
assess the credibility of the witnesses.	1	2	3	4	5	6	7
weigh the evidence.	1	2	3	4	5	6	7
evaluate the prosecution case.	1	2	3	4	5	6	7
evaluate the defence case.	1	2	3	4	5	6	7
understand the judge's instructions.	1	2	3	4	5	6	7
apply the law to the facts.	1	2	3	4	5	6	7
reach a verdict.	1	2	3	4	5	6	7

How likely is it that Mark Booth							
	Very unlikely					Very likely	
- masturbated Simon Rutter 's penis.	1	2	3	4	5	6	7
- forced Justin McCutcheon to touch and masturbate Mark Booth's penis.	1	2	3	4	5	6	7
- masturbated Justin McCutcheon 's penis.	1	2	3	4	5	6	7
- inserted his penis into Justin McCutcheon 's mouth.	1	2	3	4	5	6	7
- masturbated Timothy Lyons ' penis.	1	2	3	4	5	6	7
- inserted his finger into Timothy Lyons 's anus.	1	2	3	4	5	6	7

Which statement is <u>false</u> ?	
Mark Booth	
<input type="checkbox"/>	generally behaved like one of the boys.
<input type="checkbox"/>	took selected boys for lunch after training.
<input type="checkbox"/>	invited selected boys to lunch at his house.

Indicate your agreement with the following statements.

The accused <u>Mark Booth</u>	Strongly disagree						Strongly agree	
cared for boys when their parents needed assistance.	1	2	3	4	5	6	7	
was convincing.	1	2	3	4	5	6	7	
abused the trust of others.	1	2	3	4	5	6	7	
abused his position as a coach.	1	2	3	4	5	6	7	
was responsible for what happened to him.	1	2	3	4	5	6	7	
is a risk to other boys.	1	2	3	4	5	6	7	

Timothy Lyons

was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7

Simon Rutter

was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7

Justin McCutcheon

was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7

The trial was fair to

Mark Booth	1	2	3	4	5	6	7
------------	---	---	---	---	---	---	---

Timothy Lyons	1	2	3	4	5	6	7
Simon Rutter	1	2	3	4	5	6	7
Justin McCutcheon	1	2	3	4	5	6	7

We would have been told if		
Other charges were made against Mark Booth.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth was sexually abusive on other occasions.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth had a prior conviction for child sexual assault.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth had a prior conviction for any other crime.	<input type="checkbox"/> Yes	<input type="checkbox"/> No

Appendix E: Online juror study

Research aims and methodology for the online juror study

Research aims

The online mock juror study was a pilot study, designed to pre-test the measures and materials for the main jury study, and in particular, the relative perceived strength of the evidence of each of the three complainants.

The specific aims of the online juror study were to:

1. test whether mock jurors differentiated between the strength of evidence in the individual cases of complainants with weak, moderately strong and strong evidence, respectively, in the manner intended;
2. compare the evidence strength of a joint trial (involving three complainants with different evidence strength: weak, moderately strong and strong) against a separate trial involving one complainant with moderately strong evidence; and
3. test whether the study instruments were sensitive to differences in juror responses to the type of trial (separate vs. joint) without the influence of group deliberations.

Method

Participants

A total of 160 men and 140 women aged 18 to 87 years ($M = 48.58$, $SD = 17.22$) participated in the online mock juror study. Half of the participants (hereafter referred to as online jurors) were employed for wages, 20.7 per cent were retired, 15 per cent were unemployed for different reasons, 7.7 per cent were students, and 6.7 per cent were self-employed. Forty eight per cent of the online jurors had no children, 36.3 per cent had independent children or children who no longer resided with them, and 24.7 per cent had children who were teenagers or younger. Most of the online jurors reported completion of a tertiary degree (Bachelor's degree: 23.7%; Master's Degree: 11.3%, Professional Doctorate: 4.3%, Doctor of Philosophy: 2.3%), while 5.3 per cent reported ongoing attendance at university. Twenty-nine per cent had completed a trade certificate, technical or vocational training, 18.3 per cent had completed high school, 2.7 per cent primary school, and 3 per cent reported 'other' educational qualifications.

Materials

Trial materials

Using facts derived from a review of child sexual abuse trials, four mock trial scripts were prepared which contained allegations of sexual assault by a defendant who, as a soccer coach, had a relationship of care and control in an institutional setting with each complainant in the 1990s:

1. Separate trial with a single complainant (Simon) who gave weak evidence (779 words).
2. Separate trial with a single complainant (Justin) who gave strong evidence (1,203 words).
3. Separate trial with a single complainant (Timothy) who gave moderately strong evidence (1,217 words).
4. Separate trial with a single complainant (Timothy) who gave moderately strong evidence plus supporting relationship evidence about the prior grooming acts of the accused (1,285 words).
5. Separate trial with a single complainant (Timothy) who gave moderately strong evidence and supporting tendency evidence from two other witnesses (Simon and Justin) (2,129 words).
6. Joint trial with three complainants (Simon, Justin and Timothy) who gave weak, strong and moderately strong evidence, respectively (2,365 words).

For each trial, a summary trial transcript was prepared including opening statements by the prosecution and defence, evidence-in-chief of each complainant, cross-examination of each complainant; where relevant, evidence-in-chief of supporting witnesses, cross-examination of supporting witnesses; evidence-in-chief of the defendant and cross-examination of the defendant, followed by brief instructions by the judge about the prosecution's burden of proof and the elements of each count.

In the trial summaries of the three separate trials, each single complainant (who gave either weak, moderate or strong evidence), presented the same evidence as that presented by the three complainants in the joint trial transcript. In the separate trials, no evidence about the charges arising from the other complaints was admitted.

The fourth trial involved the complainant with moderate evidence strength (Timothy) as well as evidence by Timothy of uncharged (grooming) acts by the accused. In the fifth trial,

Timothy also gave evidence, along with evidence from two supporting witnesses (Simon and Justin) about how the defendant had sexually assault them. The joint trial involved all three complainants (Simon, Justin and Timothy) and the same defendant. The trial summaries can be found in the online materials on the Royal Commission website.

Questionnaire

The pre-trial questionnaire was used to assess online juror attitudes and biases via the Child Sexual Assault Knowledge Questionnaire,⁵⁷³ the Pre-trial Juror Attitude Questionnaire,⁵⁷⁴ and the Forensic Evidence Evaluation Bias Scale.⁵⁷⁵ The pre-trial questionnaire was the same as in the jury deliberation and reasoning study (see Appendix G).

The post-trial questionnaire sought the same information as that gathered by the post-trial questionnaire in the jury deliberation and reasoning study, except online jurors' factual recall and their perceptions of the credibility using the Observed Witness Efficacy Scale.⁵⁷⁶ Copies of the pre-trial and post-trial online juror questionnaires are attached in Appendices G and L.

Design

The online juror study employed a between-subjects design to test the influence of evidence strength (weak v. moderate v. strong evidence) and separate v. separate trials on online jurors' decisions. Online jurors were randomly assigned to one of six experimental groups (weak, moderate, strong case, relationship evidence, tendency evidence, or joint trial).

Procedure

Online jurors were recruited through a market research company and given a \$100 e-gift voucher for their participation. The study was conducted online, and participation was voluntary. After completing consent forms, online jurors were randomly assigned to one of the six experimental groups. First, online jurors responded to a pre-trial questionnaire (Appendix G) that tested their knowledge and misconceptions about child sexual assault (CSA-KQ), and other jury biases (PJAQ, FEEBS), and provided basic demographic information. Next, online jurors read a written trial summary (see online materials on the Royal Commission website) and completed the post-trial questionnaire to record their verdicts and views about the

⁵⁷³ Goodman-Delahunty, Martschuk and Cossins, above n 1.

⁵⁷⁴ Lecci and Myers, above n 149.

⁵⁷⁵ Smith and Bull, above n 151.

⁵⁷⁶ Cramer, DeCoster, Neal and Brodsky, above n 18.

key issues of interest in the case (Appendix L). Finally, the online jurors were debriefed and received payment vouchers for their time.

Analyses

Individual online jurors' responses to the pre-trial and post-trial questionnaires were analysed quantitatively using SPSS. Details of the statistical tests performed are presented in footnotes that include the specific p values indicating the significance level, and also the effect sizes, so that the magnitude of the observed effects can be assessed.

To test the impact of the experimental groups on dependent measures, analyses of variance (ANOVA) were conducted with continuous variables that were the composites of a group of items. Non-parametric Kruskal-Wallis-H tests were conducted with ordinal variables, that is, analyses of particular items on a rating scale.

Where ANOVAs were conducted, the reported coefficient η^2 reflects the magnitude of the effect: $\eta^2 = .01$ can be interpreted as a small effect size; $\eta^2 = .06$ as a medium effect size, and $\eta^2 = .14$ as a large effect size.⁵⁷⁷ The statistical significance test is the F -test. Where Kruskal-Wallis-H tests were conducted, the statistical test of significance is the Chi-square value. When necessary, post-hoc analyses were conducted using the nonparametric Mann-Whitney-U test, and the z -approximation. To control for Type-I-Error, Bonferroni adjustments were applied to the alpha (α) values by dividing $\alpha = .05$ by the number of tests conducted in the post-hoc analyses.⁵⁷⁸ The magnitude of the effect is reflected in the r value, with $r = .1$ shows a small effect, $r = .3$ a medium effect, and $r = .5$ a large effect.⁵⁷⁹

To test within-group effects in the joint trial, one-way repeated measures ANOVAs were conducted with continuous variables and nonparametric Friedman tests with ordinal variables. For repeated measures ANOVA, the reported partial eta squared (η_{partial}^2) reflects the magnitude of the effect which is interpreted as the η^2 above. Wilk's Lambda is the multivariate test for significance of within-test measures. A significant result indicates a change in the reported scores between the different parallel measures. The nonparametric Friedman test is based on

⁵⁷⁷ Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum Associates, 2nd ed, 1988).

⁵⁷⁸ Julie Pallant, *SPSS: Survival Manual* (Open University Press, 4th ed, 2010).

⁵⁷⁹ Cohen, above n 583.

the Chi-square value. Nonparametric post-hoc analyses were conducted using the Wilcoxon Signed Rank test with the Bonferroni adjusted $\alpha = .017$.

Where regression analyses were conducted, the regression coefficient β reflects the strength of the impact of an independent or predictor variable on an outcome or dependent variable.⁵⁸⁰ The statistical significance test for regression analyses is the Z-test. A Z value of 1.96 is equivalent to $p = .05$.

Preliminary analyses revealed the presence of five univariate and/or multivariate outliers. These were removed from further analyses.

The influence of online jurors' pre-trial dispositions

A series of correlational analyses were conducted to test the extent to which individual online juror pre-trial biases were associated with their post-trial responses. These results are displayed in Table 1. Mean scores by type of trial are displayed in Appendix F.

Results revealed a negative correlation between online jurors' pre-trial *CSA Knowledge* (i.e., the *Impact of SA on Children, Contextual Influences on Report*) and their pre-trial expectations and attitudes about forensic evidence (*Pro-Prosecution Bias, Pro-Defence Bias*). Similarly, online jurors' *CSA Knowledge* was negatively correlated with online juror pre-trial attitudes measured with the PJAQ (*Confidence in the Justice System, Conviction Proneness, Cynicism about the Defence, Social Justice, Racial Bias and Innate Criminality of Defendants*). The more online jurors knew about CSA, the less likely they were to endorse other types of pre-trial biases.

Post-trial responses revealed a negative correlation between online jurors' *CSA Knowledge* and *Positive* and *Negative Affect*. Conversely, online jurors' pre-trial *Pro-Prosecution* and *Pro-Defence Bias*, and other pre-trial biases as measured by the PJAQ were positively correlated with *Positive* as well as *Negative Affect*. The more online jurors knew about CSA, and the less they endorsed other types of pre-trial biases, the less likely they were to report either positive or negative affective responses.

Further, online juror *CSA Knowledge* was negatively correlated with the reported cognitive effort expended to reach decisions about the mock-trial. Similar to the above-mentioned results, online jurors' pre-trial *Pro-Prosecution Bias* (but not *Pro-Defence Bias*) and pre-trial

⁵⁸⁰ Geiser, above n 156.

attitudes measured with the PJAQ were positively correlated with post-trial reports of cognitive effort. The more accurate online jurors' *CSA Knowledge* was before the study, and the lower their susceptibility to other pre-trial biases, the less likely online jurors were to report that the mock-jury task required extensive cognitive effort.

Finally, online juror educational levels were positively correlated with their *CSA Knowledge* and negatively correlated with any type of pre-trial bias (except *Pro-Defence Bias*). Education was unrelated to the extent to which online jurors reported *Positive* or *Negative Affect* and expenditure of cognitive effort.

Table 1. Inter-correlations for Online Jurors Individual Pre-trial Biases and Post-trial Responses without Deliberation: Online Juror Study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. CSA-KQ Impact of CSA	-													
2. CSA-KQ Contextual Influences	.335**	-												
3. CSA-KQ Total	.857**	.772**	-											
4. FEEBS Pro-Prosecution	-.609**	-.269**	-.557**	-										
5. FEEBS Pro-Defence	-.530**	-.339**	-.542**	.617**	-									
6. PJAQ System Confidence	-.615**	-.233**	-.542**	.589**	.379**	-								
7. PJAQ Conviction Proneness	-.647**	-.284**	-.591**	.572**	.415**	.689**	-							
8. PJAQ Cynicism in the Defence	-.504**	-.377**	-.546**	.434**	.355**	.571**	.617**	-						
9. PJAQ Racial Bias	-.567**	-.448**	-.627**	.545**	.514**	.619**	.601**	.540**	-					
10. PJAQ Social Justice	-.339**	-.280**	-.382**	.313**	.422**	.290**	.298**	.475**	.260**	-				
11. PJAQ Innate Criminality	-.549**	-.335**	-.553**	.576**	.529**	.710**	.590**	.516**	.709**	.272**	-			
12. Positive Affect	-.231**	-.153**	-.239**	.208**	.195**	.192**	.230**	.143*	.250**	.116*	.164**	-		
13. Negative Affect	-.310**	-.114*	-.272**	.299**	.322**	.288**	.319**	.182**	.400**	.175**	.373**	.349**	-	
14. Cognitive Effort	-.159**	-.162**	-.196**	.184**	.098	.187**	.216**	.119*	.185**	.022	.126*	.142*	-.004	-
15. Education	.183**	.126*	.191**	-.149*	-.097	-.265**	-.236**	-.197**	-.228**	-.120*	-.176**	.006	-.002	.001

Note. * $p < .05$, ** $p < .01$. Pearson's correlation is reported for 1-16; Spearman's Rho for 17.

CSA-KQ = Child Sexual Assault Knowledge Questionnaire; FEEBS = Forensic Evidence Evaluation Bias Scale; PJAQ = Pre-trial Juror Attitude Questionnaire.

Online jurors' expectations about information on the defendant's criminal history

Results revealed that approximately every second online juror expected to be informed if other charges had been brought against the defendant, if the defendant had been sexually abusive on other occasions, or had prior convictions for CSA offences. Further, approximately 40 per cent of online jurors expected that they would have been informed if the defendant had a prior conviction for any other type of crime. These responses were independent of the type of trial. In other words, online jurors had expectations that they would have been informed of prior offending by the defendant⁵⁸¹ irrespective of the experimental condition or group to which they were assigned. The overall responses by trial type are reported in Table 2.

Table 2

Online Juror Expectations of Information about the Defendant's Criminal History (agreement in per cent): Study 13

We would have been informed if:	Overall	Separate trial	Relationship evidence	Tendency evidence	Joint trial
other charges were made against the defendant.	48.1	46.3	49.0	53.3	48.0
the defendant was sexually abusive on other occasions.	48.8	48.3	47.1	51.1	50.0
the defendant had a prior conviction for child sexual abuse.	46.8	47.0	43.1	51.1	46.0
the defendant had a prior conviction for any other crime.	40.0	40.9	39.2	42.2	36.0

⁵⁸¹ $p > .05$.

The influence of type of trial on online juror reasoning

In this section, results are presented for the four main types of trial (separate, relationship evidence, tendency evidence, and joint trials), focusing on online juror reasoning related to the complainant with the moderately strong case. Comparisons were made of the responses from a total of 192 online jurors, that is, 46 online jurors in the separate trial for the complainant with the moderately strong case, 51 online jurors in the relationship evidence trial, 45 online jurors in the tendency evidence trial and 50 online jurors in the joint trial.

In the analyses presented below, the main research question was the influence of variations of trial type on juror reasoning in the absence of any group influences arising from deliberations.

Criminal intent of the defendant

A composite measure of the criminal intent of the defendant was devised from responses indicating whether the defendant ‘abused the trust of others’, ‘abused his position as a coach’, ‘was responsible for what happened to him’, and was ‘a risk to other boys’. Scores reported for these items were added, comprising a single measure, based on a Principal Component Analysis which revealed that these responses were loading on the same component, sharing the same variance. Results of analyses using this composite variable provided insight into the influence of type of trial on inferences drawn by online jurors about the criminal intent of the defendant.

A one-way between-subjects ANOVA was conducted to test the influence of trial type on the perceived criminal intent of the defendant, comparing only the counts involving the complainant with the moderately strong evidence, as they were common to all trials. Results revealed no effect of type of trial on perceived criminal intent. Online jurors did not differ in their perceptions of the criminal intent inferred from the descriptions of the defendant’s conduct whether presented in the separate trial with moderately strong evidence, the separate trial with relationship evidence, the separate trial with tendency evidence, or the joint trial of three complainants.⁵⁸²

⁵⁸² $F(3, 192) = 1.308, p = .273, \eta^2_{\text{partial}} = .020$.

Perceptions of the defendant's sexual interest in boys

Online jurors' perceptions as to whether the evidence proved beyond reasonable doubt that the accused had a sexual interest in boys yielded a significant effect for type of trial.⁵⁸³ Post-hoc analyses using a Mann-Whitney-U test revealed significant differences between the separate versus the relationship evidence trial, and the separate versus the joint trial.⁵⁸⁴ Accordingly, online jurors were more likely to agree that the evidence proved beyond reasonable doubt that the accused had a sexual interest in boys in the separate trial with relationship evidence and in the joint trial than in the separate trial. Online jurors did not differentiate the defendant's sexual interest in boys in the separate versus the tendency evidence trials, and any other pairwise comparisons.

Victim blame for abuse alleged by the moderately strong complainant

The composite measure of victim blame was devised from responses indicating whether the complainant 'was responsible for what happened to him', 'misinterpreted the events', and 'should have recovered by now'. These scores comprised a single measure, based on results of a Principle Component Analysis which indicated that these three items shared the same variance, thus were loading on the same component.

A one-way between-subjects ANOVA revealed a significant effect of trial type on perceptions of victim blame for abuse alleged by the complainant with the moderately strong evidence.⁵⁸⁵ Post-hoc analyses of victim blame revealed a significant difference between blame attributed to the complainant in the separate trial and the separate trial with tendency evidence, and between the blame attributed to the complainant in the tendency evidence versus the joint trials.⁵⁸⁶ Specifically, online jurors were more likely to blame the complainant for what happened in the separate than the tendency evidence trial. Interestingly, they were also more likely to blame the complainant in the joint trial than in the tendency evidence trial, even though the sum of the evidence presented in those two trials was identical, the only difference being

⁵⁸³ $\chi^2(3, 192) = 8.208, p = .042$.

⁵⁸⁴ Separate trial v. Relationship evidence: $U = 832.0, z = -2.492, p = .013, r = -.253$; Separate trial v. Tendency evidence: $U = 852.5, z = -1.465, p = .143, r = -.154$; Separate v. Joint trial: $U = 804.0, z = -2.572, p = .010, r = -.263$; Relationship evidence v. Tendency evidence: $U = 1021.5, z = -0.938, p = .348, r = -.096$; Relationship evidence v. Joint trial: $U = 1214.0, z = -0.420, p = .674, r = -.042$; Tendency evidence v. Joint trial: $U = 1035.0, z = -0.680, p = .496, r = -.070$.

⁵⁸⁵ $F(3, 192) = 3.757, p = .012, \eta^2_{\text{partial}} = .057$.

⁵⁸⁶ Separate trial v. tendency evidence: $M_{\text{diff}} = 0.805, SE = .298, p = .038$; Tendency evidence v. joint trial: $M_{\text{diff}} = -0.904, SE = .292, p = .012$.

the number of formal charges, that is, the defendant faced two counts of sexual assault in the tendency evidence trial compared with six counts in the joint trial.

The convincingness of the key witness by type of trial

Comparisons were conducted to test the extent to which the key witness for the prosecution who testified in all four trials (Timothy) was perceived as convincing. The key evidence presented on his behalf was identical across all types of trial, but some trials included further evidence of acts of grooming by the defendant with Timothy (relationship evidence), while others included evidence of the defendant's sexual assault of two other witnesses (tendency evidence trial) or complainants (joint trial).

