Mock Juries, Real Trials: How to Solve (some) Problems with Jury Science

Abstract

Jury science is fraught with difficulty. Since legal and institutional hurdles render it all but impossible to study live criminal jury deliberation, researchers make use of various indirect methods to evaluate jury performance. But each of these methods are open to methodological criticism and, strikingly, some of the highest-profile jury research programmes in recent years have reached opposing conclusions. Uncertainty about jury performance is an obstacle for legal reform—ongoing debates about the 'justice gap' for complainants of sexual offences has rendered these problems stark. This paper proposes a way to advance the debate.

Forthcoming in the Journal of Law and Society; please cite published version when available!

1. Introduction

In the context of discussing problems facing complainants of serious sexual offences, the 2023 Law Commission *Evidence in Sexual Offences Prosecutions* Consultation contains a striking epistemological observation:

> [I]n spite of the vital role that juries play, and the thousands of jury trials that occur, *we know very little from real jurors about how juries experience a trial, interpret and weigh evidence, and deliberate to reach a verdict*. [At 2.38, p.54]

This statement, at least if we consider jury *deliberation*, is broadly accurate. The attempt to study criminal juries in a scientific manner is fraught with difficulty. In many jurisdictions, there are legal prohibitions that impact certain types of jury research.[1] With almost no exceptions, there are either insuperable legal or institutional hurdles that prevent researchers from accessing the content of live jury

[1] For example, in the United Kingdom see S.20D of the 1974 Juries Act (for England and Wales) and S.8 of the 1981 Contempt of Court Act (for Scotland and Northern Ireland). For another example, see Canada S.649 of the Criminal Code in Canada. Thomas 2013 provides fuller discussion of the relationship between legal prohibitions and jury research.

deliberations.[2,3] Transcribing or recording live jury deliberations for research purposes is widely believed to be an unacceptable infringement on the sanctity of the jury, with the privacy of what is said in jury room being regarded as paramount. The result is what has been called a 'jury-shaped hole' at the centre of empirical discussions of the criminal trial.[4]

This methodological point is a real problem for legal policy and reform. Debate over the quality of the jury on various metrics is a question of perennial controversy. Are juries racially biased?[5] Do juries understand complicated statistical or financial evidence?[6] To what extent do juries understand legal directions?[7] How does gender affect deliberation? Do juries fall prey to rape myths?[8] Does deliberation degrade the reliability of individual judgment?[9] Although we have evidence bearing on these questions, much of it, as we will see, does not look at the interaction between these issues and the distinguishing features of jury trial: decision through protracted collective deliberation following a trial. The harder it is to study jury deliberation in a rigorous and scientific way, the harder it is to reliably answer questions about jury performance. In turn, it becomes harder to know what reforms to trial procedure will effectively ameliorate problems with juries, or even to know if such reforms are necessary in the first place. Moreover, the inability to 'put to bed' concerns with jury performance creates the risk of engendering wider scepticism about the institution of jury trials. Jury trials provide a valuable mechanism for public engagement in criminal justice, providing a link between courts and the communities they serve, and even constituting a counterbalance to the power of the state. It would be deeply unfortunate if mistrust about the jury were to take root without it being warranted. The general pattern of empirical uncertainty and the shaking of public confidence can be found in one of the most pressing issues of criminal justice policy debate, an area where criticisms of the jury have been prominent, persistent and trenchant—namely, the prosecution and management of trials concerning serious sexual offences.[10]

By way of response to these methodological issues, some argue that researchers should be allowed to access *live* jury deliberations—seeing, hearing, or reading what real juries say as they decide on real criminal trials.[11] Proponents of live jury research

[2] Horan and Israel 2016 discuss institutional barriers.

[3] One exception is found in a programme of study that was permitted to access live jury deliberation in civil proceedings in Arizona, for which see Vidmar et al. 2003.

[4] This term appears in Horan and Israel (2016) and is repeated in Tinsley, Baylis, and Young 2021; Ross 2023a.

[5] For evidence from the US, see Sommers 2007; Anwar, Bayer, and Hjalmarsson 2012. In England and Wales, see, for example, The Lammy Review, or Thomas 2007; Thomas 2017.

[6] E.g. see Monaghan 2018.

[7] E.g. see Thomas 2010; Thomas 2013; Thomas 2020.

[8] E.g. see Leverick 2020. I return to this example at some length later.

[9] Hedden 2017 provides discussion.

[10] For example, see Slater 2023 for a call to abolish juries in serious sexual offence trials.