A Kruskal-Wallis-H test was not significant, indicating that online jurors perceived the complainant with moderately strong evidence (Timothy) as equally convincing, independent of the type of trial. Accordingly, online jurors' ratings of how convincing they found the complainant were independent of the type of trial and of the presence or absence of additional evidence.⁵⁸⁷

Factual culpability of the defendant by type of trial

Online jurors' perceptions of the factual culpability of the defendant for each of the alleged offences were analysed separately. A Kruskal-Wallis-H test revealed no effect of trial type on the defendant's perceived factual culpability for each count.⁵⁸⁸ Accordingly, the perceived factual culpability of the defendant for the two counts involving the complainant with moderately strong evidence (Timothy) was similar in all four types of trial.

Separate multiple regression analyses were conducted with online juror *CSA Knowledge*, perceived convincingness of the complainant, and trial type as predictors of factual culpability for the penetrative and non-penetrative offence against Timothy, respectively. The full model explained 53 per cent of the variance in factual culpability, respectively (non-penetrative offence: 53.5%; penetrative offence: 53.1%).⁵⁸⁹ Online juror *CSA Knowledge* and trial type did not predict the perceived culpability of the defendant.⁵⁹⁰ The sole variable to reach statistical significance was the perceived convincingness of the complainant. This variable alone

⁵⁸⁷ $\chi^2(3, 192) = 3.222, p = .359$.

⁵⁸⁸ Nonpenetrative offence: $\chi^2(3, 192) = 4.343, p = .227$; Penetrative offence: $\chi^2(3, 192) = 4.134, p = .247$.

⁵⁸⁹ Nonpenetrative offence: $F(3, 191) = 72.193, p < .001$; penetrative offence: $F(3, 191) = 70.939, p < .001$.

⁵⁹⁰ Nonpenetrative offence: *CSA Knowledge*: $\beta = .012, p = .804$; trial type: $\beta = .019, p = .698$.

Penetrative offence: *CSA Knowledge*: $\beta = .018, p = .722$; trial type: $\beta = .022, p = .665$.

explained 53 per cent of the variance in factual culpability, independently of the type of offence.⁵⁹¹ These results indicated a strong effect for the association of the credibility of the complainant and the culpability of the defendant. Online jurors who perceived the complainant to be more convincing were significantly more likely to find the defendant factually culpable for both the non-penetrative and penetrative offence.

Online juror reasoning about the counts and verdict consistency

Overall, 59.4 per cent of online jurors found the defendant guilty on Count 1 (masturbation of the complainant), and 58.9 per cent of online jurors found the defendant guilty on Count 2 (digital-anal penetration). Individual online juror verdicts were similar for both counts and across the different types of trial, as shown in Figure 18. The highest conviction rate emerged in the separate trial with relationship evidence, where the complainant gave evidence of uncharged acts of grooming and sexual misconduct by the defendant.

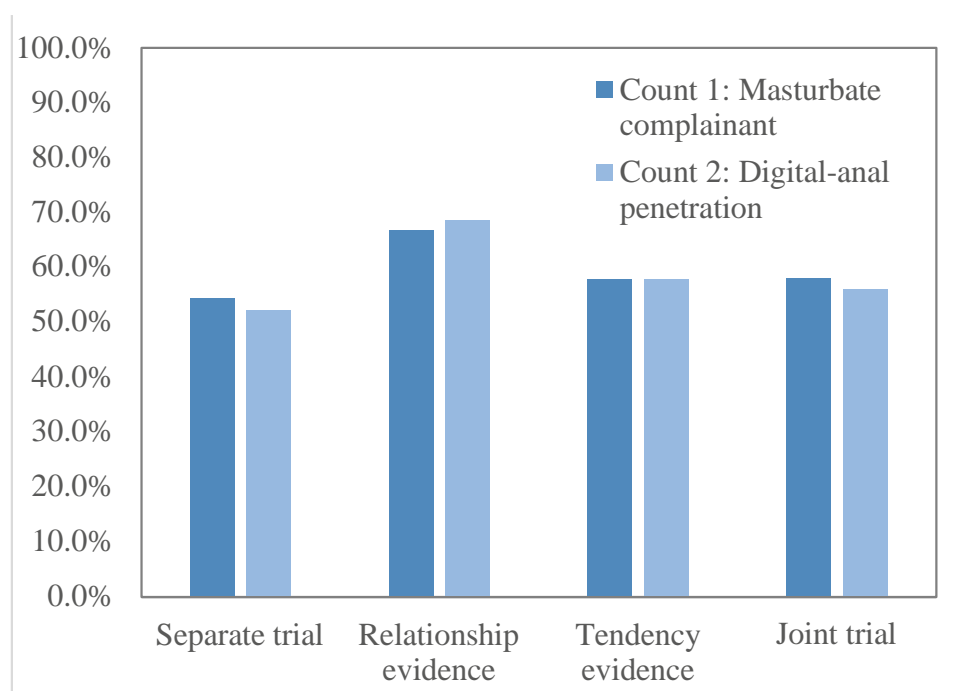


Figure 18. Online juror verdicts for non-penetrative and penetrative offences against the complainant with moderately strong evidence, by type of trial.

Separate binary regression analyses assessed verdict for penetrative and non-penetrative counts by type of trial after controlling for online juror *CSA Knowledge*, and taking perceived

⁵⁹¹ Nonpenetrative offence: $\beta = .730, p < .001$; penetrative offence: $\beta = .726, p < .001$.

convincingness of the complainant into account. Results revealed that *CSA Knowledge* and type of trial did not predict verdict.⁵⁹² The sole predictor to reach statistical significance was the convincingness of the complainant, indicating that online jurors who perceived the complainant as more convincing were 2.5 times more likely to convict the defendant than those who did not, and that this pattern held for both the non-penetrative and the penetrative offences. The model explained between 31.6 and 42.6 per cent of the variance on Count 1 and between 33.5 and 45.2 per cent of the variance on Count 2, and correctly classified a high proportion of the verdicts, 78.1 per cent and 80.2 per cent, respectively.⁵⁹³

Online juror reasoning in a separate vs. joint trial with cross-admissible evidence for weak,⁵⁹⁴ moderately strong,⁵⁹⁵ and strong claims⁵⁹⁶

The results reported in this section compare online juror reasoning in a separate trial versus a joint trial for claims of different strength. The results are discussed in relation to the complainant with weak evidence ($n = 50$ jurors), followed by the moderately strong evidence ($n = 46$ jurors), and the strong evidence ($n = 53$ jurors), compared to jury decisions in a joint trial, in which all three complainants testified ($n = 50$ jurors). The case facts were held constant for each of the complainants allowing interpretation of the results due to trial type rather than case facts for each of the complainants.

The purpose of the comparisons was to compare the effect of a single complainant versus multiple complainants with claims of different evidence strength. The analyses in this section report results for online juror *CSA Knowledge*, victim blame, factual culpability, and verdict.

Online Juror Decisions about the Complainant with the Weak Evidence⁵⁹⁷

The defendant was charged with one count of sexual assault which involved masturbation of Simon by the defendant (a non-penetrative offence). Chronologically, this alleged assault was

⁵⁹² Nonpenetrative offence: *CSA Knowledge*: Wald test (1, 192) = 0.005, $p = .942$, Odds Ratio = 1.002, 95% CI [0.954; 1.052]; Type of trial: Wald test (3, 192) = 2.840, $p = .417$; Convincingness of the complainant: Wald test = 45.960, $p < .001$, Odds Ratio = 2.448, 95% CI [1.890; 3.172].

Penetrative offence: *CSA Knowledge*: Wald test (1, 192) = 1.533, $p = .216$, Odds Ratio = 0.969, 95% CI [0.923; 1.018]; Type of trial: Wald test (3, 192) = 4.285, $p = .232$; Convincingness of the complainant: Wald test = 47.268, $p < .001$, Odds Ratio = 2.532, 95% CI [1.943; 3.301].

⁵⁹³ Nonpenetrative offence: $\chi^2(5, 192) = 72.847$, $p < .001$; penetrative offence: $\chi^2(5, 192) = 78.348$, $p < .001$.

⁵⁹⁴ Trials 1 v 6.

⁵⁹⁵ Trials 2 v 6.

⁵⁹⁶ Trials 3 v 6.

⁵⁹⁷ Trials 1 v 6.

the first (compared to those reported by the other complainants) and involved no witnesses. The evidence by Simon included a number of inconsistencies, and thus comprised relatively weak evidence against the defendant.

Victim blame for the alleged abuse

Multiple regression analysis assessed the impact of online juror *CSA Knowledge* and trial type on victim blame. The model was significant and explained 13.7 per cent of the variance in victim blame.⁵⁹⁸ Online juror *CSA Knowledge* significantly predicted victim blame, indicating that the more online jurors knew about child sexual abuse, the less likely they were to blame Simon for the alleged abusive events.⁵⁹⁹ Trial type was not significant, indicating that the blame attributed to the complainant was similar irrespective of whether his evidence was presented in a separate or a joint trial.⁶⁰⁰ Figure 19 displays perceived victim blame for all three complainants (with weak, moderately strong, and strong evidence).

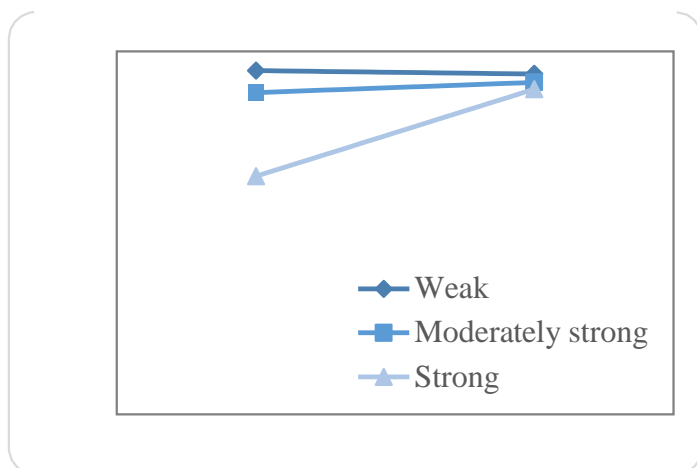


Figure 19. Perceived victim blame for complainants with weak, moderately strong and strong evidence by type of trial 2.

How convincing was the complainant by type of trial?

The extent to which the complainant with weak evidence was perceived as convincing was not dependent on the type of trial.⁶⁰¹ Online jurors perceived the complainant as convincing in the separate trial as in the joint trial, indicating that the additional evidence provided by two other

⁵⁹⁸ $F(2, 99) = 7.703, p = .001$.

⁵⁹⁹ $\beta = -.377, p < .001$.

⁶⁰⁰ $\beta = -.083, p = .389$.

⁶⁰¹ $U = 1041.0, z = -1.46, p = .143$.

complainants in relation to their own cases did not influence online jurors' perceptions of the complainant with weak evidence strength.

Factual culpability of the defendant

In contrast to victim blame, online juror *CSA Knowledge* did not predict the factual culpability of the defendant.⁶⁰² However, the type of trial was significant in that online jurors were more likely to hold the defendant factually culpable for the alleged offence against the complainant with the weak evidence (Simon) in the joint than in the separate trial.⁶⁰³ Thus, the evidence of the other two complainants in the joint trial benefited Simon by increasing the extent to which the defendant was viewed as factually culpable for his complaints of sexual assault.

Verdict

Analyses revealed that half as many online jurors (24 per cent) found the defendant guilty of the offence against Simon when their verdict was delivered in a separate rather than a joint trial (50 per cent). Hierarchical logistic regression analyses revealed that after controlling for online juror *CSA Knowledge*, trial type was a significant predictor of verdict.⁶⁰⁴ Specifically, online jurors were three times more likely to convict the defendant for the count involving Simon in the joint trial than in the separate trial. When convincingness of the complainant was added to the model, trial type became non-significant.⁶⁰⁵ Instead, online jurors who perceived the complainant as more convincing were 2.3 times more likely to convict than jurors who perceived the complainant as less convincing. The full model explained between 34.2 and 46.7 per cent of the variance in verdict and correctly classified 79.0 per cent of the jury decisions.⁶⁰⁶

The defendant was charged with two counts involving the complainant with moderately strong evidence (Timothy): (a) masturbation of the complainant, and (b) digital-anal penetration. Compared to the other two complainants, chronologically, these claims were the most recent and were supported by evidence in the form of a witness called by the prosecution who in part corroborated the Timothy's evidence about the location of both alleged acts of sexual assault. However, no eyewitnesses directly supported the evidence of Timothy.

⁶⁰² $\beta = -.088, p = .374$.

⁶⁰³ $\beta = .255, p = .012$.

⁶⁰⁴ *CSA Knowledge*: Wald test (1, 100) = 1.390, $p = .238$, Odds Ratio = 0.969, 95% CI [0.919; 1.021]; Type of trial: Wald test (1, 100) = 5.901, $p = .015$, Odds Ratio = 2.923, 95% CI [1.230; 6.945].

⁶⁰⁵ *CSA Knowledge*: Wald test (1, 100) = 0.720, $p = .396$, Odds Ratio = 0.973, 95% CI [0.913; 1.037]; Type of trial: Wald test (1, 100) = 3.293, $p = .070$, Odds Ratio = 2.673, 95% CI [0.924; 7.732]; Convincingness of the complainant: Wald test = 22.003, $p < .001$, Odds Ratio = 2.334, 95% CI [1.638; 3.326].

⁶⁰⁶ $\chi^2(3, 100) = 41.841, p < .001$.

Victim blame for the alleged abuse

Multilevel regression analysis revealed that online juror *CSA Knowledge* significantly predicted victim blame for the alleged abuse of Timothy.⁶⁰⁷ The greater an online juror's *CSA Knowledge* score, the less likely they were to blame the complainant for the alleged abuse. After controlling for *CSA Knowledge*, there was no effect of type of trial on victim blame.⁶⁰⁸ Blame of Timothy was equivalent in the separate and joint trials. The model containing both predictors was statistically significant and explained 8.5 per cent of the variance in victim blame.⁶⁰⁹

How convincing was the complainant by type of trial?

The extent to which the complainant with moderately strong evidence was perceived as convincing did not change with additional independent evidence from additional complainants.⁶¹⁰ Online jurors perceived the complainant with moderately strong evidence to be just as convincing in the separate trial as in the joint trial.

Factual culpability of the defendant

Perceived factual culpability of the defendant was assessed for each of the counts separately using multiple regression analyses. The results revealed that online juror *CSA Knowledge* and type of trial had no effect on the factual culpability of the defendant for the non-penetrative or the penetrative offence.⁶¹¹ The perceived convincingness of the complainant was a significant predictor in that the more convincing online jurors perceived the complainant to be, the more likely they were to perceive the defendant as factually culpable. The model including all three predictors explained 56 per cent of the variance in the factual culpability of the defendant for both claims.⁶¹²

Verdict

Over fifty per cent of online jurors found the defendant guilty on Count 1 (masturbation of the complainant: separate trial: 54.3 per cent; joint trial: 58.0 per cent) and on Count 2 (digital-anal

⁶⁰⁷ $\beta = -.291, p = .004$.

⁶⁰⁸ $\beta = .004, p = .968$.

⁶⁰⁹ $F(2, 95) = 4.324, p = .016$.

⁶¹⁰ $U = 1025.5, z = -0.93, p = .354$.

⁶¹¹ Nonpenetrative offence: *CSA Knowledge*: $\beta = .066, p = .344$; trial type: $\beta = .053, p = .452$; convincingness of the complainant: $\beta = .739, p < .001$.

Penetrative offence: *CSA Knowledge*: $\beta = .045, p = .523$; trial type: $\beta = .054, p = .440$; convincingness of the complainant: $\beta = .742, p < .001$.

⁶¹² Nonpenetrative offence: $F(3, 95) = 38.224, p < .001$ Penetrative offence: $F(3, 95) = 38.719, p < .001$.

penetration: separate trial: 52.2 per cent; joint trial: 56.0 per cent), with slightly fewer convictions for the penetrative than the non-penetrative offence. Logistic regression analyses revealed that online juror *CSA Knowledge* and type of trial had no influence on the conviction rate.⁶¹³ In line with the above-mentioned results, online jurors who perceived the complainant as convincing were 2.2 times more likely to convict the defendant on each of the counts than those who did not. The model with all three predictors explained between 25.6 per cent and 34.3 per cent of the variance for Count 1 and between 27.5 and 36.7 per cent of the variance in verdict for Count 2, and correctly classified 78.1 per cent of the verdicts, respectively.⁶¹⁴

Online juror decisions about the complainant with the strong evidence⁶¹⁵

The case for the complainant with the strong evidence (Justin) included two prosecution witnesses who supported the complainant's allegations of sexual assault. The defendant was charged with three counts involving Justin: Count 1 (masturbate defendant), Count 2 (masturbate complainant) and Count 3 (oral penetration).

Victim blame for the alleged abuse

Multiple regression analysis assessed the influence of online juror *CSA Knowledge* and type of trial on perceived victim blame. The model was significant and explained 10.8 per cent of the variance in victim blame.⁶¹⁶ In contrast to the weak and moderately strong cases, it was not online juror *CSA Knowledge* but type of trial that predicted victim blame for the alleged abuse of Justin.⁶¹⁷ The effect was, however, in an unexpected direction, in that online jurors attributed more blame to the complainant with the strong evidence for the alleged abuse he experienced in the joint trial compared to the separate trial.

However, inspection of the rates of victim blame attributed to all three complainants in the joint trial showed that online jurors blamed all three complainants equivalently.

⁶¹³ Nonpenetrative offence: *CSA Knowledge*: Wald test (1, 96) = 0.042, $p = .590$, Odds Ratio = 0.981, 95% CI [0.913; 1.053]; Type of trial: Wald test (1, 96) = 3.293, $p = .837$, Odds Ratio = 1.105, 95% CI [0.424; 2.880]; Convincingness of the complainant: Wald test (1, 96) = 19.719, $p < .001$, Odds Ratio = 2.172, 95% CI [1.542; 3.058].

Penetrative offence: *CSA Knowledge*: Wald test (1, 96) = 1.805, $p = .174$, Odds Ratio = 0.951, 95% CI [0.884; 1.023]; Type of trial: Wald test (1, 96) = 0.065, $p = .799$, Odds Ratio = 1.134, 95% CI [0.432; 2.976]; Convincingness of the complainant: Wald test (1, 96) = 20.161, $p < .001$, Odds Ratio = 2.222, 95% CI [1.568; 3.148].

⁶¹⁴ Nonpenetrative offence: $\chi^2(3, 96) = 28.415$, $p < .001$; Penetrative offence: $\chi^2(3, 96) = 30.826$, $p < .001$.

⁶¹⁵ Trials 3 v 6.

⁶¹⁶ $F(2, 102) = 6.070$, $p = .003$.

⁶¹⁷ *CSA Knowledge*: $\beta = -.133$, $p = .163$; trial type: $\beta = .301$, $p = .002$.

How convincing was the complainant by type of trial?

Analyses revealed no effect of trial type on perceived convincingness of the complainant.⁶¹⁸ The complainant with strong evidence was perceived as convincing in the separate trial, where he was the only complainant, as he was in the joint trial with two other complainants with weaker evidence.

Factual culpability of the defendant

Multiple regression analysis revealed no effect of online juror *CSA Knowledge* and type of trial on the perceived factual culpability of the defendant, independently of type of trial.⁶¹⁹ However, the more online jurors perceived Justin as convincing, the more likely they were to perceive the defendant as factually culpable for the alleged offences. These variables explained between 44 per cent (penetrative offence) and 55 per cent (non-penetrative offences) of the variance in factual culpability.⁶²⁰ The effect was greater for the non-penetrative offences than for the penetrative offence.

Verdict

The defendant was charged with three different counts involving Justin. The majority of online jurors convicted the defendant on each of the counts, but the percentage of convictions for each varied. In the separate trial, more online jurors found the defendant guilty for the non-penetrative offences (Count 1: 71.1 per cent; Count 2: 67.9 per cent) than for the penetrative offence (58.5 per cent), suggesting that online jurors were more reluctant to convict the defendant for the more serious offence. In the joint trial, by contrast, the conviction rate for the non-penetrative offences was slightly lower, with 58 per cent conviction rate for the non-penetrative offences and 60 per cent conviction rate for the penetrative offence. The differences in the conviction rates between the separate and the joint trials were, however, not statistically significant.

⁶¹⁸ $U = 1074.0, z = -1.69, p = .092.$

⁶¹⁹ Masturbate defendant: *CSA Knowledge*: $beta = .013, p = .851$; trial type: $beta = .055, p = .421$; convincingness of the complainant: $beta = .752, p < .001.$

Masturbate complainant: *CSA Knowledge*: $beta = .044, p = .514$; trial type: $beta = .055, p = .425$; convincingness of the complainant: $beta = .750, p < .001.$

Oral-penile penetration: *CSA Knowledge*: $beta = .011, p = .883$; trial type: $beta = .063, p = .413$; convincingness of the complainant: $beta = .671, p < .001$

⁶²⁰ Masturbate defendant: $F(3, 102) = 40.905, p < .001$; Masturbate complainant: $F(3, 102) = 40.344, p < .001$; Oral-penile penetration: $F(3, 102) = 25.845, p < .001.$

In line with the above-mentioned results, the perceived convincingness of the complainant was the sole variable that predicted verdict on each of the counts.⁶²¹ Specifically, online jurors who perceived the complainant as more convincing were twice as likely to convict the defendant on each of the counts compared to online jurors who perceived the complainant as less convincing. The effect was slightly larger for the non-penetrative than the penetrative offences.

Each of the models that included online jurors' *CSA Knowledge*, type of trial and the convincingness of the complainant as predictors of verdict was statistically significant, correctly classifying about 70 per cent of the verdicts (Count 1: 71.8 per cent; Count 2: 69.9 per cent; Count 3: 73.8 per cent), and explaining between 22.0 per cent and 30.4 per cent of the variance in verdict on Count 1, between 26.3 per cent and 35.9 per cent of the variance in verdict on Count 2, and between 20.8 per cent and 28.1 per cent of variance in verdict on Count 3.⁶²²

Weak v. medium v. strong evidence in a joint v. separate trial

The results in this section were based on analyses conducted with separate trials of different evidence strength (weak: $n = 50$; moderately strong: $n = 46$; strong: $n = 52$) and with the joint trial in which evidence of all three complainants was presented ($n = 50$). Where possible, all four experimental conditions are compared (e.g., criminal intent of the defendant) as these analyses are not dependent on the complainants. When results are based on complainants (i.e., convincingness of the complainant, fairness to the complainant, blame of the complainant) or complaints (i.e., factual culpability, verdict), the results are presented as (1) between-subjects analyses for separate trials of different evidence strength, and (2) as within-subjects analyses for single complainants within the joint trial.

⁶²¹ Masturbate defendant: *CSA Knowledge*: Wald test (1, 103) = 0.028, $p = .867$, Odds Ratio = 1.005, 95% CI [0.943; 1.071]; Type of trial: Wald test (1, 103) = 0.616, $p = .433$, Odds Ratio = 1.451, 95% CI [0.573; 3.676]; Convincingness of the complainant: Wald test (1, 103) = 17.467, $p < .001$, Odds Ratio = 2.014, 95% CI [1.450; 2.796].

Masturbate complainant: *CSA Knowledge*: Wald test (1, 103) = 0.002, $p = .963$, Odds Ratio = 1.002, 95% CI [0.939; 1.068]; Type of trial: Wald test (1, 103) = 0.054, $p = .816$, Odds Ratio = 1.119, 95% CI [0.433; 2.891]; Convincingness of the complainant: Wald test (1, 103) = 20.412, $p < .001$, Odds Ratio = 2.277, 95% CI [1.593; 3.254].

Oral-penile penetration: *CSA Knowledge*: Wald test (1, 103) = 1.213, $p = .271$, Odds Ratio = 0.968, 95% CI [0.914; 1.026]; Type of trial: Wald test (1, 103) = 0.992, $p = .319$, Odds Ratio = 0.629, 95% CI [0.253; 1.566]; Convincingness of the complainant: Wald test (1, 103) = 17.038, $p < .001$, Odds Ratio = 1.946, 95% CI [1.419; 2.669].

⁶²² Masturbate defendant: $\chi^2(3, 103) = 25.640$, $p < .001$; masturbate complainant: $\chi^2(3, 103) = 31.373$, $p < .001$; oral-penile penetration: $\chi^2(3, 103) = 24.070$, $p < .001$.

Criminal intent of the defendant

A one-way between-subjects ANOVA was conducted to assess the influence of evidence strength (separate trial with weak evidence vs. moderately strong evidence vs. strong evidence vs. joint trial) on the perceived criminal intent of the defendant. The results revealed a significant effect of evidence strength on the perceived criminal intent of the defendant.⁶²³ Post-hoc analyses confirmed that online jurors were less likely to perceive criminal intent in the defendant's behaviour in the separate trial with the weak evidence ($M = 3.70$) than in the separate trial with strong evidence ($M = 5.09$) and in the joint trial ($M = 4.83$).⁶²⁴ No other comparisons were significant. Inspection of the mean scores of each experimental group in more detail revealed a trend for ratings of the criminal intent of the defendant to increase in line with the evidence strength in the separate trials, as shown in Figure 20. Interestingly, online jurors' ratings of the criminal intent of the defendant in the joint trial ($M = 4.83$) fell between those in the separate trial with moderately strong evidence ($M = 4.44$) and those in separate trial with strong evidence ($M = 5.09$).⁶²⁵

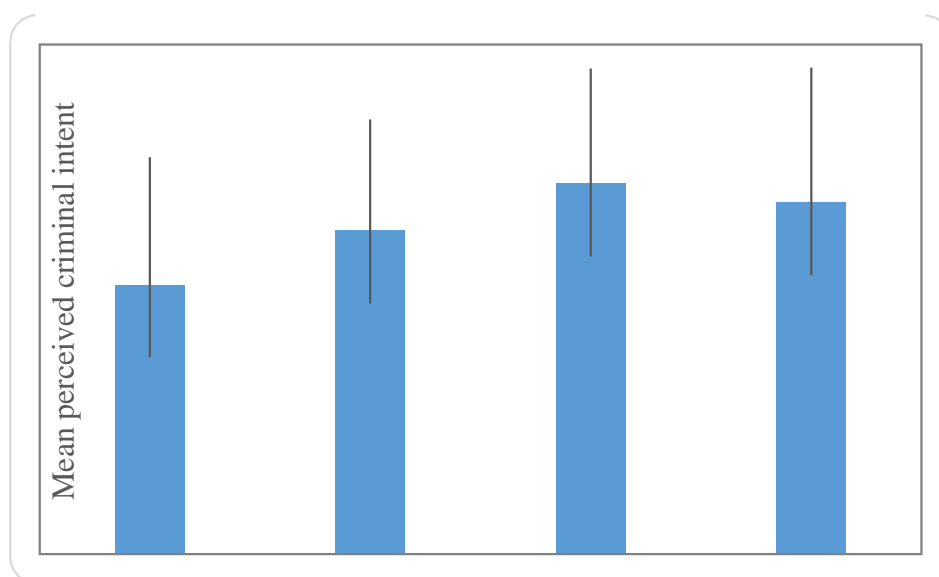


Figure 20. Mean perceived criminal intent of the defendant by evidence strength and type of trial.

⁶²³ $F(3, 199) = 6.583, p < .001, \eta^2_{\text{partial}} = .092$.

⁶²⁴ Separate weak v. separate strong: $M_{\text{diff}} = 1.394, \text{Std Err} = .332, p < .001$; Separate weak v. joint trial: $M_{\text{diff}} = 1.125, \text{Std Err} = .337, p = .006$.

⁶²⁵ Separate weak: $\text{Std Dev} = 1.75$; separate moderately strong: $\text{Std Dev} = 1.53$; separate strong: $\text{Std Dev} = 1.58$; joint trial: $\text{Std Dev} = 1.85$.

Perceptions of the defendant's sexual interest in boys

Online jurors were asked to which extent the complainants' testimony proved beyond reasonable doubt that the defendant had a sexual interest in (a) boys and (b) the particular complainant.

Analyses of the perception of sexual interest in boys without referring to the particular complainant revealed a significant effect of evidence strength and trial type.⁶²⁶ Post-hoc analyses⁶²⁷ revealed that online jurors were less likely to agree that the defendant had sexual interest in boys in the weak separate trial than in the moderately strong and strong separate trials and the joint trial.⁶²⁸ All other comparisons were not significant.

In contrast to the results reported above in section 3.2.4.1, the analyses presented here focused on the perceived sexual interest by the defendant in these particular complainants and not a general sexual interest in boys.

Figure 21 presents the perceived sexual interest in each complainant by evidence strength and trial type. First, perceived sexual interest in each of the complainants was compared between the separate trials of different evidence strength, next perceived sexual interest in each of the complainants within the joint trial was compared.

⁶²⁶ $\chi^2(3, 199) = 33.795, p < .001$.

⁶²⁷ using the Bonferroni-adjusted $\alpha = .008$

⁶²⁸ Weak v. moderately strong: $U = 722.5, z = -3.190, p = .001$; weak v. moderately strong: $U = 658.0, z = -4.470, p < .001$; moderately strong v. strong: $U = 1019.0, z = -1.423, p = .155$; weak separate v. joint trial: $U = 501.5; z = -5.225, p < .001$; moderately strong separate v. joint trial: $U = 804.0, z = -2.572, p = .010$; strong separate v. joint trial: $U = 1155.0, z = -1.140, p = .254$.

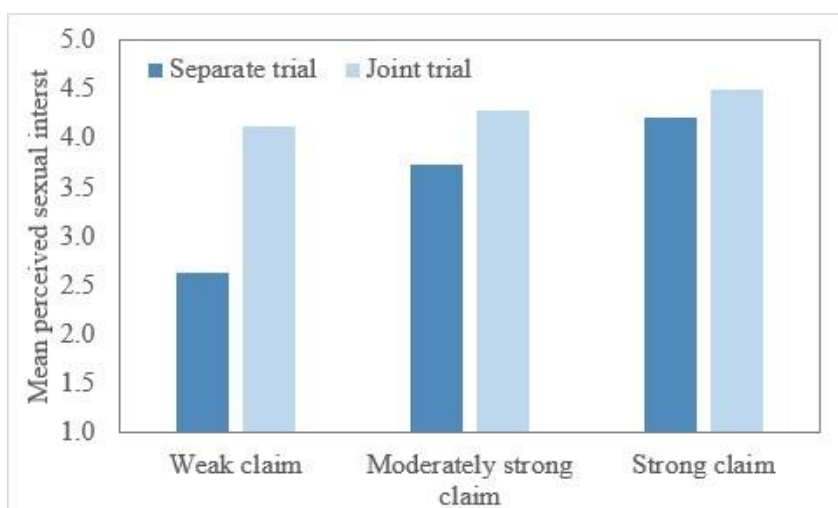


Figure 21. Mean perceived sexual interest in complainants with different evidence strength.

When comparing the separate trials of different evidence strength, the Kruskal-Wallis-H test revealed a significant effect of the evidence strength on perceived sexual interest in the boys.⁶²⁹ Post-hoc analyses revealed that online jurors were more likely to agree that the defendant had sexual interest in the complainant with the moderately strong and strong claim than the weak claim.⁶³⁰ There were no differences between the complainants with moderately strong and strong claims. See Figure 15.

Within-group analyses in the joint trial revealed significant difference in perceived sexual interest between the complainants.⁶³¹ Post-hoc analyses revealed only a significant difference between the weak and the strong claim.⁶³² Online jurors were more likely to agree that the defendant had sexual interest in the complainant with the strong claim than in the complainant with the weak claim. The perceived sexual interest of the defendant in boys in the case of the complainant with the moderately strong claim did not differ significantly from that in the case of the complainant with weak or strong evidence after taking into account the Bonferroni-adjusted $\alpha = .017$.

In sum, when the complainants testified in a separate trial without any additional evidence from other complainants, online jurors perceived more of a sexual interest in the complainant by the defendant as the strength of the evidence by the particular complainant against the defendant

⁶²⁹ $\chi^2(2, 149) = 19.199, p < .001$.

⁶³⁰ Weak v. moderately strong: $U = 721.5, z = -3.208, p = .001$; weak v. strong: $U = 710.0, z = -4.119, p < .001$; moderately strong v. strong: $U = 1055.5, z = -1.164, p = .245$.

⁶³¹ $\chi^2(2, 50) = 9.869, p = .007$.

⁶³² Weak v. moderately strong: $Z = -1.137, p = .256$, weak v. strong: $Z = -3.014, p = .003$; moderately strong v. strong: $Z = -2.326, p = .020$.

increased. This effect persisted for the complainant with weak versus the strong claims only when all three complainants testified in the same trial. The perceived sexual interest in the complainants with moderately strong and strong claims did not differ from each other in the joint trial.

Victim blame for abuse alleged by complainants

Analyses were conducted to test the impact of the evidence strength on blame of the complainants in the separate trials with different evidence strength. One-way between-subjects ANOVA revealed a significant effect of evidence strength on victim blame.⁶³³ Post-hoc analyses indicated that online jurors were more likely to blame the complainants with weak evidence and moderately strong evidence than the complainant with strong evidence.⁶³⁴ No differences were found between the blame allocated to the complainant in separate trials with weak and moderately strong claims.

Online jurors who were exposed to the evidence-in-chief and cross-examination of all three complainants in a joint trial blamed all three complainants to the same extent, i.e., the results of the within-subjects ANOVA were not significant, indicating no differences in perceived blameworthiness of these complainants.⁶³⁵ Figure 19 above illustrates the differences in victim blame between and within the experimental groups.

How convincing were the complainants with different evidence strength perceived to be?

Between-subjects analysis testing the impact of evidence strength in separate trials on online jurors' perceptions of how convincing the complainant was, yielded significant differences.⁶³⁶ Post-hoc analyses⁶³⁷ revealed significant differences between the weak and the strong case and the moderately strong and strong case.⁶³⁸ Specifically, online jurors in the separate trial with strong evidence perceived the complainant as more convincing than the complainants in the separate trials with weak and moderately strong evidence. The difference between the

⁶³³ $F(2, 149) = 8.540, p < .001, \eta^2_{\text{partial}} = .105.$

⁶³⁴ Weak v. moderately strong: $Mdiff = 0.212, SE = 0.270, p = .712$; weak v. strong: $Mdiff = 1.018, SE = 0.260, p < .001$; moderately strong v. strong: $Mdiff = 0.806, SE = 0.270, p = .008.$

⁶³⁵ Wilk's Lambda = .961, $F(2, 48) = 0.979, p = .383.$

⁶³⁶ $\chi^2(2, 149) = 19.165, p < .001.$

⁶³⁷ using the Bonferroni adjusted $\alpha = .017$

⁶³⁸ Weak v. moderately strong: $U = 896.0, z = -1.892, p = .059$; weak v. strong: $U = 689.0, z = -4.252, p < .001$; moderately strong v. strong: $U = 859.5, z = -2.561, p = .010.$

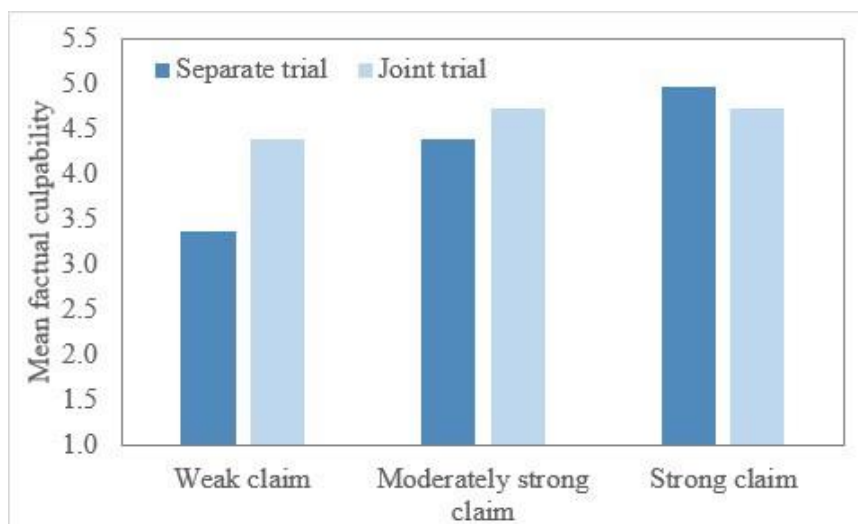
perceived convincingness of the three complainants with evidence of different strength when presented in a joint trial in which all three complainants testified, was not significant.⁶³⁹

Perceived fairness to the complainants

Between- and within-subjects analyses revealed that the extent to which the trial was perceived as fair to the complainants was not significant.⁶⁴⁰ Accordingly, independently of the evidence strength, the trial was perceived as equivalently fair to the complainant with weak, the moderately strong and strong evidence. This effect was not significant either when separate trials were compared with each other, or when the fairness of the trial to the three individual complainants appearing within the joint trial was compared.

Factual culpability of the defendant

Assessment of the perceived factual culpability of the defendant by evidence strength was analysed by comparing the same type of offence (masturbation of the complainant) between the complainants with different evidence strength. Figure 22 presents the perceived factual culpability by evidence strength of each complainant for the separate trials and the joint trial.



⁶³⁹ $\chi^2(2, 50) = 2.835, p = .242$.

⁶⁴⁰ Between-subjects analyses (separate trials): $\chi^2(2, 149) = 4.671, p = .097$; within-subjects analyses (joint trial): $\chi^2(2, 50) = 1.333, p = .513$.

Figure 22. Perceived factual culpability of the defendant for cases of different evidence strength.

Between-subjects analyses revealed a significant effect of evidence strength on the perceived factual culpability of the defendant in the separate trials.⁶⁴¹ Post-hoc analyses revealed significant differences between the case with the weak evidence and those with moderately strong and strong evidence.⁶⁴² Online jurors who were exposed to the separate trial with weak evidence were significantly less likely to perceive the defendant as factually culpable for the alleged nonpenetrative offence than were online jurors who were exposed to the separate trial with moderately strong and strong evidence. No differences were found between the perceived factual culpability of the defendant for the cases with moderately strong and strong evidence.

Hierarchical multilevel regression analysis was conducted to test the impact of online juror *CSA Knowledge*, evidence strength and the extent to which the complainant was perceived to be convincing on the factual culpability of the defendant. Online juror *CSA Knowledge* did not predict perceived factual culpability, explaining less than one per cent of the variance in factual culpability.⁶⁴³ Evidence strength was a significant predictor, explaining 15.3 per cent of the variance in perceived factual culpability of the defendant.⁶⁴⁴ The extent to which the complainant was perceived to be convincing was the strongest predictor of perceived factual culpability, explaining an additional 41.4 per cent of the variance in the defendant's factual culpability.⁶⁴⁵ These findings demonstrated that the stronger the evidence against the complainant, and the more convincing the complainant was perceived to be, the more likely online jurors were to find the defendant factually culpable of the nonpenetrative offence of masturbation of the complainant. The full model containing all three predictors explained 57.5 per cent of the variance in the perceived factual culpability of the defendant.

The perceived factual culpability of the defendant for the nonpenetrative offence was likewise dependent on evidence strength in the joint trial. Post-hoc analyses revealed a significant

⁶⁴¹ $\chi^2(2, 149) = 22.993, p < .001$.

⁶⁴² Weak v. moderately strong: $U = 748.5, z = -2.990, p = .003$; weak v. strong: $U = 632.5, z = -4.634, p < .001$; moderately strong v. strong: $U = 955.5, z = -1.881, p = .060$.

⁶⁴³ *CSA Knowledge*: $\beta = -.086, t = -1.043, p = .299, R^2 = .007$. . $F(1, 148) = 1.087, p = .299$.

⁶⁴⁴ *CSA Knowledge*: $\beta = -.021, t = -0.278, p = .781$; evidence strength: $\beta = .397, t = 5.162, p < .001$; $R^2 = .161, R^2_{change} = .153$. . $F(2, 148) = 13.964, p < .001$.

⁶⁴⁵ *CSA Knowledge*: $\beta = .039, t = 0.707, p = .480$; evidence strength: $\beta = .163, t = 2.788, p = .006$; convincingness of the complainant: $\beta = .691, t = 11.890, p < .001$; $R^2 = .575, R^2_{change} = .414$. .

$F(3, 148) = 65.387, p < .001$.

difference between factual culpability ratings in the case with the weak evidence and the moderately strong evidence, and a significant difference between the case with the weak evidence and that with the strong evidence.⁶⁴⁶ Similarly to the between-group analyses, the perceived factual culpability of the defendant was lower for the case of the complainant with weak evidence than that of the complainants with stronger evidence. No regression analyses could be conducted for within-subjects results due to the dependency of the data.

Online juror reasoning about the claims and verdict consistency

Parallel to the analyses on the perceived factual culpability of the defendant, analyses on verdict were conducted for offences of the same type only, that is, masturbation of the complainant. Between-subjects analyses revealed a significant effect of evidence strength on verdict.⁶⁴⁷ Whereas one quarter of the online jurors convicted the defendant in the separate trial with weak evidence (24.5 per cent), the majority of the online jurors convicted the defendant in the separate trial with moderately strong (54.3% per cent) and with strong evidence (67.9 per cent). Moreover, the difference in conviction rates was most extreme between the trial with the weakest and the trial with the strongest evidence.

Logistic regression analysis was conducted to test the effect on verdicts in the separate trials of online juror *CSA Knowledge*, evidence strength, and perceived convincingness of the complainants. The full model was statistically significant,⁶⁴⁸ and explained between 38.6 per cent and 51.4 per cent of the variance, and correctly classified 80.5 per cent of the responses. Whereas online jurors' *CSA Knowledge* did not predict verdict, evidence strength and the perceived convincingness of the complainant did predict verdict.⁶⁴⁹ Online jurors were three times more likely to convict the defendant in the trial with moderately strong and strong evidence than in the trial with weak evidence. Parallel analyses revealed no difference in conviction rates between the trials with moderately strong and strong evidence.⁶⁵⁰ Further, online jurors who perceived the complainant as convincing were 2.5 times more likely to convict the defendant than jurors who did not.

⁶⁴⁶ Weak v. moderately strong: $Z = -2.359, p = .018$, weak v. strong: $Z = -2.654, p = .008$; moderately strong v. strong: $Z = 0.000, p = 1.000$.