[11] Ross 2023a defends this idea in general. The idea that jury deliberations should be recorded has been advanced in non-research contexts too. For example, it has been advanced as a way to safeguard against racially motivated jurors (see Daly and Pattenden 2005).

argue that the objections against live jury research are overstated.[12] However, live jury research remains deeply controversial and, in light of institutional inertia, may not provide a short-term way to make progress in ongoing methodological disputes about jury science.

In this paper, I outline a less radical alternative that solves many of the problems that afflict the attempt to study jury deliberation. The proposal is that 'mock' juries be permitted to watch and deliberate about *real* criminal trials. This *mock jury, real trial* methodology has some (historic) antecedents that speak to its feasibility, is underutilised, and represents an excellent opportunity to make progress in filling the epistemological lacuna that the Law Commission identifies regarding jury deliberation.

2. Jury Research without the Jury

Researchers use a variety of methods for studying the criminal jury.[13] The most prominent of these methods include:

- General population surveys about attitudes relevant for criminal trials (e.g. on rape myth acceptance; on the relevance of 'bad character' evidence; on attitudes about punishment severity; and so on).
- Quantitative analysis of data about criminal trials: e.g., relating to conviction rates, jury composition, hung jury incidence, deliberation time, etc.
- Surveys of real jurors following a criminal trial ('**post-trial surveys**').
- Experiment participants discussing simulated trials ('**mock juries**').

For certain questions, these methods are entirely appropriate. If we wish to know about a juror's subjective experience of the trial process, post-trial surveys are an excellent method. If we wish to know about (say) differences in conviction rates for various crimes, quantitative analysis is beyond reproach. If we want to isolate and change certain variables (say, different jury gender compositions) to see what effect it has in different situations, then highly reproducible mock jury studies have an important role to play. Contrary to the pessimism of the Law Commission, researchers have learned a great deal about the jury using each of these empirical methods.

However, given the impossibility for researchers to access the jury room, these methods must also be used as the main source of conclusions about jury *deliberation*. It is here that we find ourselves on much shakier ground. The first two methods—general attitude surveys and quantitative data on trial outcomes—provide only the most tenuous information about deliberation. Neither attitude surveys nor

[12] Ross 2023a attempts to defuse various objections to live research: e.g., that it would constitute unfair interference in the right to a fair trial; that it would put jurors at risk; that it would lack ecological validity; that it would undermine the political value of jury secrecy.

[13] For methodological discussion of the advantages and weaknesses of different research methods, see Saks 1997; Chalmers and Leverick 2016; Bornstein et al. 2017.

quantitative data allow researchers to see how deliberation affects jury performance. For example, quantitative data might reveal something about variable conviction rates, the effect of racial composition of panels on verdicts, or length of deliberation between crime-types. But these findings enable only conjecture about what interpersonal deliberative factors explain such results. Attitude surveys might speak to the sorts of views that a randomly selected panel might have when they go *into* the jury room, but they do not reveal whether these views are hardened or undermined in the fire of collective deliberation. Quite simply, these methods do not speak to the contextual pressures and epistemic opportunities that juries encounter when deliberating about real cases.

Let us turn to the third and fourth research methods—post-trial surveys and mock juries. Each of these research methods yield much more targeted evidence about real juries. But what do they tell us about jury deliberation?

Post-trial surveys can shed a great deal of light on subjective aspects of a juror's personal experience of jury service. They are an invaluable way to study certain elements of juror experience, such as juror satisfaction rates or perceived comprehension of evidence or legal directions.[14] However, such methods do not provide objective data about deliberation and experimenters are often limited in what they can ask about deliberation. But even if researchers had free rein to inquire about deliberation, self-report data has considerable limitations. The most obvious issue concerns memory. Moreover, interacting with the fallibility of memory, is the fact that self-report data is susceptible to various presentational biases. For example, post-trial surveys are vulnerable to 'experimenter effects' in which participants provide answers that they believe are socially desirable rather than truly honest.[15] As Bornstein et al. (2017: 210) put it, the post-trial survey method tells us "How jurors *think* they make their decisions, which is not necessarily the same thing [as how they actually do], although it is interesting in its own right". As such, post-trial surveys are not a particularly good way to study how juries make up their minds through deliberation during criminal trials. Although there are high-profile exceptions, post-trial surveys have not been the most common way for UK-based researchers to address controversies about the jury system.[16] One reason is that conducting post-trial surveys requires a considerable degree of institutional buy-in to facilitate researcher access to real jurors, even if only as a logistical rather than legal matter.