⁶⁴⁷ $\chi^2(2, 149) = 20.627, p < .001, Phi = .372$.

⁶⁴⁸ $\chi^2(4, 149) = 72.581, p < .001$.

⁶⁴⁹ *CSA Knowledge*: Wald test (1, 149) = 2.388, $p = .122$, Odds Ratio = 0.957, 95% CI [0.904; 1.012]; Type of trial: Wald test (2, 149) = 5.787, $p = .055$; Type of trial (weak v. moderately strong): Wald test (1, 149) = 4.782, $p = .029$, Odds Ratio = 3.309, 95% CI [1.132; 9.674]; Type of trial (weak v. strong): Wald test (1, 149) = 4.154, $p = .042$, Odds Ratio = 3.047, 95% CI [1.044; 8.898]; Convincingness of the complainant: Wald test = 45.960, $p < .001$, Odds Ratio = 2.448, 95% CI [1.890; 3.172].

⁶⁵⁰ Wald test (1, 149) = 0.025, $p = .874$, Odds Ratio = 1.085, 95% CI [0.392; 3.007].

When considering verdict of the same offence type but different evidence strength within the joint trial, the conviction rate was again elevated for the complaint with the moderately strong (58.0 per cent) and strong claim (58.0 per cent) compared to the weak claim. However, the conviction rate was the same for the complaints with the moderately strong and strong claims, supporting the results of the perceived factual culpability of the defendant. Inspection of verdict consistency revealed that 84 per cent of online jurors returned the same verdict for the alleged offence in the weak and moderately strong cases, compared to 80 per cent of online jurors who returned the same verdicts for both the weak and strong claims. The consistency of verdicts was higher for the offences alleged by the complainants with the moderately strong and strong claims (96.0 per cent).

Self-reported cognitive effort expended by online jurors

After reaching their verdicts, each online juror was asked to rate the extent to which they expended mental effort to complete different aspects of the online juror tasks and decision making. Specifically, they rated the mental effort required to: understand the charges against the defendant, remember the case facts, assess the credibility of the witnesses, evaluate the prosecution and the defence case, understand the judge's instructions, apply the law to the facts and reach a verdict. Further, online jurors were asked how confident they were about their verdicts.

The analyses presented below compare reported cognitive effort by type of trial, followed by evidence strength each of the complainants.

Reported cognitive effort by trial type

Results of analyses of the reported cognitive effort by type of trial are presented in Figure 23. Inspection of the overall mean for reported cognitive effort, independently of the specific questions (Separate trial: $M = 4.82$; Relationship evidence: $M = 4.70$; Tendency evidence: $M = 5.01$; Joint trial: $M = 5.16$)⁶⁵¹ indicated that the least cognitive effort was demanded of online jurors in the relationship evidence trial, and the most cognitive effort was demanded in the tendency evidence and the joint trials. A one-way between-subjects ANOVA testing the influence of trial type on the overall reported mean cognitive effort showed no differences between the experimental groups.⁶⁵² Accordingly, the type of trial did not affect the amount

⁶⁵¹ Separate trial: $Std Dev = 1.22$; Relationship evidence: $Std Dev = 1.28$; Tendency evidence: $Std Dev = 1.30$; Joint trial: $Std Dev = 1.16$.

⁶⁵² $F(3, 191) = 1.345, p = .261$.

of cognitive effort experienced by online jurors. This effect remained constant and was not significant when each of the questions about different types of cognitive effort were considered separately.⁶⁵³ Similarly, online jurors did not differ in their reported confidence in relation to their verdicts by type of trial.⁶⁵⁴

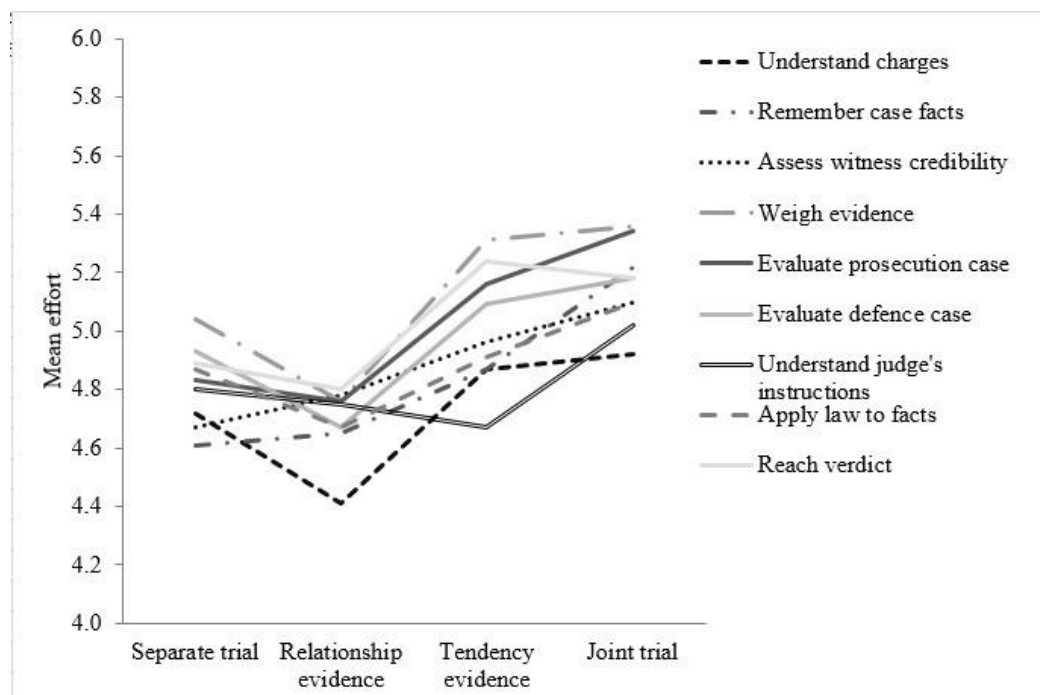


Figure 23. Reported cognitive effort by type of trial and juror task.

Reported cognitive effort by case strength

The overall ratings of reported cognitive effort, independent of the type of juror task, were moderate to high ($M = 4.93$)⁶⁵⁵ and varied by type of trial.⁶⁵⁶ Online jurors reported that the most cognitive effort was expended in the case with the weak evidence ($M = 5.20$), somewhat less in the case with the moderately strong evidence ($M = 4.82$) and the least in the case with the strong evidence ($M = 4.55$).⁶⁵⁷ Further analyses revealed that the difference in cognitive

⁶⁵³ $p > .100$.

⁶⁵⁴ $\chi^2(3, 192) = 4.95, p = .176$.

⁶⁵⁵ Possible score range: 1-7. $Std Dev = 1.29$.

⁶⁵⁶ $F(3, 198) = 2.963, p = .033, \eta_{\text{partial}}^2 = .044$.

⁶⁵⁷ Weak evidence: $Std Dev = 1.27$, moderately strong evidence: $Std Dev = 1.22$, strong evidence $Std Dev = 1.41$.

effort expended in response to the weak and the strong cases was statistically significant, showing a moderate effect size.⁶⁵⁸

The extent to which the differences were significant or not, varied by task type. The reported cognitive effort on each of the measures is presented in Figure 24. Kruskal-Wallis tests revealed significant between cases of different strength in the reported cognitive effort expended to (a) understand the charges against the defendant;⁶⁵⁹ (b) weigh the evidence;⁶⁶⁰ (c) evaluate the prosecution case;⁶⁶¹ and (d) apply the law to the facts.⁶⁶² Post-hoc analyses revealed significantly higher reports of cognitive effort were required to apply the law to the facts in the weak case compared to the cases with moderately strong and strong evidence.⁶⁶³ Further, online jurors who read the weak evidence case reported expending significantly more cognitive effort than their counterparts who read the strong evidence case (but not the moderately strong evidence case) in response to the remaining three questions mentioned above, i.e., (a) understanding the charges against the defendant;⁶⁶⁴ (b) weighing the evidence;⁶⁶⁵ and (c) evaluating the prosecution case.⁶⁶⁶

⁶⁵⁸ $M_{diff} = 0.646$, $SE = .251$, $p = .052$, $d = 0.48$, 95% CI [0.089; 0.873].

⁶⁵⁹ $\chi^2(2, 199) = 5.78$, $p = .056$.

⁶⁶⁰ $\chi^2(2, 199) = 7.28$, $p = .026$.

⁶⁶¹ $\chi^2(2, 199) = 5.94$, $p = .051$.

⁶⁶² $\chi^2(2, 199) = 8.20$, $p = .017$.

⁶⁶³ Weak v. moderately strong: $U = 887.5$, $z = -2.00$, $p = .045$; weak v. strong: $U = 932.0$, $z = -2.65$, $p = .008$; moderately strong v. strong: $U = 1067.0$, $z = -1.09$, $p = .274$.

⁶⁶⁴ $\chi^2(2, 199) = 5.78$, $p = .056$.

⁶⁶⁵ $\chi^2(2, 199) = 7.28$, $p = .026$.

⁶⁶⁶ Understand charges against defendant: Weak v. moderately strong: $U = 1022.0$, $z = -0.97$, $p = .333$; weak v. strong: $U = 997.5$, $z = -2.20$, $p = .028$; moderately strong v. strong: $U = 986.5$, $z = -1.67$, $p = .096$.

Weigh the evidence: Weak v. moderately strong: $U = 953.5$, $z = -1.49$, $p = .136$; weak v. strong: $U = 941.5$, $z = -2.59$, $p = .010$; moderately strong v. strong: $U = 1034.0$, $z = -1.33$, $p = .182$.

Evaluate the prosecution case: Weak v. moderately strong: $U = 911.5$, $z = -1.80$, $p = .072$; weak v. strong: $U = 989.0$, $z = -2.27$, $p = .023$; moderately strong v. strong: $U = 1125.0$, $z = -0.68$, $p = .499$.

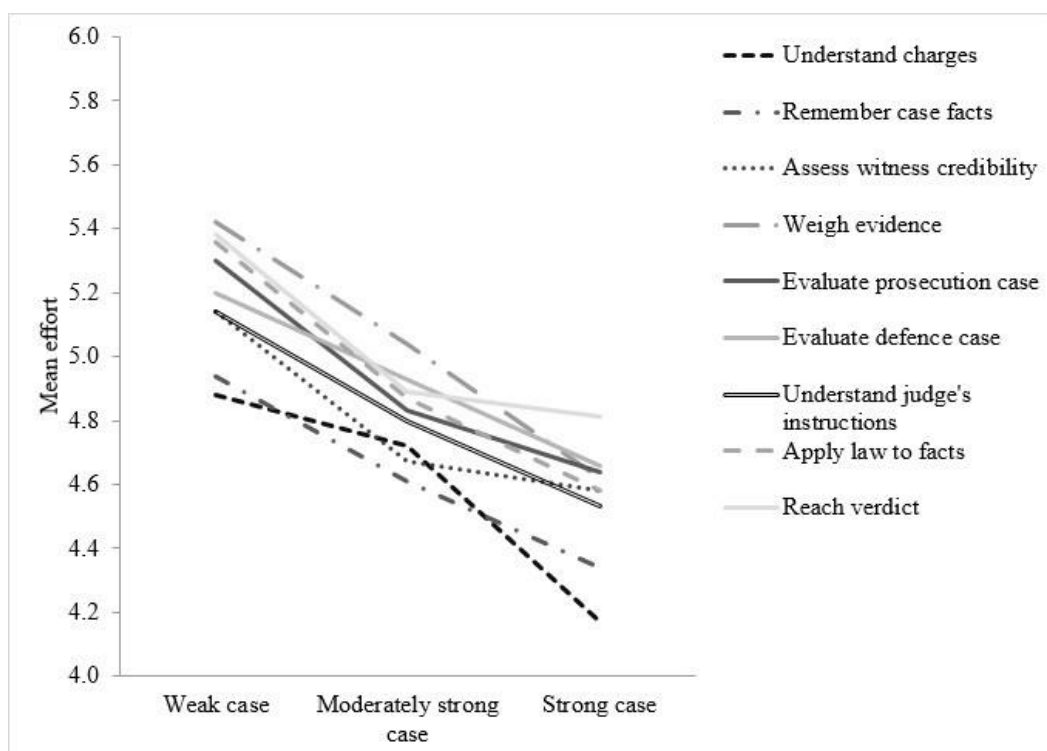


Figure 24. Reported cognitive effort by evidence strength and juror task.

The influence of online jurors' emotional responses on juror reasoning

Online jurors' emotional responses were assessed using two scales contained within the PANAS, namely, *Positive Affect* and *Negative Affect*. Results in Section 3.3.1 indicated a significant correlation between online jurors' pre-trial *CSA Knowledge* and their emotional responses post-trial, in that online jurors reported less emotional arousal when they knew more about CSA. In this section, we considered the extent to which (1) trial type and evidence strength had an effect on online jurors' emotional responses and (2) these emotional responses impacted decisions on the factual culpability of the defendant and verdict.

The influence of the type of trial on emotional responses

Multivariate ANOVA testing the impact of trial type (separate trial with moderately strong evidence, relationship evidence, tendency evidence and joint trial) on reported emotional arousal was not significant.⁶⁶⁷ Online jurors' emotional responses post-trial were independent of the type of trial. Similarly, multivariate ANOVA testing the impact of evidence strength (weak v. moderately strong v. strong) on reported emotional arousal was not significant.⁶⁶⁸

⁶⁶⁷ Wilk's Lambda = .990, $F(6, 374) = 0.323$, $p = .925$, $\eta^2_{\text{partial}} = .005$.

⁶⁶⁸ Wilk's Lambda = .989, $F(6, 388) = 0.364$, $p = .902$, $\eta^2_{\text{partial}} = .006$.

Online jurors' emotional responses post-trial were similar, and independent of the strength of the evidence presented in the trials. Overall, online jurors' emotional responses as measured by the *Positive* and *Negative Affect* Scales did not differ in response to trials with evidence of a different strength and or type of trial, even when all three complainants testified against the defendant in the same trial.

The influence of emotional responses on online jurors' decision making

To test the impact of online jurors' emotional responses on their decisions, a series of comparisons was conducted. First, comparisons were made of the trials of the complainant with moderately strong evidence, whether presented in the context of a separate trial, or a trial with relationship evidence, tendency evidence, or a joint trial. Next, analyses were conducted for each complainant separately to assess online jurors' responses to complaints of different evidential strength.

Factual culpability of the defendant

Multiple regression analyses were conducted to test the effect of *Positive* and *Negative Affect* on the factual culpability of the defendant for the penetrative and the nonpenetrative offence. Results revealed that *Positive Affect* but not *Negative Affect* predicted perceived factual culpability for the nonpenetrative offence. The model was statistically significant and predicted 3.9 per cent of the variance in factual culpability ratings.⁶⁶⁹ In contrast, neither *Positive* nor *Negative Affect* predicted factual culpability of the defendant for the penetrative offence.⁶⁷⁰ Overall, online jurors' reported levels of emotional arousal at the end of the trial had little influence on the perceived factual culpability of the defendant.

Verdict

Logistic regression analysis revealed that online jurors' *Negative Affect* but not their *Positive Affect* significantly predicted verdict for the nonpenetrative and penetrative offence, showing a small but significant effect for both counts.⁶⁷¹ The model correctly classified 60.4 per cent of the cases for the nonpenetrative offence and 63.5 per cent of the cases for the penetrative

⁶⁶⁹ Positive Affect: $\beta = .170, p = .027$; Negative Affect: $\beta = .056, p = .464$; $F(2, 191) = 3.827, p = .023$.

⁶⁷⁰ Positive Affect: $\beta = .126, p = .101$; Negative Affect: $\beta = .105, p = .173$; $F(2, 191) = 3.570, p = .030$.

⁶⁷¹ Nonpenetrative offence: Positive Affect: Wald test (1, 192) = 0.646, $p = .422$, Odds Ratio = 1.015, 95% CI [0.979; 1.053]; Wald test (1, 192) = 6.630, $p = .012$, Odds Ratio = 1.053, 95% CI [1.012; 1.096].

Penetrative offence: positive Affect: Wald test (1, 192) = 1.204, $p = .273$, Odds Ratio = 1.021, 95% CI [0.984; 1.043]; Wald test (1, 192) = 15.135, $p < .001$, Odds Ratio = 1.053, 95% CI [1.045; 1.143].

offence, and explained 5.1 to 6.9 per cent and 12.0 to 16.2 per cent of the variance in verdict for each offence type, respectively.⁶⁷² These results differed from those for decision making with respect to the perceived factual culpability of the defendant, showing that *Negative Affect*, or negative emotional arousal in response to the, nature of the charges had a small effect on decision making with respect to verdict.

Fairness to the defendant

To assess the online jurors' ratings of the fairness of the trial to the defendant, comparative analyses were conducted with all online jurors, i.e., in all six experimental groups, wherever feasible.

Fairness of the trial to the defendant

Online jurors were asked how fair the trial was to the defendant. Responses were reported on a 7-point rating scale, with higher scores indicating higher perceived fairness to the defendant. On average, the trial was rated as more fair than unfair ($M = 5.03$),⁶⁷³ with average scores ranging between $M = 4.78$ and $M = 5.38$.⁶⁷⁴ Nonparametric analyses revealed no differences between the different trials in the perceived fairness of the proceedings to the defendant.⁶⁷⁵ Independently of the variations in the type of trial and strength of evidence, the different trials were perceived as equally fair to the defendant.

The perceived convincingness of the defendant

The extent to which the defendant was perceived as convincing varied according to the strength of evidence presented, but not the type of trial.⁶⁷⁶ Specifically, whereas there were no differences in the perceived convincingness of the defendant in the separate trials with moderately strong evidence ($M = 4.07$), relationship evidence ($M = 4.47$), tendency evidence ($M = 4.33$) and the joint trial ($M = 4.80$), the defendant was perceived as more convincing in the separate trial with weak evidence ($M = 5.02$) than in the separate trials with moderately strong ($M = 4.07$)⁶⁷⁷ and strong evidence ($M = 4.09$).⁶⁷⁸

⁶⁷² Nonpenetrative offence: $\chi^2(2, 192) = 10.05, p = .007$; penetrative offence: $\chi^2(2, 192) = 24.60, p < .001$.

⁶⁷³ $Std Dev = 1.43$.

⁶⁷⁴ $Std Dev = 1.28$ and $Std Dev = 1.51$, respectively.

⁶⁷⁵ $\chi^2(3, 295) = 8.44, p = .133$.

⁶⁷⁶ Evidence strength: $\chi^2(2, 199) = 16.66, p < .001$; Type of trial: $\chi^2(2, 199) = 6.36, p = .095$.

⁶⁷⁷ $U = 637.5, z = -3.89, p < .001$.

⁶⁷⁸ $U = 847.0, z = -3.22, p = .001$.

In sum, the defendant’s evidence was perceived as most convincing in the separate trial with weak evidence and as least convincing in the other two separate trials with stronger evidence against the defendant. Interestingly, online jurors’ perceptions of how convincing the defendant was in the trials with relationship evidence, tendency evidence and in joint trials were slightly more favourable than in the separate trial with moderately strong evidence against the defendant.

Results of the Manipulation Check

The specific aims of the online juror study were accomplished. With respect to verdict, as the evidence increased in strength, the conviction rate increased concomitantly. The main manipulation check was based on a comparison of conviction rates for the charge of indecency in the form of masturbation of the complainant by the defendant, as this charge was common to all three complainants. Significantly different conviction rates for this offence were obtained in the predicted direction in the separate trials of the complainant with the weak evidence (23.5%), the moderately strong evidence (53.2%) and the strong evidence (67.9%).

In the separate trial of the complainant with the moderately strong evidence, when additional evidence of uncharged sexually abusive acts against the complainant was admitted in the form of relationship evidence about grooming behaviours, the conviction rate for the charged offences increased by approximately 15 per cent. An increase in the conviction rate of approximately 7 per cent was observed when evidence of uncharged acts by the defendant against two other boys was admitted in the separate trial of the complaint with the moderately strong evidence (tendency evidence).

Evidence type and charge		Mean conviction rate by type of trial			
Evidence strength	Offence type	Separate trial	Joint trial	Separate relationship evidence	Separate tendency evidence
Weak	Masturbate complainant	23.5	51.0		
Moderate	Masturbate complainant	53.2	58.8	66.7	59.6
	Digital anal penetration	51.1	56.9	68.6	59.6
Strong	Masturbate complainant	67.9	58.8		

Weak evidence: *Std Dev* = 1.12; moderately strong evidence: *Std Dev* = 1.18; strong evidence: *Std Dev* = 1.72; relationship evidence: *Std Dev* = 1.49, tendency evidence: *Std Dev* = 1.61; the joint trial: *Std Dev* = 1.49.

	Masturbate defendant	71.7	58.8
	Penile-oral penetration	58.5	60.8

Overall, the manipulation check confirmed that:

1. Mock jurors differentiated between the strength of evidence for the same offence in the individual cases of complainants with weak, moderately strong and strong evidence, respectively, in the manner intended. In the separate trials, the conviction rate for indecency in the form of masturbation by the defendant of the complainant with the weak evidence was 23.5 per cent, 53.2 per cent for the complainant with the moderately strong evidence, and 67.9 per cent for the complainant with the strong evidence. Thus, this manipulation of the evidence was successful and suitable for replication and extension in a more elaborate simulated video trial.
2. Comparisons of the reasoning of individual mock jurors in response to a joint trial (involving three complainants with different evidence strength: weak, moderately strong and strong) versus a separate trial involving one complainant with moderately strong evidence showed some similarities and some differences, indicating that this design would be effective to examine reasoning and decision-making processes in a larger scale study with juries rather than jurors as the unit of analysis, and with jury deliberations as a further source of information about their reasoning.
3. The study instruments, the pre-trial and post-trial questionnaires, were sensitive to differences in juror responses to the type of trial (separate vs. joint) without the influence of group deliberations, and provided useful measures to complement the deliberation analyses, i.e., these measures could be analysed at both the level of the jury and the level of the juror, using multi-level modelling.

Appendix F: Descriptive statistics for pre-trial and post-trial measures by experimental group in the online juror study

	Separate trial weak (<i>N</i> = 50)				Separate trial moderate (<i>N</i> = 46)				Separate trial strong (<i>N</i> = 51)				Relationship evidence (<i>N</i> = 51)			
	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
CSA-KQ Impact on Children	7.00	34.00	22.10	6.12	9.00	28.00	20.20	4.41	10.00	33.00	19.60	5.43	9.00	32.00	19.84	5.18
CSA-KQ Contextual Influences	6.00	25.00	16.98	4.47	9.00	25.00	17.09	3.57	8.00	28.00	16.32	3.71	4.00	25.00	16.10	4.70
CSA-KQ Total	18.00	58.00	39.08	9.22	21.00	50.00	37.28	6.59	18.00	60.00	35.92	8.03	14.00	54.00	35.94	7.84
FEEBS Pro-Prosecution	6.00	32.00	19.32	6.08	9.00	35.00	20.02	4.91	6.00	29.00	18.68	4.80	9.00	31.00	19.20	5.46
FEEBS Pro-Defence	5.00	33.00	16.02	5.79	10.00	26.00	16.80	4.71	5.00	29.00	17.34	5.24	8.00	30.00	17.78	4.97
PJAQ System Confidence	2.17	6.33	3.99	0.92	2.50	5.50	3.96	0.75	1.50	5.83	3.92	0.94	1.17	6.50	3.80	1.14
PJAQ Conviction Proneness	1.40	6.80	4.10	1.26	2.20	5.80	4.30	0.85	2.00	6.40	4.06	1.08	1.00	7.00	4.25	1.23
PJAQ Cynicism in the Defence	1.43	6.71	4.25	1.10	2.00	6.00	4.26	0.83	2.57	6.57	4.21	0.92	2.00	6.43	4.22	1.12
PJAQ Racial Bias	1.00	6.50	3.12	1.14	1.75	5.50	3.43	1.01	1.00	5.50	3.31	1.10	1.00	6.75	3.55	1.20
PJAQ Social Justice	1.50	6.25	3.97	1.09	1.75	7.00	3.85	1.07	1.25	6.00	4.17	0.95	1.50	6.50	3.97	1.11
PJAQ Innate Criminality	1.00	6.75	3.21	1.23	1.50	5.50	3.53	0.97	1.00	6.00	3.17	1.15	1.00	6.25	3.32	1.22
Case Facts	1.00	3.00	1.84	0.93	1.00	3.00	2.09	0.94	1.00	3.00	1.74	0.90	1.00	3.00	2.22	0.90
Positive Affect	10.00	45.00	25.18	8.37	10.00	38.00	24.22	6.92	10.00	48.00	24.53	8.40	10.00	40.00	25.31	8.63
Negative Affect	10.00	39.00	16.86	7.55	10.00	38.00	17.00	8.36	10.00	36.00	18.26	8.19	10.00	37.00	18.14	8.35
Cognitive Effort	2.00	7.00	5.20	1.27	1.00	7.00	4.82	1.22	1.00	7.00	4.55	1.41	1.00	7.00	4.69	1.28

APPENDIX F CONTINUED

	Tendency evidence (N = 45)				Joint trial (N = 50)			
	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
CSA-KQ Impact on Children	9.00	33.00	20.56	5.51	9.00	33.00	19.30	5.39
CSA-KQ Contextual Influences	4.00	27.00	16.69	5.36	6.00	28.00	16.58	4.45
CSA-KQ Total	22.00	58.00	37.24	8.61	15.00	54.00	35.88	7.41
FEEBS Pro-Prosecution	10.00	30.00	18.13	5.67	11.00	30.00	21.32	4.54
FEEBS Pro-Defence	5.00	30.00	15.80	5.14	8.00	30.00	18.48	4.75
PJAQ System Confidence	1.50	5.67	3.92	1.08	2.50	6.50	4.14	0.91
PJAQ Conviction Proneness	1.00	6.20	3.98	1.18	2.60	6.60	4.48	1.06
PJAQ Cynicism in the Defence	1.00	6.57	4.03	1.08	1.86	6.43	4.20	1.08
PJAQ Racial Bias	1.00	5.50	3.23	1.17	1.75	7.00	3.50	1.19
PJAQ Social Justice	1.00	6.25	4.07	1.34	2.75	6.75	4.20	0.92
PJAQ Innate Criminality	1.00	5.25	3.06	1.05	1.25	6.25	3.51	1.17
Case Facts	1.00	3.00	2.33	0.88	1.00	3.00	1.98	0.89
Positive Affect	10.00	47.00	25.80	9.60	10.00	46.00	25.54	8.92
Negative Affect	10.00	40.00	16.98	7.81	10.00	40.00	18.34	9.43
Cognitive Effort	2.00	7.00	5.01	1.30	2.56	7.00	5.16	1.16

Appendix G: Pre-trial juror questionnaire for the jury deliberation and reasoning study

Mock Jury Registration

Section 1 Demographic information

S1. Are you...

Male	1
Female	2

S2. Which of the following best describes you/ your current household?

Single with no children	1
Single parent	2
Young couple with no children	3
Young couple with children	4
Mid-life family (families with teenage children)	5
Mid-life household (family with independent children)	6
Mid-life couple with no children	7
Empty nester (parents who no longer live with their children)	8

S3. Professional or employment status

Employed for wages (full time or part time)	
Self-employed	
Out of work and looking for work	
Out of work and not currently looking for work	
A homemaker	
A student	
Military	
Retired	
Unable to work	

S4. What is the highest level of education you have completed?

Primary School	
High School (HSC)	
Some university (not completed)	
Trade/technical/vocational training	
Bachelor's Degree	
Master's Degree	

Doctoral Degree	
Professional Degree (MD, JD, etc)	
Other (please specify)	

Section 2 Pretrial Juror Section

Q1. Please rate whether you agree or disagree with each statement below about the **justice system**

	Strongly disagree 1	2	3	4	5	6	Strongly Agree 7
A suspect who runs from police, probably committed a crime							
A defendant should be found guilty if 11 out of 12 jurors vote guilty							
Too often jurors hesitate to convict someone who is guilty out of pure sympathy							
In most cases where the accused presents a strong defense, it is only because of a good lawyer							
Out of every 100 people brought to trial, at least 75 are guilty of the crime charged							
For serious crimes like murder, a defendant should be found guilty so long as there is a 90% chance that person committed the crime							
Defense lawyers don't really care about guilt or innocence, they are just in business to make money							
Generally, the police make an arrest only when they are sure who committed the crime							
Many accident claims filed against insurance companies are phony							
The defendant is often a victim of his or her own bad reputation							
Extenuating circumstances should not be considered; if a person commits a crime, that person should be							
If the defendant committed a victimless crime like gambling or possession of marijuana, that person should never be convicted							
Defense lawyers are too willing to defend individuals they know are guilty							
Police routinely lie to protect other police officers							
Once a criminal, always a criminal							
Lawyers will do whatever it takes, even lie, to win a							
Criminals should be caught and convicted by any means necessary							
A prior record of conviction is the best indicator of a person's guilt in the present case							
Wealthy individuals are almost never convicted of their crimes							
A defendant who is a member of a gang, is definitely guilty of the crime							
Ethnic minorities use the 'race issue' only when they are guilty							

When it is the suspect's word against the police officer's, I believe the police							
Men are more likely to be guilty of crimes than women							
The large number of Australian Aboriginals and Torres Strait Islanders currently in prison is an example of the innate criminality of that subgroup							
A black person on trial with a predominantly white jury will always be found guilty							
Ethnic minority suspects are likely to be guilty, more often than not							
If a witness refuses to take a lie detector test, that person is hiding something							
Defendants who change their story are almost always guilty							
Famous people are often considered to be above the law							

Section 3 Forensic Evidence Evaluation Bias Scale

Q2. Please read each of the statements below about **forensic evidence** and rate whether you agree or disagree with each statement

	Strongly disagree 1	2	3	4	5	6	Strongly Agree 7
Every crime can be solved with forensic science							
If no forensic evidence is recovered then the defendant is probably innocent							
Every criminal leaves trace evidence at every scene							
Forensic evidence alone is enough to convict							
No forensic evidence means investigators did not look hard enough							
Forensics always identifies the guilty person							
If there is no forensic evidence, the jury should not							
Forensics always provides a conclusive answer							
Police should not charge someone without forensic evidence							
Science is always the most reliable way to identify perpetrators							

Section 4 Child Sexual Assault Knowledge

Q3. Please read each of the statements below about **child sexual abuse in relation to court cases** and rate whether you agree or disagree with each statement

	strongly disagree 1	2	3	4	5	6	strongly agree 7
A medical examination almost always shows whether or not a child was sexually abused							

A child who has been sexually abused will tell someone soon afterwards							
Child victims of sexual abuse respond in a similar way to the abuse							
Children sometimes make false claims of sexual abuse to get back at an adult							
A child victim of sexual abuse will avoid his or her							
A sexually abused child typically cries out for help and tries to escape							
Repeating questions such as: "what happened? What else happened?" leads children to make false abuse							
Children are easily coached to make false claims of sexual abuse							
Children who are sexually abused display strong emotional reactions afterwards							
Children sometimes make up stories about having been sexually abused when they actually have not							
Children are sometimes led by an adult to report they have been sexually abused when they have not							
Children aged 7-8 years are easily manipulated to give false reports of sexual abuse							
Sexual assault is the worst offence that can be committed against a child							
Child sexual offenders deserve life imprisonment							

Thank you for taking the time to complete our survey. We look forward to your participation in the mock jury.

Appendix H: Verdict form for separate trial with two counts

VERDICT

Count 1: Act of indecency towards a person under the age of 16 years, namely Timothy Lyons, between 1 and 31 December 1997

We the jury, find Mark Booth

Guilty

Not guilty

Count 2: Sexual intercourse against a child above the age of 10 years and under the age of 14 years, namely Timothy Lyons, between 1 and 31 December 1997

We the jury, find Mark Booth

Guilty

Not guilty

Appendix I: Verdict form for joint trial with six counts

VERDICT

Count 1: Act of indecency towards a person under the age of 16 years, namely **Simon Rutter** between 1 March and 1 September 1993

We the jury, find Mark Booth Guilty
 Not guilty

Count 2: Act of indecency towards a person under the age of 16 years, namely **Justin McCutcheon**, between 1 September and 31 October 1995.

We the jury, find Mark Booth Guilty
 Not guilty

Count 3: Act of indecency towards a person under the age of 16 years, namely **Justin McCutcheon**, between 1 September and 31 October 1995.

We the jury, find Mark Booth Guilty
 Not guilty

Count 4: Sexual intercourse against a child above the age of 10 years and under the age of 14 years, namely **Justin McCutcheon**, between 1 September and 31 October 1995.

We the jury, find Mark Booth Guilty
 Not guilty

Count 5: Act of indecency towards a person under the age of 16 years, namely **Timothy Lyons**, between 1 and 31 December 1997

We the jury, find Mark Booth Guilty
 Not guilty

Count 6: Sexual intercourse against a child above the age of 10 years and under the age of 14 years, namely **Timothy Lyons**, between 1 and 31 December 1997

We the jury, find Mark Booth Guilty
 Not guilty

Appendix J: Question trail for separate trial with relationship evidence

QUESTIONS FOR THE JURY - TRIAL OF MARK BOOTH

Read in conjunction with the judge's directions.

Count 1 – act of indecency (masturbated Timothy Lyons' penis)

Indecent means contrary to the standards of ordinary, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A child under 16 years of age cannot consent to an act of indecency.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between December 1 and 31 in 1997? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth committed an act of indecency? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that the act was indecent? *If yes, go to question 4. If no, find the accused not guilty.*

Question 4. Did the Crown prove beyond reasonable doubt that at the relevant time, Timothy Lyons was under the age of 16 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 2 – sexual intercourse against a child above the age of 10 years and under the age of 14 years (inserted his finger into Timothy Lyons' anus)

Sexual intercourse is defined under our Crimes Act to include sexual connection occasioned by the penetration, to any extent, of the anus of any person by any part of the body of another person. A person under the age of 16 years cannot consent to sexual intercourse.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between December 1 and 31 in 1997? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth had sexual intercourse with Timothy Lyons? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that at the relevant time, Timothy Lyons was under the age of 14 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Appendix K: Question trail for joint trial

QUESTIONS FOR THE JURY - TRIAL OF MARK BOOTH

Read in conjunction with the judge's directions.

Count 1 – act of indecency (masturbated Simon Rutter's penis)

Indecent means contrary to the standards of ordinary, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A child under 16 years of age cannot consent to an act of indecency.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between 1 March and 1 September in 1993? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth committed an act of indecency? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that the act was indecent? *If yes, go to question 4. If no, find the accused not guilty.*

Question 4. Did the Crown prove beyond reasonable doubt that at the relevant time, Simon Rutter was under the age of 16 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 2 – act of indecency (forced Justin McCutcheon to touch and masturbate the accused's penis)

Indecent means contrary to the standards of ordinary, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A child under 16 years of age cannot consent to an act of indecency.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between 1 September and 31 October in 1995? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth committed an act of indecency? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that the act was indecent? *If yes, go to question 4. If no, find the accused not guilty.*

Question 4. Did the Crown prove beyond reasonable doubt that at the relevant time, Justin McCutcheon was under the age of 16 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 3 – act of indecency (masturbated Justin McCutcheon’s penis)

Indecent means contrary to the standards of ordinary, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A child under 16 years of age cannot consent to an act of indecency.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between 1 September and 31 October 1995? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth committed an act of indecency? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that the act was indecent? *If yes, go to question 4. If no, find the accused not guilty.*

Question 4. Did the Crown prove beyond reasonable doubt that at the relevant time, Justin McCutcheon was under the age of 16 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 4 – sexual intercourse against a child above the age of 10 years and under the age of 14 years (inserted his penis into Justin McCutcheon’s mouth)

Sexual intercourse is defined under our Crimes Act to include sexual connection occasioned by the by the introduction of any part of the penis of a person into the mouth of another person. A person under the age of 16 years cannot consent to sexual intercourse.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between 1 September and 31 October in 1995? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth had sexual intercourse with Justin McCutcheon? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that at the relevant time, Justin McCutcheon was under the age of 14 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 5 – act of indecency (masturbated Timothy Lyons’ penis)

Indecent means contrary to the standards of ordinary, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A child under 16 years of age cannot consent to an act of indecency.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between December 1 and 31 in 1997? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth committed an act of indecency? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that the act was indecent? *If yes, go to question 4. If no, find the accused not guilty.*

Question 4. Did the Crown prove beyond reasonable doubt that at the relevant time, Timothy Lyons was under the age of 16 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Count 6 – sexual intercourse against a child above the age of 10 years and under the age of 14 years (inserted his finger into Timothy Lyons’ anus)

Sexual intercourse is defined under our Crimes Act to include sexual connection occasioned by the penetration, to any extent, of the anus of any person by any part of the body of another person. A person under the age of 16 years cannot consent to sexual intercourse.

Question 1. Did the Crown prove beyond reasonable doubt that the events occurred between December 1 and 31 in 1997? *If yes, go to question 2. If no, find the accused not guilty.*

Question 2. Did the Crown prove beyond reasonable doubt that Mark Booth had sexual intercourse with Timothy Lyons? *If yes, go to question 3. If no, find the accused not guilty.*

Question 3. Did the Crown prove beyond reasonable doubt that at the relevant time, Timothy Lyons was under the age of 14 years? *If yes, find the accused guilty. If no, find the accused not guilty.*

Appendix L: Post-trial juror questionnaire in the jury deliberation and reasoning study



Jury Study Questionnaire

Are you satisfied beyond reasonable doubt that Mark Booth, the accused	
Count 1 – masturbated Simon Rutter 's penis between 1 March and 1 September 1993?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 2 – forced Justin McCutcheon to touch and masturbate the accused's penis between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 3 – masturbated Justin McCutcheon 's penis between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 4 – inserted his penis into Justin McCutcheon 's mouth between 1 September and 31 October 1995?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 5 – masturbated Timothy Lyons ' penis between 1 and 31 December 1997?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>
Count 6 – inserted his finger into Timothy Lyons 's anus between 1 and 31 December 1997?	<input type="checkbox"/> <i>Yes, guilty.</i> <input type="checkbox"/> <i>No, not guilty.</i>

Mark the number that best represents your answer.							
	Not at all						Very much
How confident are you about your verdicts?	1	2	3	4	5	6	7
Are you satisfied that the evidence proved beyond reasonable doubt that the accused had a sexual interest in boys?	1	2	3	4	5	6	7
Did Simon Rutter's testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Justin McCutcheon ?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Timothy Lyons ?	1	2	3	4	5	6	7
Did Justin McCutchen's testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Simon Rutter ?	1	2	3	4	5	6	7
influence your verdicts in relation to the charges involving Timothy Lyons ?	1	2	3	4	5	6	7
Did Timothy Lyons' testimony							
prove beyond reasonable doubt that Mark Booth had a sexual interest in him?	1	2	3	4	5	6	7
influence your verdicts in relation to the counts involving Simon Rutter ?	1	2	3	4	5	6	7
influence your verdicts in relation to the charges involving Justin McCutcheon?	1	2	3	4	5	6	7

The accused is presumed innocent unless the prosecution has proved his guilt beyond reasonable doubt.

What number between 0% and 100% represents "beyond reasonable doubt". %

Please fill in the blanks.

(3) What was the main reason for your verdict?

(4) What other factors went into your decision?

Mark the number that best represents your mental effort to							
	Very low mental effort			Very high mental effort			
understand the charges against Mark Booth.	1	2	3	4	5	6	7
remember the facts.	1	2	3	4	5	6	7
assess the credibility of the witnesses.	1	2	3	4	5	6	7
weigh the evidence.	1	2	3	4	5	6	7
evaluate the prosecution case.	1	2	3	4	5	6	7
evaluate the defence case.	1	2	3	4	5	6	7
understand the judge's instructions.	1	2	3	4	5	6	7
apply the law to the facts.	1	2	3	4	5	6	7
reach a unanimous verdict.	1	2	3	4	5	6	7

How likely is it that Mark Booth

	Very unlikely					Very likely	
- masturbated Simon Rutter's penis.	1	2	3	4	5	6	7
- forced Justin McCutcheon to touch and masturbate Mark Booth's penis.	1	2	3	4	5	6	7
- masturbated Justin McCutcheon's penis.	1	2	3	4	5	6	7
- inserted his penis into Justin McCutcheon's mouth.	1	2	3	4	5	6	7
- masturbated Timothy Lyons' penis.	1	2	3	4	5	6	7
- inserted his finger into Timothy Lyons's anus.	1	2	3	4	5	6	7

Please tick the box with the best response:

Which statement is false? Mark Booth:

- generally behaved like one of the boys.
- took selected boys for lunch after training.
- invited selected boys to lunch at his house.

Timothy stayed at Mark Booth's house overnight, because

- he wanted to watch TV and eat pizza.
- they went to a movie that night.
- his mother was in hospital.

Ellen Samuels testified that Mark Booth

- looked after her townhouse when she was away.
- entered her house without her knowledge.
- and she dated a few times.

Mark Booth

- had tea at Timothy's house after driving him home.
- was a friend of Timothy's father.
- offered to look after Timothy when his mother was working late.

Please tick the box with the best response:

Timothy said that Mark Booth touched him

<input type="checkbox"/> when Timothy tried on his uniform.
<input type="checkbox"/> after Timothy stayed overnight at Mark Booth's house.
<input type="checkbox"/> after he pushed Timothy into the swimming pool.
Timothy said that he
<input type="checkbox"/> knew about his mother's relationship with Mark Booth.
<input type="checkbox"/> disapproved of his mother going out with Mark Booth.
<input type="checkbox"/> did not know about his mother's relationship with Mark Booth.