Mock juries are a dominant method for studying juries. There is a great deal of variation in how mock jury studies are conducted. The majority of such studies create a simulation far from the experience of participating in a real trial. Of crucial importance for our interest in jury deliberation, most mock jury studies contain no

[14] When combined with field experiments, these surveys can be especially powerful. E.g. see Mott 2003 or Thomas 2020.

[15] See Chalmers, Leverick, and Munro 2021a, Daly et al. 2023 on one side, and Thomas 2021 on the other, for sharply differing views.

[16] For some exceptions, see Zander and Henderson 1993 and the various works of Cheryl Thomas cited in the bibliography.

deliberative element (and indeed participants often do not meet other experimental subjects). Many do not contain any element of trial re-enactment, simply asking participants to read a written vignette, while other studies may ask participants to watch a short video.[17] Time for reflection is often limited. The experience of the typical mock juror involves no courtroom, no trial, nor any interaction with other jurors. This raises serious questions about the external validity of mock jury studies—whether the findings of such studies are a reliable guide to how people react when they attend a courtroom, experience a lengthy criminal trial, and then deliberate together to reach a verdict.[18] Yet, mock jury trials do not need to be so unrealistic.[19] The ideal mock jury study would include a thorough re-enactment of a criminal trial: judges would be asked to preside over an ersatz trial in a real courtroom, with the different roles played by actors or, better still, by legal practitioners and other professionals assisting with the research. The subjects of the experiment would be given as long as they want to deliberate, doing so as a collective, in a surrounding that closely mimics that of a real juror. Mock jurors would have the chance to submit questions to the judge, just as they would in a real trial. Such a process would take multiple days, with actors being retained throughout. All the while, courtroom premises would be used to ensure a maximally realistic experience.

Some barriers to such high-quality mock jury research are economic and logistical: such experiments are extremely costly and take a huge amount of effort to organise. Actors, scripts, courtrooms, and experimental participants are expensive—especially when retained over multiple days for live performance. This expense quickly multiplies if we wish to conduct such mock trials in a way that maximises their scientific usefulness, by (i) having mock juries deliberate about a variety of different types of case, and (ii) having different mock juries deliberate about the same case. From the perspective of the individual researcher, it is often not worthwhile to engage in such realistic mock jury studies when the results of considerably less resource-intensive (and hence less realistic) methods can still be published in academic journals.

The more intractable issue with mock jury research concerns its realism and hence validity. Concerns about the (ir)realism of mock jury studies have been extensively discussed so I will not belabour the issue here.[20] Worries about realism have at least

[17] One long-standing worry with mock juries studies concerns inadequate sampling. Many older studies exclusively rely on undergraduate populations, which may not provide a representative group, especially when what is being investigated engages the moral or political views of participants. Happily, many more recent mock jury studies employ much-improved sampling techniques. For an optimistic discussion of sampling issues, see Bornstein et al. 2017.

[18] For general discussion and references, see Bornstein 2017: especially at 217; Saks 1997: especially at 7 and Wiener, Krauss, and Lieberman 2011. Worries about validity are heightened by concerns, within recent decades, that areas of social psychology have been undergoing a 'replication crisis' that threatens many research findings. For discussion see, for example, Wiggins and Christopherson 2019.

[19] Thomas 2007 & 2010; Ormston et al. 2019; and a series of studies by Vanessa Munro and collaborators are among the highest-quality recent examples in the UK. Given their focus on sexual criminality, the latter two are discussed in more depth at pp. 7-8.

[20] See Ross 2023a for an article-length discussion.

two sides. The first is the challenge of generating realistic scripts[21] and, more importantly, having them convincingly performed by actors. Convincing writing and acting is a difficult business and any slip detracts from the validity of the mock jury study. The second deeper component of the realism issue is that *mock juries are fully aware they are not evaluating a real trial*. Mock jurors are primed to notice the fictional nature of the scenario, they know that there is not really a 'correct' answer about what happened, they are fully aware that they are not in a normal courtroom environment, and they know nobody who appears in front of them experienced a crime or could be punished in any way. Although mock jury studies can attempt to replicate the solemnity of the courtroom experience, it can only be a replication. Even audiences who attend the most prestigious theatres to watch performances by the world's best actors (commanding million-dollar fees) rarely tend to forget that they are not watching real-life. The distance between a researcher-led replication of a trial and the real thing invariably remains large.