In the witness box to what degree did <u>Timothy Lyons</u>					
	Not well		Moderately well		Very well
remain calm under cross-examination.	1	2	3	4	5
control his emotions under cross-examination.	1	2	3	4	5
maintain a stable tone of voice.	1	2	3	4	5
avoid fidgeting.	1	2	3	4	5
maintain good posture.	1	2	3	4	5
appear comfortable.	1	2	3	4	5
remain poised.	1	2	3	4	5
maintain eye contact with people in the courtroom.	1	2	3	4	5
maintain eye contact with the questioning barrister.	1	2	3	4	5
hide his nervousness.	1	2	3	4	5
convey confidence.	1	2	3	4	5
organize his thoughts.	1	2	3	4	5
comfortably admit uncertainty of an answer.	1	2	3	4	5
sit upright.	1	2	3	4	5
lean slightly forward when answering some questions.	1	2	3	4	5
give more than "yes/no" answers.	1	2	3	4	5
behave naturally.	1	2	3	4	5

The accused <u>Mark Booth</u>	Strongly disagree						Strongly agree
cared for the boys when their parents needed assistance.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7
abused the trust of others.	1	2	3	4	5	6	7
abused his position as a coach.	1	2	3	4	5	6	7
was responsible for what happened to him.	1	2	3	4	5	6	7
is a risk to other boys.	1	2	3	4	5	6	7
The complainant <u>Timothy Lyons</u>	Strongly disagree						Strongly agree
was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7
The complainant <u>Simon Rutter</u>	Strongly disagree						Strongly agree
was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7
The complainant <u>Justin McCutcheon</u>	Strongly disagree						Strongly agree
was responsible for what happened to him.	1	2	3	4	5	6	7
misinterpreted the events.	1	2	3	4	5	6	7
was severely harmed by Mark Booth.	1	2	3	4	5	6	7
should have recovered by now.	1	2	3	4	5	6	7
was convincing.	1	2	3	4	5	6	7

To what extent do you agree with the following statements?

The judge's instructions

helped me understand the case.	1	2	3	4	5	6	7
were easy to understand.	1	2	3	4	5	6	7
changed my mind about the verdict.	1	2	3	4	5	6	7
were confusing.	1	2	3	4	5	6	7
were fair to Mark Booth.	1	2	3	4	5	6	7

The trial was fair to

Mark Booth	1	2	3	4	5	6	7
Timothy Lyons	1	2	3	4	5	6	7
Simon Rutter	1	2	3	4	5	6	7
Justin McCutcheon	1	2	3	4	5	6	7

To what extent do you agree with the following statements?

Deliberation	Strongly disagree				Strongly agree	
helped me to understand the case.	1	2	3	4	5	6 7
changed my mind about the verdict.	1	2	3	4	5	6 7
In deliberation I disagreed with other jurors.	1	2	3	4	5	6 7

We would have been informed, if...

Other charges were made against Mark Booth.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth was sexually abusive on other occasions.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth had a prior conviction for child sexual assault.	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Mark Booth had a prior conviction for any other crime.	<input type="checkbox"/> Yes	<input type="checkbox"/> No

Right now, indicate to what extent you feel					
	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Interested	1	2	3	4	5
Distressed	1	2	3	4	5
Excited	1	2	3	4	5
Upset	1	2	3	4	5
Strong	1	2	3	4	5
Guilty	1	2	3	4	5
Scared	1	2	3	4	5
Hostile	1	2	3	4	5
Enthusiastic	1	2	3	4	5
Proud	1	2	3	4	5
Irritable	1	2	3	4	5
Alert	1	2	3	4	5
Ashamed	1	2	3	4	5
Inspired	1	2	3	4	5
Nervous	1	2	3	4	5
Determined	1	2	3	4	5
Attentive	1	2	3	4	5
Jittery	1	2	3	4	5
Active	1	2	3	4	5
Afraid	1	2	3	4	5

Thank you for participating in this survey.

If you are distressed by the survey, please contact Lifeline Australia on 13 11 14
for referral to counselling.

Appendix M: Quantitative coding scheme for transcribed deliberations

Unless specified otherwise, variables were coded each time the topic arose in conversation.

Topic	Code	Notes
Juror reasoning about judicial directions		
Delay in complaint direction	CDD1 correct understanding CDD2 incorrect understanding CDD3 applied direction CDD4 incorrectly applied direction CDD5 delay did not harm complainant's credibility CDD6 delay did harm complainant's credibility	CDD3 & CDD4 if conclusion on credibility based on delay. CDD5 & CDD6 if discussed but no conclusion.
Delay in complaint -forensic disadvantage	ADD1 correct understanding ADD2 incorrect understanding ADD3 correctly applied direction ADD4 incorrectly applied direction	
Delay in complaint other	DID1 correct understanding DID2 incorrect understanding	See instruction below
<i>I am not telling you that these problems for the accused make it impossible for the prosecution to prove its case. If, after carefully considering my warnings and scrutinising the complainants' evidence with great care (in the context of all the other evidence) you are satisfied of the truth and accuracy of the complainants' evidence and you are satisfied beyond reasonable doubt that any of the alleged offences did occur, then it is your duty to return a verdict of guilty in relation to that matter.</i>		
Burden of proof	OP1 understand onus on Pros to prove the case OP2 understand no onus on Defence OP3 believe onus of proof on Def OP4 understand innocent until proven guilty	
Relationship evidence direction	ReID1 correct understanding ReID2 incorrect understanding ReID3 correctly applied	

	ReID4 incorrectly applied	
Tendency evidence direction	TD1 correct understanding TD2 incorrect understanding TD3 correctly applied TD4 incorrectly applied TD5 mention pattern among complainants but not applied as reason for guilt/innocence	
Reasonable doubt	RD1 (reasonable) doubt unrelated to self RD2 defined reasonable doubt RD3 referred (reasonable) doubts re the evidence	
Justin elements	JE11a time indecency occurred JE11b place the indecency occurred JE12 was an act of indecency JE13 act indecent (per community standards) JE14 Justin under 16 years **JE15a time sexual assault occurred JE15b place sexual assault occurred JE16 was sexual intercourse JE17 Justin 10-14 years.	Referred to elements as needing proof (not mere talk about time, place, if act of indecency occurred).
Simon elements	SimE11a time indecency occurred SimE11b place indecency occurred SimE12 was an act of indecency SimE13 act indecent (per community standards) SimE14 Simon under 16 years	
Timothy elements	TimE11a time indecency occurred TimE11b place indecency occurred TimE12 an act of indecency	

	<p>TimE13 act indecent (per community standards)</p> <p>TimE14 Timothy under 16 years</p> <p>TimE15a time sexual assault occurred</p> <p>TimE15b place sexual assault occurred</p> <p>TimE16 was sexual intercourse</p> <p>TimE17 = Timothy between 10 -14 years</p>	
Standard other direction	<p>JD1 correct understanding</p> <p>JD2 incorrect understanding</p>	Any directions not captured above
Time deliberating	<p>WT total words deliberating</p> <p>DWT total words discussing judge's directions</p> <p>CWT total words discussing directions re counts</p>	<p>DWT re (ADD, CDD, DID, JD, OP, ReID, TD) and RD2 (not RD1 or RD3).</p> <p>CWT reading count, what it relates to, discuss elements (JEI, SimEI, TimEI).</p>
Juror reasoning about the evidence		
<p>Identification with the Defendant</p>	<p>AID1 identifies with defendant and reason towards guilt.</p> <p>AID2 identifies with defendant but this doesn't drive assessment of evidence</p> <p>AID3 identifies with defendant and reasons towards acquittal</p>	
Complainant's distress	<p>CD1 more believable</p> <p>CD2 less believable</p> <p>CLD1 lack of distress = more believable</p> <p>CLD2 lack of distress = less believable</p>	
Reference to other child sexual abuse cases	<p>CSAO1 refers to other CSA case when assessing evidence to support conviction</p> <p>CSAO2 refers to other CSA case when assessing evidence to support acquittal</p>	
Prior allegations	<p>PA5 would have been told of the defendant's prior allegations/offences</p> <p>PA6 would not have been told of the defendant's prior allegations/offences</p>	

Questions about trial		
Evidence	E2 desire DNA evidence E3 desire forensic evidence E4 desire medical/psychiatric evidence E5 desire Timothy's mother evidence E6 desire hospital records/ mother's stay E7 desire photos or video of what occurred in the locker room E8 desire other additional evidence E9 belief case is word against word E10 oral evidence is not real evidence	
Fact errors	EF1 error of fact in case (including conduct the subject of each charge) EF2 corrects error	Coded per error (not repeats of same error).
Other charges	OC1 questions why charges not laid re locker room incident OC2 question why charges not laid re Simon and Justin	
Speculation	Spec engaged in speculation	Hypothesise/suggest facts not raised in trial.
Strategy	Strat uses strategy to respond to difficulty in deliberation	
Reasoning about verdict		
Category	Codes	Notes
Verdict without assessing facts	NEV final verdict on a count without discussing evidence	
Not guilty reasons	NGJ1 "gut feeling" guilty, not enough evidence NGJ2 believe cannot convict on oral evidence NGJ3 persuaded BRD of guilt, will not convict (nullification) NGJ4 reluctant to convict for penetration	
Unfair prejudice	UPemo1 reason emotional reaction to case UPemo2 convict for emotional reaction UPlogic1 evidence logically unconnected to reason towards guilty verdict UPlogic2 convict logically unconnected UPlow1 reason lower threshold than BRD should be used UPlow2 convict lower threshold than BRD	

Verdict Order	Order of the final vote for each count: SJT Simon Justin Timothy STJ Simon Timothy,Justin JST Justin, Simon Timothy JTS Justin Timothy Simon TSJ Timothy Simon Justin TJS Timothy Justin Simon T = Timothy (no other complainants) Other = other order	
---------------	--	--

Appendix N: Descriptive statistics for pre-trial and post-trial measures by experimental group in the jury deliberation and reasoning study

	Separate trial standard instructions (<i>N</i> = 105)				Relationship evidence standard instructions (<i>N</i> = 135)				Relationship evidence rel + basic instructions (<i>N</i> = 103)				Relationship evidence rel + basic instructions QT (<i>N</i> = 107)			
	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
CSA-KQ Impact on Children	5.00	34.00	20.41	5.58	7.00	34.00	21.87	5.68	8.00	35.00	21.13	5.50	10.00	35.00	21.37	5.90
CSA-KQ Contextual Influences	4.00	28.00	17.85	4.77	4.00	28.00	17.90	4.77	9.00	28.00	17.78	3.93	6.00	28.00	18.20	4.62
CSA-KQ Total	9.00	59.00	38.26	8.61	17.00	60.00	39.77	9.00	23.00	62.00	38.91	7.66	17.00	62.00	39.57	9.02
FEEBS Pro-Prosecution	5.00	35.00	17.52	6.74	5.00	34.00	16.92	6.50	5.00	30.00	17.76	5.35	5.00	35.00	18.06	6.34
FEEBS Pro-Defence	5.00	35.00	15.08	5.91	5.00	28.00	14.95	5.60	5.00	29.00	14.85	4.88	5.00	28.00	15.08	5.26
PJAQ System Confidence	1.33	7.00	3.90	1.14	1.00	6.50	3.61	1.14	1.67	6.33	3.75	0.92	1.33	7.00	3.81	0.95
PJAQ Conviction Proneness	1.40	7.00	4.09	1.09	1.00	6.40	3.64	1.16	1.20	7.00	3.83	1.10	1.00	7.00	3.73	1.09
PJAQ Cynicism in the Defence	2.00	7.00	4.10	1.09	1.00	6.43	3.87	1.14	1.71	6.00	3.89	0.92	1.14	6.14	3.89	0.96
PJAQ Racial Bias	1.25	6.00	3.30	1.11	1.00	5.25	3.00	0.92	1.00	5.75	3.18	0.89	1.00	5.25	3.27	0.97
PJAQ Social Justice	1.75	6.25	4.11	0.91	1.00	6.00	4.04	1.01	2.25	5.75	4.13	0.83	1.00	6.00	4.13	0.96
PJAQ Innate Criminality	1.00	7.00	3.09	1.32	1.00	6.50	2.82	1.16	1.00	5.25	3.03	0.93	1.00	6.50	2.97	1.21
Case Facts	0.00	6.00	4.58	1.41	1.00	6.00	4.68	1.50	0.00	6.00	4.48	1.52	1.00	6.00	4.42	1.32
Witness Credibility Poise	1.23	4.92	2.95	0.69	1.62	4.92	3.04	0.62	1.15	4.92	3.09	0.76	1.00	4.67	3.02	0.70
Witness Credibility Communication Style	1.40	5.00	3.18	0.65	2.00	5.00	3.31	0.59	1.30	4.90	3.41	0.70	1.10	4.63	3.27	0.64
Overall Witness Credibility	1.24	4.82	3.03	0.65	1.80	4.82	3.13	0.58	1.24	4.88	3.19	0.71	1.12	4.67	3.11	0.65
Positive Affect	10.00	50.00	30.09	8.80	14.00	49.00	29.40	7.57	11.00	50.00	29.85	7.99	10.00	50.00	29.03	9.18
Negative Affect	10.00	39.00	16.56	7.72	10.00	47.00	16.66	6.70	10.00	42.00	16.61	6.52	10.00	40.00	16.61	6.95
Cognitive Effort	1.00	7.00	4.83	1.32	1.00	7.00	5.10	1.27	2.11	7.00	5.40	1.23	1.56	7.00	5.12	1.26

APPENDIX N CONTINUED

	Joint tendency instructions (<i>N</i> = 93)				Joint Tendency instructions QT (<i>N</i> = 100)				Joint 2 fewer witnesses (<i>N</i> = 108)				Joint basic instructions (<i>N</i> = 83)			
	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
CSA-KQ Impact on Children	9.00	32.00	22.31	5.39	7.00	35.00	20.58	5.62	5.00	35.00	21.00	6.93	5.00	35.00	20.63	6.22
CSA-KQ Contextual Influences	4.00	28.00	18.28	4.86	5.00	28.00	17.60	4.37	4.00	28.00	18.11	4.71	4.00	28.00	17.26	4.91
CSA-KQ Total	15.00	59.00	40.59	8.38	20.00	63.00	38.19	8.20	9.00	59.00	39.11	9.97	9.00	63.00	37.89	9.47
FEEBS Pro-Prosecution	5.00	35.00	17.55	6.19	5.00	33.00	17.37	6.39	5.00	35.00	17.78	6.33	5.00	35.00	18.90	6.51
FEEBS Pro-Defence	5.00	35.00	14.49	5.64	5.00	31.00	14.86	5.86	5.00	34.00	15.62	5.31	5.00	32.00	15.25	5.40
PJAQ System Confidence	1.67	7.00	3.75	1.01	1.50	6.33	3.82	1.09	1.33	6.33	3.77	1.05	1.83	6.67	3.99	1.03
PJAQ Conviction Proneness	1.00	7.00	3.94	1.10	1.20	6.20	3.75	1.22	1.40	6.40	3.87	1.21	1.80	7.00	3.87	1.11
PJAQ Cynicism in the Defence	1.71	7.00	4.07	1.04	1.57	6.29	3.86	1.07	1.71	6.57	3.97	1.01	1.71	6.14	4.16	1.04
PJAQ Racial Bias	1.00	5.50	3.18	0.99	1.25	6.00	3.18	1.01	1.00	6.50	3.30	1.06	1.50	5.50	3.39	0.96
PJAQ Social Justice	2.00	7.00	4.29	1.01	1.25	6.00	4.04	0.95	1.50	7.00	4.05	1.00	1.75	7.00	4.31	0.93
PJAQ Innate Criminality	1.00	7.00	2.87	1.18	1.00	5.75	2.95	1.09	1.00	6.25	2.93	1.15	1.25	6.25	2.99	1.12
Case Facts	0.00	6.00	3.30	1.72	1.00	6.00	3.70	1.37	0.00	6.00	3.50	1.49	0.00	6.00	3.60	1.52
Witness Credibility Poise	1.92	5.00	3.40	0.74	2.00	5.00	3.53	0.63	2.00	5.00	3.67	0.77	1.54	4.85	3.45	0.68
Witness Credibility Communication Style	2.00	5.00	3.55	0.70	1.70	5.00	3.64	0.58	2.00	5.00	3.79	0.70	1.30	5.00	3.58	0.63
Overall Witness Credibility	2.00	5.00	3.45	0.70	1.94	5.00	3.56	0.59	2.00	5.00	3.69	0.73	1.41	4.82	3.50	0.63
Positive Affect	10.00	50.00	29.96	9.34	10.00	48.00	29.51	9.31	10.00	50.00	30.21	8.61	10.00	45.00	28.89	8.04
Negative Affect	10.00	43.00	18.11	8.09	9.00	43.00	16.92	7.01	10.00	44.00	18.28	7.95	10.00	41.00	17.57	6.71
Cognitive Effort	1.11	7.00	5.24	1.25	2.00	7.00	5.50	1.05	2.00	7.00	5.48	1.16	1.11	7.00	5.36	1.14

APPENDIX N CONTINUED

	Tendency evidence basic instructions (N = 85)				Tendency evidence tend + basic instructions (N = 112)			
	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
CSA-KQ Impact on Children	5.00	34.00	21.16	5.75	5.00	33.00	20.92	5.77
CSA-KQ Contextual Influences	4.00	28.00	18.36	4.92	4.00	28.00	17.42	4.97
CSA-KQ Total	9.00	60.00	39.52	8.84	14.00	57.00	38.34	8.68
FEEBS Pro-Prosecution	5.00	31.00	15.82	6.33	6.00	33.00	18.50	5.82
FEEBS Pro-Defence	5.00	27.00	14.79	5.35	5.00	27.00	15.27	5.06
PJAQ System Confidence	1.33	6.00	3.49	0.99	1.33	6.00	3.74	0.96
PJAQ Conviction Proneness	1.00	6.40	3.76	1.19	1.80	6.20	3.88	1.10
PJAQ Cynicism in the Defence	1.29	6.14	3.96	1.09	1.29	6.29	3.97	1.04
PJAQ Racial Bias	1.25	5.25	3.06	0.92	1.00	5.50	3.38	0.96
PJAQ Social Justice	1.00	6.25	3.98	1.06	2.00	6.50	4.26	0.94
PJAQ Innate Criminality	1.00	7.00	2.79	1.19	1.00	6.25	2.94	1.06
Case Facts	0.00	6.00	3.75	1.57	0.00	6.00	3.68	1.43
Witness Credibility Poise	1.38	5.00	3.31	0.70	1.77	5.00	3.46	0.70
Witness Credibility Communication Style	1.60	5.00	3.49	0.67	1.50	5.00	3.62	0.69
Overall Witness Credibility	1.47	5.00	3.37	0.66	1.82	5.00	3.51	0.67
Positive Affect	12.00	47.00	28.42	8.37	10.00	47.00	29.56	8.83
Negative Affect	10.00	39.00	16.34	7.06	10.00	42.00	17.11	7.41
Cognitive Effort	1.00	7.00	4.93	1.49	1.44	7.00	5.31	1.32

Appendix O. Jury reasoning in separate trials with appropriate jury directions⁶⁷⁹

The analyses in Chapter 5.2 explored jury reasoning as more inculpatory evidence was admitted at trial without confounding those changes in the evidence with changes in the jury directions. In the following analyses, the jury directions that are typically given in the respective types of separate cases in tangent with the additional evidence were included. As in Chapter 5.2, the focus of the analyses is the complainant with the moderately strong evidence (Timothy), with results about mock juror *CSA Knowledge*, the criminal intent of the defendant, witness credibility assessments, victim blame, factual culpability and verdict. The influence of jury directions is examined in Chapter 5.5.

These results are based on responses from a total of 27 juries and 320 mock jurors who viewed the three types of separate trials and received the appropriate legal directions accompanying those trials. In all, there were 9 juries ($n = 105$ jurors) who viewed the basic separate trial and received standard jury directions; 9 juries ($n = 103$ jurors) who viewed the trial with relationship evidence and received standard directions plus a jury direction on context evidence, and 9 juries ($n = 112$ jurors) who viewed the tendency evidence trial and received standard directions plus a jury direction on tendency evidence.

Criminal intent of the defendant

Two-level regression analysis was conducted using the trial or type of trial (separate trial vs. relationship evidence; separate trial vs. tendency evidence) as a predictor of the extent of criminal intent inferred on the part of the defendant. As more evidence was presented for the prosecution, the inferred criminal intent of the defendant increased, with significantly more intent inferred in the relationship evidence ($M = 4.97$) and tendency evidence trials ($M = 5.70$) compared to the simple separate trial ($M = 3.94$).⁶⁸⁰ The intra-class correlation ($ICC = .255$) revealed that 25.5 per cent of the variance in criminal intent was attributable to jury group clusters, indicating a strong design effect.

⁶⁷⁹ Trials 1 versus 3 versus 6.

⁶⁸⁰ Separate trial v relationship evidence: $\beta = 1.052$, $SE = .267$, $Z = 3.935$, $p < .001$; Separate trial v tendency evidence: $\beta = 1.755$, $SE = .274$, $Z = 6.407$, $p < .001$.
Separate trial: $Std Dev = 1.48$; Relationship evidence: $Std Dev = 1.57$; Tendency evidence: $Std Dev = 1.53$.

Factual culpability of the defendant

Two-level regression analysis tested the effect of mock juror *CSA Knowledge* and trial type on perceived factual culpability of the defendant. The intra-class correlation indicated that 27.1 per cent of the variance in factual culpability for the non-penetrative offence and 29.7 per cent of the variance in factual culpability for the penetrative offence was unexplained by trial type and could be attributed to the jury group variable.

Juror and mean jury *CSA Knowledge* did not predict factual culpability.⁶⁸¹ However, trial type was a significant Level-2 predictor of factual culpability.⁶⁸² Juries deliberating about the trial which included relationship evidence and tendency evidence perceived the defendant as significantly more factually culpable on both counts (relationship evidence non-penetrative offence: $M = 5.51$, penetrative offence: $M = 5.47$; tendency evidence non-penetrative offence: $M = 5.77$, penetrative offence: $M = 5.64$) than did mock jurors who deliberated about a separate trial without additional Crown evidence (non-penetrative offence: $M = 4.39$, penetrative offence: $M = 4.12$).⁶⁸³

Verdict

Verdicts by type of trial are presented in Table 3. The effect of additional evidence on verdict reduced the high proportion of acquittals in the most straightforward separate trial (77.8%-88.9%), compared to the more complex separate trial with tendency evidence trial, in which a higher proportion of guilty verdicts were returned (55.6%). Almost half of the juries who viewed the relationship evidence trial were hung, while only one third convicted the defendant.

⁶⁸¹ Juror level: Non-penetrative offence: $\beta = 0.017$, $SE = .012$, $Z = 1.431$, $p = .162$; Penetrative offence: $\beta = 0.020$, $SE = .012$, $Z = 1.755$, $p = .079$. Jury level: Non-penetrative offence: $\beta = 0.024$, $SE = .091$, $Z = 0.266$, $p = .790$; Penetrative offence: $\beta = 0.095$, $SE = .083$, $Z = 1.140$, $p = .254$.

⁶⁸² Non-penetrative offence: Separate trial v relationship evidence: $\beta = 1.098$, $SE = .317$, $Z = 3.464$, $p = .001$; Separate trial v tendency evidence: $\beta = 1.347$, $SE = .365$, $Z = 3.695$, $p < .001$. Penetrative offence: Separate trial v relationship evidence: $\beta = 1.295$, $SE = .286$, $Z = 4.533$, $p < .001$; Separate trial v tendency evidence: $\beta = 1.493$, $SE = .347$, $Z = 4.299$, $p < .001$.

⁶⁸³ Separate trial non-penetrative offence: $Std Dev = 1.63$, penetrative offence: $Std Dev = 1.61$; relationship evidence non-penetrative offence: $Std Dev = 1.61$, penetrative offence: $Std Dev = 1.62$; tendency evidence non-penetrative offence: $Std Dev = 1.46$, penetrative offence: $Std Dev = 1.63$.

Table 3. Jury and mock juror verdicts by count and by type of trial (per cent)

	Count 1: Masturbate complainant			Count 2: Digital-anal penetration		
	Not Guilty		Hung jury	Not Guilty		Hung jury
(a) Jury verdict						
Separate trial	77.8	11.1	11.1	88.9	0.0	11.1
Relationship evidence	22.2	33.3	44.4	22.2	33.3	44.4
Tendency evidence	22.2	55.6	22.2	22.2	55.6	22.2
(b) Juror verdict						
Separate trial	81.0	19.0	--	90.5	9.5	--
Relationship evidence	31.1	68.9	--	30.1	69.9	--
Tendency evidence	24.5	75.5	--	23.6	76.4	--

Note. Separate trial: nine juries, $n = 105$ jurors; Relationship evidence: nine juries, $n = 103$ jurors; Tendency evidence: nine juries, $n = 112$ jurors.

Due to the high proportion of hung juries, multilevel analyses of jury verdicts were not feasible. Therefore, mock jurors' individual post-deliberation verdicts were used as a dependent measure. Separate two-level regression analyses were conducted to test the impact of mock juror pre-trial *CSA Knowledge*, the perceived convincingness of the complainant, and to test the type of trial on individual post-deliberation verdicts for Count 1 (non-penetrative offence) and Count 2 (penetrative offence).

Count 1: Masturbate complainant

Multilevel regression analyses revealed that mock juror *CSA Knowledge* (juror level) did not predict verdict on Count 1 (masturbate complainant),⁶⁸⁴ compared to perceived convincingness of the complainant which did predict verdict. The odds of conviction were 2.2 times greater when mock jurors perceived the complainant as more convincing.⁶⁸⁵

⁶⁸⁴ $\beta = 0.035$, $SE = 0.026$, $Z = 1.338$, $p = .181$, Odds Ratio = 1.036, 95% CI [0.984; 1.091].

⁶⁸⁵ $\beta = 0.789$, $SE = 0.129$, $Z = 6.125$, $p < .001$, Odds Ratio = 2.201, 95% CI [1.710; 2.833].

Jury level analyses revealed that mean jury *CSA Knowledge* did not predict verdict on Count 1.⁶⁸⁶ However, trial type predicted verdict for Count 1 as the conviction rates in trials with relationship evidence and tendency evidence were significantly higher than those in the separate trial without relationship evidence⁶⁸⁷ or tendency evidence.⁶⁸⁸

Count 2: Digital-anal penetration

Neither juror *CSA Knowledge*⁶⁸⁹ nor jury *CSA Knowledge*⁶⁹⁰ predicted verdict on the penetrative charge, compared to the perceived convincingness of the complainant which did predict verdict at the juror level.⁶⁹¹ Specifically, the odds of conviction were 2.5 times greater by mock jurors who perceived the complainant as more convincing. After controlling for mock juror *CSA Knowledge* and the perceived convincingness of the complainant, trial type predicted verdict in that juries who viewed the relationship evidence and tendency evidence trials were more likely to convict than jurors who viewed the basic separate trial.⁶⁹² Further, the effect was larger for the tendency evidence than the relationship evidence trial.

Victim blame for the alleged abuse

Two-level regression analysis for victim blame was conducted with mock juror pre-trial *CSA Knowledge* and type of trial as predictors. At the juror level, the simple regression with a random slope was used, with mock juror *CSA Knowledge* scores as a predictor of victim blame. At the jury level, the mean group *CSA Knowledge* and type of trial were used to predict victim blame.

Analyses revealed mock juror *CSA Knowledge* predicted victim blame at the juror level.⁶⁹³ This effect disappeared at the jury level, such that the mean jury *CSA Knowledge* did not predict verdict.⁶⁹⁴ However, type of trial significantly predicted victim blame at the jury level in that mock jurors exposed to relationship ($M = 2.02$)⁶⁹⁵ and tendency evidence ($M = 2.03$)⁶⁹⁶ were

⁶⁸⁶ $\beta = -0.200$, $SE = 0.293$, $Z = -0.684$, $p = .494$.

⁶⁸⁷ $\beta = 4.569$, $SE = 1.973$, $Z = 3.523$, $p < .001$.

⁶⁸⁸ $\beta = 4.864$, $SE = 1.260$, $Z = 3.860$, $p < .001$.

⁶⁸⁹ $\beta = 0.049$, $SE = .025$, $Z = 1.921$, $p = .055$, Odds Ratio = 1.050, 95% CI [0.999; 1.104].

⁶⁹⁰ $\beta = 0.004$, $SE = 0.286$, $Z = 0.013$, $p = .989$.

⁶⁹¹ $\beta = 0.909$, $SE = 0.168$, $Z = 5.404$, $p < .001$.

⁶⁹² Separate trial v relationship evidence: $\beta = 5.716$, $SE = 1.294$, $Z = 4.418$, $p < .001$; Separate trial v tendency evidence: $\beta = 6.075$, $SE = 1.162$, $Z = 5.228$, $p < .001$.

⁶⁹³ $\beta = -0.029$, $SE = 0.010$, $Z = -2.882$, $p = .004$.

⁶⁹⁴ $\beta = -0.018$, $SE = 0.030$, $Z = -0.594$, $p = .553$.

⁶⁹⁵ $\beta = -0.410$, $SE = 0.110$, $Z = -3.708$, $p < .001$.

⁶⁹⁶ $\beta = -0.465$, $SE = 0.131$, $Z = -3.548$, $p < .001$.

less likely to blame the complainant, Timothy, for the alleged abuse than were jurors in the separate trial without relationship or tendency evidence ($M = 2.47$).⁶⁹⁷

⁶⁹⁷ Separate trial: *Std Dev* = 1.03; relationship evidence: *Std Dev* = 1.03; tendency evidence: *Std Dev* = 1.10.

Appendix P Jury directions

Judge's summing up in the basic separate trial.

Judge: The accused stands before you upon an indictment that he committed one count of act of indecency towards a person under the age of 16 years, namely Timothy Lyons, and one count of sexual intercourse against a child above the age of 10 years but under the age of 14 years, namely Mr Timothy Lyons.

In relation to both counts the accused has pleaded "not guilty". It is your duty and your responsibility, therefore, to consider whether the accused is "guilty" or "not guilty" of each count of sexual assault and to return your verdict on each count according to the evidence you have heard.

You have heard addresses from counsel for the Crown and counsel for the accused. I remind you that in no sense do those submissions amount to evidence in the case.

You must, as a jury, act impartially, dispassionately and fearlessly. You must not let sympathy or emotion sway your judgment. You are expected to use your individual qualities of reasoning, experience, common sense, as well as your understanding of people and human affairs during the course of your deliberation

A critical part of the criminal justice system is the presumption of innocence. This means that a person who is charged with a criminal offence is presumed to be innocent unless and until the Crown persuades a jury that the person is guilty beyond reasonable doubt.

That burden rests upon the Crown in respect of every element or essential fact that makes up the offences with which the accused has been charged. There is no obligation whatsoever on the accused to prove any fact or issue that is in dispute before you. It is not for the accused to prove his innocence but for the Crown to prove his guilt in relation to each offence and to prove it beyond reasonable doubt. The Crown does not have to prove, however, every single fact in the case beyond reasonable doubt. The Crown's onus is to prove, beyond a reasonable doubt, the elements of the charges against the accused. I shall shortly outline for you the elements of each charge that the Crown must prove beyond reasonable doubt.

That is the high standard of proof that the Crown must achieve before you can convict the accused.

You will be required to deliver two different verdicts. Whether your verdicts are "guilty" or "not guilty", they ought to be unanimous. That does not mean each of you must agree upon the same reasons for your verdicts. You may individually rely upon different parts of the evidence or place a different emphasis upon parts of the evidence. However you arrive at your decision of "guilty" or "not guilty", it must be the decision of all of you, unanimously, before it can become your verdict.

In this trial, the Crown must prove each element of each of the two counts against the accused. I will now describe these elements or essential facts that have to be proved by the Crown beyond reasonable doubt.

Count 1 – act of indecency

For count 1, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;

2. The accused committed an act of indecency;
3. That the act was indecent; and
4. At the relevant time, Timothy Lyons was under the age of 16 years.

Indecent: the word indecent means contrary to the standards of ordinary and, therefore, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A consensual act between those old enough to consent that occurs in privacy would not generally be regarded as an act of indecency. The law also provides that a child under 16 years of age cannot consent to an act of indecency.

Count 2 – sexual intercourse against a child above the age of 10 years and under the age of 14 years

For count 2, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;
2. The accused had sexual intercourse with Timothy Lyons; and
3. At the time Timothy Lyons was above the age of 10 years and under the age of 16 years.

Sexual intercourse: “sexual intercourse” is defined under our Crimes Act to include “sexual connection occasioned by the penetration, to any extent, of the anus of any person by any part of the body of another person.” The law provides that a person under the age of 16 years does not have the capacity to consent to sexual intercourse.

Finally, I remind you that the failure to prove any element of any count beyond reasonable doubt would mean that the accused is not guilty of that count.

You must consider each count separately and return a separate verdict of guilty or not guilty on each of the counts. This means that you are entitled to bring in a verdict of guilty on one count and not guilty on the other count if there is a good reason in the evidence for that outcome.

If you find Timothy’s evidence to be truthful, it is sufficient to satisfy the elements of act of indecency that the prosecution has to prove beyond reasonable doubt for Count 1.

Furthermore, if you find Timothy’s evidence to be truthful, it is sufficient to satisfy the elements of sexual intercourse against a child above the age of 10 years and under the age of 14 years that the prosecution has to prove beyond reasonable doubt for Count 2.

However, just because the accused testifies does not mean that the onus of proof shifts to the accused. Even if you reject the evidence of the accused, you must still be satisfied that the prosecution has established beyond reasonable doubt that the accused has committed each element of the two charges against him.

In relation to both counts, there was delay of decades between the alleged incidents and the matters being reported to police.

However I am required by law to direct you that a delay in complaint, even a long delay, does not necessarily mean that a complaint is false. There may be good reasons why a person who alleges they have been sexually assaulted may fail to report, to delay in reporting, such an offence.

It is for you to evaluate such considerations. You should also take into account the fact that the accused has denied each of the offences.

It is most important that you appreciate the effects of delay on the ability of the accused to defend himself by testing the prosecution's evidence or bringing forward evidence in his own case, to establish a reasonable doubt about his guilt.

One of the effects is the inability to properly inspect Mr Booth's former residences to determine the layout of Mr Booth's bedroom about which Timothy Lyons gave evidence. There is also the inability to determine the existence of another house or apartment that Mr Booth may have used. These difficulties put the accused at a significant disadvantage in testing the prosecution evidence, or in bringing forward evidence to establish a reasonable doubt about his guilt, or both.

Had the allegations been brought to light and the prosecution commenced much sooner, the complainant's memory for details would have been clearer, such as when the alleged incidents occurred. This may have enabled his evidence to be checked against independent sources so as to verify it, or disprove it. The complainant's inability to recall precise details of the alleged incidents makes it difficult for the accused to throw doubt on the complainant's evidence by pointing to the circumstances which may contradict him.

Another aspect of the accused's disadvantage is that if he had learned of the allegations at a much earlier time he may have been able to find more witnesses or items of evidence that might have either contradicted the complainant or supported his case, or both. He may have been able to recall with some precision what he was doing and where he was at particular times on particular dates and to have been able to bring forward evidence to support him which could have been used in the cross-examination of the complainant.

With a final reminder that the verdicts you reach must be unanimous, I ask you to retire to the jury room to consider your verdicts.

Judge's summing up in the relationship evidence trial

Judge: Members of the jury, it is now my job to instruct you about the legal principles that you must apply to the facts of this case. The accused stands before you upon an indictment that he committed one count of act of indecency towards a person under the age of 16 years, namely Timothy Lyons, and one count of sexual intercourse against a child above the age of 10 years but under the age of 14 years, namely Mr Timothy Lyons.

In relation to both counts the accused has pleaded “not guilty”. It is your duty and your responsibility, therefore, to consider whether the accused is “guilty” or “not guilty” of each count of sexual assault and to return your verdict on each count according to the evidence you have heard.

The verdicts you give are for you and you alone, because you alone are the judges of the facts. I am the judge of the law. I have nothing to do with the facts or your decisions in relation to them. I have nothing to do with what evidence is accepted by you as truthful, or what evidence you reject; nor indeed the weight you might give to any one particular part of the evidence or what inferences you draw from that evidence. However, the principles of law which I will give to you, you are bound to accept. You are bound to apply them to the facts of the case as you find them to be.

In my summing up, I do not propose to try to persuade you one way or the other—that is not my task. It is necessary for you to consider the totality of all the evidence and not only the evidence highlighted by counsel.

It is for you to assess the evidence of the Crown witnesses and the accused. You are not obliged to accept the whole of their evidence. You may, if you choose, accept parts of their evidence and reject other parts of their evidence. Your ultimate decision as to what parts of their evidence you accept or reject may be based on all manner of things, including what they said; the manner in which they said it; and the general impression which they made when giving evidence.

You have heard addresses from counsel for the Crown and counsel for the accused. I remind you that in no sense do those submissions amount to evidence in the case.

You have very important matters to decide in this case. The privilege which you have of sitting in judgment upon your fellow citizen is one which carries with it corresponding duties and obligations. You must, as a jury, act impartially, dispassionately and fearlessly. You must not let sympathy or emotion sway your judgment. You are expected to use your individual qualities of reasoning, experience, common sense, as well as your understanding of people and human affairs during the course of your deliberations.

Let me now say something to you about the onus of proof. A critical part of the criminal justice system is the presumption of innocence. This means that a person who is charged with a criminal offence is presumed to be innocent unless and until the Crown persuades a jury that the person is guilty beyond reasonable doubt.

The obligation to prove the guilt of the accused is placed squarely on the Crown. That burden rests upon the Crown in respect of every element or essential fact that makes up the offences

with which the accused has been charged. That burden never shifts to the accused. There is no obligation whatsoever on the accused to prove any fact or issue that is in dispute before you. It is not for the accused to prove his innocence but for the Crown to prove his guilt in relation to each offence and to prove it beyond reasonable doubt. The Crown does not have to prove, however, every single fact in the case beyond reasonable doubt. The Crown's onus is to prove, beyond a reasonable doubt, the elements of the charges against the accused. I shall shortly outline for you the elements of each charge that the Crown must prove beyond reasonable doubt.

At the end of your consideration of the evidence in the trial and the submissions made to you by the parties you must ask yourself: 'Has the Crown proved the guilt of the accused beyond reasonable doubt?' That is the high standard of proof that the Crown must achieve before you can convict the accused. If the answer is "Yes", the appropriate verdict is "Guilty". If the answer is "No", the verdict must be "Not guilty".

You will be required to deliver three different verdicts. Whether your verdicts are "guilty" or "not guilty", they ought to be unanimous. That does not mean each of you must agree upon the same reasons for your verdicts. You may individually rely upon different parts of the evidence or place a different emphasis upon parts of the evidence. However you arrive at your decision of "guilty" or "not guilty", it must be the decision of all of you, unanimously, before it can become your verdict.

In this trial, the Crown must prove each element of each of the three counts against the accused. I will now describe these elements or essential facts that have to be proved by the Crown beyond reasonable doubt.

Count 1 – act of indecency

For count 1, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;
2. The accused committed an act of indecency;
3. That the act was indecent; and
4. At the relevant time, Timothy Lyons was under the age of 16 years.

Indecent: the word indecent means contrary to the standards of ordinary and, therefore, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. A consensual act between those old enough to consent that occurs in privacy would not generally be regarded as an act of indecency. The law also provides that a child under 16 years of age cannot consent to an act of indecency.

Count 2 – sexual intercourse against a child above the age of 10 years and under the age of 14 years

For count 2, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;
2. The accused had sexual intercourse with Timothy Lyons; and

3. At the time Timothy Lyons was above the age of 10 years and under the age of 16 years.

Sexual intercourse: “sexual intercourse” is defined under our Crimes Act to include “sexual connection occasioned by the penetration, to any extent, of the anus of any person by any part of the body of another person.” The law provides that a person under the age of 16 years does not have the capacity to consent to sexual intercourse.

Finally, I remind you that the failure to prove any element of any count beyond reasonable doubt would mean that the accused is not guilty of that count.

There are two counts in this trial which are being heard together as a matter of convenience. You must consider each count separately and return a separate verdict of guilty or not guilty on each of the counts. This means that you are entitled to bring in a verdict of guilty on one count and not guilty on the other count if there is a good reason in the evidence for that outcome.

If you accept beyond reasonable doubt that what Timothy Lyons said actually occurred then the case on each count would be proved. But if you have a reasonable doubt concerning the truthfulness or reliability of the evidence of Timothy Lyons in relation to one or other of the counts, that doubt must be taken into account in assessing the truthfulness or reliability of his evidence generally. If you were to find the accused not guilty on one count, particularly if that was because you had doubts about the reliability of Timothy Lyons’ evidence, you would have to consider whether or how that conclusion affected your consideration of the remaining count. While you must consider each count separately, there must be a logical consistency in your verdicts.

If you find Timothy’s evidence to be truthful, it is sufficient to satisfy the elements of act of indecency that the prosecution has to prove beyond reasonable doubt for Count 1.

Furthermore, if you find Timothy’s evidence to be truthful, it is sufficient to satisfy the elements of sexual intercourse against a child above the age of 10 years and under the age of 14 years that the prosecution has to prove beyond reasonable doubt for Count 2.

However, it is up to you to decide whether the evidence given by Timothy in cross-examination and the inconsistencies in his evidence, are sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

Based on the cross-examination of the other prosecution witness, Mrs Ellen Samuels, it is also up to you to decide whether her evidence in cross-examination and any inconsistencies in her evidence, are sufficient to undermine her credibility as a witness such that you decide that some or all of her evidence is unreliable.

Finally, the evidence of the accused should be treated in same way as that of other witnesses. This means that it is up to you to decide whether the evidence Mr Booth gave in cross-examination is sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

However, just because the accused testifies does not mean that the onus of proof shifts to the accused. Even if you reject the evidence of the accused, you must still be satisfied that the

prosecution has established beyond reasonable doubt that the accused has committed each element of the two charges against him.

In addition to the evidence led by the Crown specifically on the counts in the indictment, the Crown has led evidence of other alleged misconduct by the accused towards the complainant. I shall, for the sake of convenience, refer to this evidence as evidence of “other acts”. The evidence of the other acts is as follows:

1. That the accused encouraged Timothy to parade up and down the equipment room like a model, half-naked and then naked;
2. That the accused filmed and photographed Timothy in various naked poses;
3. That the accused touched Timothy’s bottom and genitals while encouraging him to perform naked poses.

Without the evidence of these other acts the Crown says, you may wonder about the likelihood of apparently isolated acts occurring suddenly without any reason or any circumstance to link them in any way. If you had not heard about the evidence of other acts, you may have thought that Timothy Lyons’ evidence was less credible because it was less understandable. So the evidence is placed before you only to answer questions that might otherwise arise in your mind about the particular allegations in the charges in the indictment.

If, for example, the particular acts charged are placed in a wider context, that is, a context of what the complainant alleges was an ongoing history of the accused’s conduct toward him, then what might appear to be a curious feature of the complainant’s evidence — that he willingly stayed at the accused’s house and slept in the accused’s bed even after the first occasion when he says he was sexually assaulted — would disappear. It is for that reason that the law permits a complainant to give an account of the alleged sexual history between himself and an accused person in addition to the evidence given in support of the charges in the indictment. It is to avoid any artificiality or unreality in the presentation of the evidence from the complainant. The complainant’s account of other acts by the accused allows him to more naturally and intelligibly explain his account of what allegedly took place.

However, I must give you some important warnings with regard to the use of this evidence of other acts. Firstly, you must not use this evidence of other acts as establishing a tendency on the part of the accused to commit offences of the type charged. You cannot act on the basis that the accused is likely to have committed the offences charged because the complainant made other allegations against him. This is not the reason that the Crown placed the evidence before you. The evidence has a very limited purpose as I have explained it to you, and it cannot be used for any other purpose or as evidence that the particular allegations contained in the charges have been proved beyond reasonable doubt.

Secondly, you must not substitute the evidence of the other acts for the evidence of the specific allegations contained in the charges in the indictment. The Crown is not charging a course of misconduct by the accused but has charged particular allegations arising in what the

complainant says, was a course of sexual misconduct. You are concerned with the particular and precise occasion alleged in each charge.

You must not reason that, just because the accused may have done something wrong to the complainant on some or other occasion, he must have done so on the occasions alleged in the indictment. You cannot punish the accused for other acts attributed to him by finding the accused guilty of the charges in the indictment. Such a line of reasoning would amount to a misuse of the evidence and would not be in accordance with the law.

The delay in making a complaint is a matter that you may take into account in assessing the credibility of the complainant's evidence. In relation to both counts, there was delay of decades between the alleged incidents and the matters being reported to police.

However I am required by law to direct you that a delay in complaint, even a long delay, does not necessarily mean that a complaint is false. There may be good reasons why a person who alleges they have been sexually assaulted may fail to report, to delay in reporting, such an offence.

From your own knowledge of the world, you might think that there could be cases where embarrassment, guilt or worry about the reactions of family or friends might cause a person who alleges sexual abuse to suppress what has taken place.

In this trial, there may be good reasons from the evidence why, if the complainant was sexually abused in his youth, as alleged, he might maintain silence for a long time.

It is for you to evaluate such considerations. You should also take into account the fact that the accused has denied each of the offences.

There is a further warning I should give you relating the delay in complaint. It is most important that you appreciate the effects of delay on the ability of the accused to defend himself by testing the prosecution's evidence or bringing forward evidence in his own case, to establish a reasonable doubt about his guilt.

One of the effects is the inability to properly inspect Mr Booth's former residences to determine the layout of Mr Booth's bedroom about which Timothy Lyons gave evidence. There is also the inability to determine the existence of another house or apartment that Mr Booth may have used. These difficulties put the accused at a significant disadvantage in testing the prosecution evidence, or in bringing forward evidence to establish a reasonable doubt about his guilt, or both.

Had the allegations been brought to light and the prosecution commenced much sooner, the complainant's memory for details would have been clearer, such as when the alleged incidents occurred. This may have enabled his evidence to be checked against independent sources so as to verify it, or disprove it. The complainant's inability to recall precise details of the alleged incidents makes it difficult for the accused to throw doubt on the complainant's evidence by pointing to the circumstances which may contradict him.

Another aspect of the accused's disadvantage is that if he had learned of the allegations at a much earlier time he may have been able to find more witnesses or items of evidence that might have either contradicted the complainant or supported his case, or both. He may have

been able to recall with some precision what he was doing and where he was at particular times on particular dates and to have been able to bring forward evidence to support him which could have been used in the cross-examination of the complainant.

As a result, I warn you that before you convict the accused you must give the prosecution case the most careful scrutiny and bear in mind the matters I have just been speaking about—the fact that the complainant’s evidence has not been tested to the extent that it otherwise might have been and the diminished ability of the accused to bring forward evidence to challenge it, or to support his defence.

However, I am not telling you that these problems for the accused make it impossible for the prosecution to prove its case. If, after carefully considering my warnings and scrutinising the complainant’s evidence with great care (in the context of all the other evidence) you are well satisfied of the truth and accuracy of the complainant’s evidence and you are satisfied beyond reasonable doubt that any of the alleged offences did occur, then it is your duty to return a verdict of guilty in relation to that matter.

I have now completed my summing-up. With a final reminder that the verdicts you reach must be unanimous, I ask you to retire to the jury room to consider your verdicts on both counts.

Judge's summing up in the joint trial.

Judge: Members of the jury, it is now my job to instruct you about the legal principles that you must apply to the facts of this case. The accused stands before you upon an indictment that he committed one count of act of indecency against a male person under the age of 16 years, namely Mr Simon Rutter, the first complainant in this case. He also stands before you on two counts of act of indecency towards a person under the age of 16 years, namely Justin McCutcheon, the second complainant in this case, and one count of sexual intercourse against a child above the age of 10 years and under the age of 14 years, namely Justin McCutcheon. Finally, he stands before you on one count of act of indecency towards a person under the age of 16 years, namely Timothy Lyons, the third complainant in this case, and one count of sexual intercourse against a child above the age of 10 years but under the age of 14 years, namely Mr Timothy Lyons.

In relation to all six counts the accused has pleaded “not guilty”. It is your duty and your responsibility, therefore, to consider whether the accused is “guilty” or “not guilty” of each count of sexual assault and to return your verdict on each count according to the evidence you have heard.

The verdicts you give are for you and you alone, because you alone are the judges of the facts. I am the judge of the law. I have nothing to do with the facts or your decisions in relation to them. I have nothing to do with what evidence is accepted by you as truthful, or what evidence you reject; nor indeed the weight you might give to any one particular part of the evidence or what inferences you draw from that evidence. However, the principles of law which I will give to you, you are bound to accept. You are bound to apply them to the facts of the case as you find them to be.

In my summing up, I do not propose to try to persuade you one way or the other—that is not my task. It is necessary for you to consider the totality of all the evidence and not only the evidence highlighted by counsel.

It is for you to assess the evidence of each of the complainants, the other Crown witnesses and the accused. You are not obliged to accept the whole of their evidence. You may, if you choose, accept parts of their evidence and reject other parts of their evidence. Your ultimate decision as to what parts of their evidence you accept or reject may be based on all manner of things, including what they said; the manner in which they said it; and the general impression they made when giving evidence.

You have heard addresses from counsel for the Crown and counsel for the accused. I remind you that in no sense do those submissions amount to evidence in the case.

You have very important matters to decide in this case. The privilege which you have of sitting in judgment upon your fellow citizen is one which carries with it duties and obligations. You must, as a jury, act impartially, dispassionately and fearlessly. You must not let sympathy or emotion sway your judgment. You are expected to use your individual qualities of reasoning, experience, common sense, as well as your understanding of people and human affairs during the course of your deliberations.

Let me now say something to you about the onus of proof. A critical part of the criminal justice system is the presumption of innocence. This means that a person who is charged with

a criminal offence is presumed to be innocent unless and until the Crown persuades a jury that the person is guilty beyond reasonable doubt.

The obligation to prove the guilt of the accused is placed squarely on the Crown. That burden rests upon the Crown in respect of every element or essential fact that makes up the offences with which the accused has been charged. That burden never shifts to the accused. There is no obligation whatsoever on the accused to prove any fact or issue that is in dispute before you. It is not for the accused to prove his innocence but for the Crown to prove his guilt in relation to each offence and to prove it beyond reasonable doubt. The Crown does not have to prove, however, every single fact in the case beyond reasonable doubt. The Crown's onus is to prove, beyond a reasonable doubt, the elements of the charges against the accused. I shall shortly outline for you the elements of each charge that the Crown must prove beyond reasonable doubt.

At the end of your consideration of the evidence in the trial and the submissions made to you by the parties you must ask yourself: 'Has the Crown proved the guilt of the accused beyond reasonable doubt?' That is the high standard of proof that the Crown must achieve before you can convict the accused. If the answer is "Yes", the appropriate verdict is "Guilty". If the answer is "No", the verdict must be "Not guilty".

You will be required to deliver six different verdicts. Whether your verdicts are "guilty" or "not guilty", they ought to be unanimous. That does not mean each of you must agree upon the same reasons for your verdicts. You may individually rely upon different parts of the evidence or place a different emphasis upon parts of the evidence. However you arrive at your decision of "guilty" or "not guilty", it must be the decision of all of you, unanimously, before it can become your verdict.

In this trial, the Crown must prove each element of each of the six counts against the accused. I will now describe these elements or essential facts that have to be proved by the Crown beyond reasonable doubt.

Counts 1, 2, 3 and 5 – act of indecency

For these four counts, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in each count;
2. The accused committed an act of indecency;
3. That the act was indecent; and
4. At the relevant times, Simon Rutter, Justin McCutcheon and Timothy Lyons were under the age of 16 years.

Indecent: The word indecent means contrary to the standards of ordinary and, therefore, respectable people in this community. For an act to be indecent it must have a sexual connotation or overtone. It is for you to determine the standards prevailing in our community. The law provides that a child under 16 years of age cannot consent to an act of indecency.

Count 4 – sexual intercourse against a child above the age of 10 years and under the age of 14 years

For this count, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;
2. The accused had sexual intercourse with Justin McCutcheon; and
3. At the time Justin McCutcheon was above the age of 10 years and under the age of 16 years.

Sexual intercourse: “sexual intercourse” is defined under our Crimes Act to include “sexual connection occasioned by the introduction of any [part](#) of the penis of a [person](#) into the mouth of another [person](#)”. The law provides that a person under the age of 16 years old does not have the capacity to consent to sexual intercourse.

Count 6 – sexual intercourse against a child above the age of 10 years and under the age of 14 years

For this count, the Crown must prove the following elements beyond reasonable doubt:

1. That at the time and place alleged in the count;
2. The accused had sexual intercourse with Timothy Lyons; and
3. At the time Timothy Lyons was above the age of 10 years and under the age of 16 years.

Sexual intercourse: “sexual intercourse” is defined under our Crimes Act to include “sexual connection occasioned by the penetration, to any extent, of the anus of any person by any part of the body of another person.” The law provides that a person under the age of 16 years does not have the capacity to consent to sexual intercourse.

There are six counts in this trial which are being heard together as a matter of convenience. Subject to a later direction about “tendency evidence”, which I will explain to you shortly, you must consider each count separately and return a separate verdict of guilty or not guilty on each of the counts. This means that you are entitled to bring in verdicts of guilty on some counts and not guilty on other counts if there is a good reason in the evidence for that outcome.

If you accept beyond reasonable doubt that what each complainant said actually occurred then the case on each count would be proved. But if you have a reasonable doubt concerning the truthfulness or reliability of the evidence of one of the complainants in relation to one or more of the counts pertaining to them, that doubt must be taken into account in assessing the truthfulness or reliability of that complainant’s evidence generally. If you were to find the accused not guilty on one count, particularly if that was because you had doubts about the reliability of one of the complainant’s evidence, you would have to consider whether or how that conclusion affected your consideration of the remaining counts involving that particular

complainant. While you must consider each count separately, there must be a logical consistency in your verdicts.

If you find Simon's evidence to be truthful, it is sufficient to satisfy the elements of act of indecency that the prosecution must prove beyond reasonable doubt for Count 1. However, it is up to you to decide whether the evidence Simon gave in cross-examination and the inconsistencies in his evidence are sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

If you find Justin's evidence to be truthful, it is sufficient to satisfy the elements of act of indecency that the prosecution has to prove beyond reasonable doubt for Counts 2 and 3.

In addition, if you find Justin's evidence to be truthful, it is sufficient to satisfy the elements of sexual intercourse against a child above the age of 10 years and under the age of 14 years that the prosecution has to prove beyond reasonable doubt for Count 4. However, it is up to you to decide whether the evidence given by Justin in cross-examination and the inconsistencies in his evidence, are sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

If you find Timothy's evidence to be truthful, it is sufficient to satisfy the elements of act of indecency that the prosecution has to prove beyond reasonable doubt for Count 5.

Finally, if you find Timothy's evidence to be truthful, it is sufficient to satisfy the elements of sexual intercourse against a child above the age of 10 years and under the age of 14 years that the prosecution has to prove beyond reasonable doubt for Count 6. However, it is up to you to decide whether the evidence given by Timothy in cross-examination and the inconsistencies in his evidence, are sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

Based on the cross-examination of all the other prosecution witnesses, including Mrs McCutcheon, Aaron Sorkin and Mrs Samuels, it is also up to you to decide whether their evidence in cross-examination and any inconsistencies in their evidence, are sufficient to undermine their credibility as witnesses such that you decide that some or all of their evidence is unreliable.

Finally, the evidence of the accused should be treated in same way as that of other witnesses. This means that it is up to you to decide whether the evidence Mr Booth gave in cross-examination is sufficient to undermine his credibility as a witness such that you decide that some or all of his evidence is unreliable.

However, just because the accused testifies does not mean that the onus of proof shifts to the accused. Even if you reject the evidence of the accused, you must still be satisfied that the prosecution has established beyond reasonable doubt that the accused has committed each element of the six charges against him.

As you are aware, the accused is charged only with the offences stated in the indictment. Generally, juries may only consider the evidence that is directly related to each charge against the accused.

In certain cases the Crown may be allowed to add to the evidence it calls to prove guilt of a particular charge. The law calls this added evidence, "tendency evidence". Here, the Crown

can assert, in proof of any (or all) of the charges, a pattern of behaviour revealing that the accused had a tendency to act in a particular way. Here, the Crown asserts that at the time of the alleged offences, the accused:

1. had a tendency to have a sexual interest in young boys under the age of 12;
2. had a tendency to engage in sexual activities with young boys under the age of 12;
3. had a tendency to use his position of authority as a soccer coach to gain access to young boys under the age of 12 so that he could engage in sexual activity with them.

The Crown asserts that the tendencies of the accused are drawn from the evidence related to all the charges. These include:

1. using his position as soccer coach to befriend the parents of young boys and gain their trust;
2. offering to act as a surrogate father to young boys whose fathers were absent;
3. offering to take young boys on outings;
4. grabbing or touching young boys in the crotch;
5. taking young boys to his house, alone; and
6. getting young boys naked while they were in his home alone.

How can you use this evidence? The Crown relies upon this evidence to prove beyond reasonable doubt that the accused had a sexual interest in each complainant and was willing to act upon it in the way that each complainant alleges. The Crown argues that you will find the accused's sexual interest proved beyond reasonable doubt and therefore you can use it to prove the allegations in the indictment beyond reasonable doubt. First of all, the Crown says that you will be satisfied that the accused had a sexual interest in Simon Rutter on the basis of the acts of a sexual nature committed against Simon Rutter.

The Crown also says that you will be satisfied that the accused had a sexual interest in Justin McCutcheon on the basis of the acts of a sexual nature committed against Justin McCutcheon.

Finally, the Crown says that you will be satisfied that the accused had a sexual interest in Timothy Lyons on the basis of the acts of a sexual nature committed against Timothy Lyons.

Before you can use the evidence of these other acts of a sexual nature in the way the Crown asks you to do so, you must make two findings beyond reasonable doubt. The first is that you must be satisfied beyond reasonable doubt that the sexual act against Simon Rutter occurred. In making that finding, you consider all of Simon's evidence and ask yourself whether you are satisfied that the particular act took place.

If you cannot find that this act described by Simon is proved by the Crown beyond reasonable doubt, then you must put aside any suggestion that the accused had a sexual interest in Simon

Rutter, and decide the case on rest of the evidence. If you are satisfied beyond reasonable doubt that this sexual act against Simon occurred, then you go on to consider the second finding. You ask yourself whether, from the act that you have found proved, you can infer or conclude beyond reasonable doubt that the accused had a sexual interest in Simon Rutter.

If you cannot draw that inference or conclusion beyond reasonable doubt, you must put aside any suggestion that the accused had a sexual interest in Simon Rutter.

If you decide that the sexual act committed *against Simon Rutter* is proved beyond reasonable doubt and you can infer or conclude beyond reasonable doubt that the accused had a sexual interest in Simon Rutter, you may use that fact in determining whether the accused committed the offences against Justin McCutcheon and Timothy Lyons.

Similarly, if you decide that one or more of the acts *against Justin McCutcheon* is proved beyond reasonable doubt and you can infer or conclude beyond reasonable doubt that the accused had a sexual interest in Justin McCutcheon, you may use that fact in determining whether the accused committed the offences against Simon Rutter and Timothy Lyons.

And finally, if you decide that one or more of the acts *against Timothy Lyons* is proved beyond reasonable doubt and you can infer or conclude beyond reasonable doubt that the accused had a sexual interest in Timothy Lyons, you may use that fact in determining whether the accused committed the offences against Simon Rutter or Justin McCutcheon.

The evidence must not be used in any other way. It would be completely wrong to reason that, because the accused has committed one offence or is guilty of one piece of misconduct, he is therefore generally a person of bad character and for that reason must have committed all the offences charged. That is not the purpose of the evidence being placed before you and you must not reason in that way.

You cannot use the tendency evidence in any way prejudicial to the accused unless you accept the Crown's argument that it shows that the accused had a sexual interest in, for example, Simon Rutter which therefore makes it more likely that the accused committed the other offences charged against him, that is those involving Justin McCutcheon and Timothy Lyons.

Remember that you are required to find that the elements of each specific charge are proved beyond reasonable doubt before you can find the accused guilty of that charge.

The delay in making a complaint is a matter that you may take into account in assessing the credibility of each of the complainant's evidence. For all three complainants, there was delay of decades between the alleged incidents and the matters being reported to police.

However I am required by law to direct you that a delay in complaint, even a long delay, does not necessarily mean that a complaint is false. There may be good reasons why a person who alleges they have been sexually assaulted may fail to report, to delay in reporting, such an offence.

From your own knowledge of the world, you might think that there could be cases where embarrassment, guilt or worry about the reactions of family or friends might cause a person who alleges sexual abuse to suppress what has taken place.

In this trial, there may be good reasons from the evidence why, if the complainants were sexually abused in their youth, as alleged, they might maintain silence for a long time. It is for you to evaluate such considerations. You should also take into account the fact that the accused has denied each of the offences.

There is a further warning I should give you relating to the delays in complaint. It is most important that you appreciate the effects of delay on the ability of the accused to defend himself by testing the prosecution's evidence or bringing forward evidence in his own case, to establish a reasonable doubt about his guilt.

One of the effects is the inability to properly inspect Mr Booth's former residences to determine the presence or absence of a swimming pool about which Simon Rutter gave evidence and the layout of Mr Booth's bedroom about which Timothy Lyons gave evidence. There is also the inability to determine the existence of another house or apartment that Mr Booth may have used. These difficulties put the accused at a significant disadvantage in testing the prosecution evidence, or in bringing forward evidence to establish a reasonable doubt about his guilt, or both.

Had the allegations been brought to light and the prosecution commenced much sooner, each complainant's memory for details would have been clearer, such as what dates they joined the soccer team. This may have enabled their evidence to be checked against independent sources so as to verify it, or disprove it. The complainants' inability to recall precise details of all the alleged incidents makes it difficult for the accused to throw doubt on their evidence by pointing to the circumstances which may contradict them.

Another aspect of the accused's disadvantage is that if he had learned of the allegations earlier he may have been able to find more witnesses or items of evidence that might have contradicted the complainants or supported his case, or both. He may have been able to recall with some precision what he was doing and where he was at particular times on particular dates and to bring forward evidence to support him which could have been used in cross-examination of the complainants.

As a result, I warn you that before you convict the accused you must give the prosecution case the most careful scrutiny and bear in mind the matters I have just been speaking about—the fact that each complainant's evidence has not been tested to the extent that it otherwise might have been and the diminished ability of the accused to bring forward evidence to challenge it, or to support his defence.

However, I am not telling you that these problems for the accused make it impossible for the prosecution to prove its case. If, after carefully considering my warnings and scrutinising the complainants' evidence with great care (in the context of all the other evidence) you are satisfied of the truth and accuracy of the complainants' evidence and you are satisfied beyond reasonable doubt that any of the alleged offences did occur, then it is your duty to return a verdict of guilty in relation to that matter.

I have now completed my summing-up. With a final reminder that the verdicts you reach must be unanimous, I ask you to retire to the jury room to consider your verdicts on all eight counts.