Case-Study: Rape Myth Acceptance Among Jurors

Concerns about the methodological credentials of jury science are not merely an abstract issue. They are of huge contemporary import, afflicting some of the most prominent policy debates about criminal justice. In the legal jurisdictions comprising the UK, along with many societies across the world, there is currently a vigorous debate about justice in sexual offence cases. Some allege that there is a 'justice gap' for complainants, with perhaps the most headline-grabbing statistic being a claimed complaint-to-conviction ratio of under 2% for rape accusations.[22,23] Researchers, campaigners, legal practitioners, civil servants and politicians dispute the cause of unsatisfactory management of sexual offences within the criminal justice system.[24] One prominent suggestion lays some blame at the door of the jury room, claiming that an alleged propensity of jurors to acquit in such trials is due to their being influenced by 'rape myths'.[25] In broad terms, rape myths are (often societally prevalent) false

[21] One way to generate realistic scripts is to use materials from real trials. See footnote 44 for discussion of imperfect attempts to condense real trials into short experimental stimuli.

[22] See: https://www.gov.uk/government/publications/end-to-end-rape-review-report-on-findings-and-actions. The relevance of this research for the jury should be read in light of recent analysis of conviction rates in rape cases up to 2021. In the case of non-historic rape of an adult female, the 15-year average conviction rate is 50%, but recent years have seen considerably higher rates (63% in 2019, 72% in 2020, 67% in 2021). See Thomas 2023 for comprehensive analysis.

[23] The 'justice gap' term appears at least as far back as Temkin and Krahé 2008 and is repeated in, inter alia, Brown and Walklate 2012; Chalmers, Leverick, and Munro 2021a.

[24] In addition to the aforementioned 'End to End Rape Review' in England & Wales, see in Scotland the 2021 Scottish Courts and Tribunals Service 'Improving the Management of Sexual Offence Cases' Report ('The Dorrian Review').

[25] The extent to which juries believe rape myths also has ramifications for other aspects of criminal evidence procedure such as the admissibility and proper role of sexual history evidence. For a recent book-length treatment, see Conaghan and Russell 2023.

beliefs that inhibit the proper evaluation of accusations of rape.[26] Canonical rape myths include inaccurate views about the 'proper' behaviour of victims, about the types of people who are susceptible to commit or be victims of sexual crimes, or about the way that sexual consent or dissent must be communicated. Many rape myths serve to downplay or undercut accusations concerning male sexual violence against women, others apply to cases involving same-sex incidents, and yet others pertain to cases with male victims.

Whether jury verdicts are substantially influenced by rape myths is an empirical question, one that we should look to jury science to answer. Unfortunately, policymakers in the UK have found few conclusive answers when looking to empirical research. Two of the highest-profile empirical programmes of jury research have reached what, on the surface, seem like **opposing** conclusions.[27] A programme of post-trial surveys conducted by the UCL Jury Project in England and Wales—one of the only research programmes that has consistently gained access to real jurors in English and Welsh courts—finds that "hardly any jurors believe widespread myths and stereotypes about rape and sexual assault".[28] This evidence has been widely reported on in debates about criminal justice reform in the UK.[29] By contrast, a useful systematic review of a variety of mock jury studies—relied on by the Scottish Government—finds there is 'overwhelming evidence' that jurors believe rape myths.[30] Unfortunately, the quantitative studies canvassed by this systematic review are highly unrealistic. Ross (2023a: 113) summarises these limitations as follows[31]:

> On the relationship between rape myth acceptance and victim blaming in particular instances, 29 studies are cited and 28 of these show a statistically significant relationship between the two. But *none* contained a deliberative element and none used a realistic trial re-enactment. On the relationship between rape myth acceptance and reluctance to convict, 28 studies are cited and 25 suggest a statistically significant relationship between the two. But only two contained a deliberative component, most were not trial re-enactments, and many

[26] Jenkins 2021 provides a fuller discussion of the idea of rape myths. Empirically, susceptibility to rape myths is often studied by asking experimental participants to complete attitude surveys that allow researchers to place them on a 'rape myth acceptance scale'. A number of such scales exist. For review, see Burt 1980; Lonsway and Fitzgerald 1994; Payne, Lonsway, and Fitzgerald 1999; Gerger et al. 2007

[27] Although, see, for example, para 5.38-5.44 of the 'Dorrian Review' mentioned in footnote 23 for a discussion of the results that finds less disagreement than is often reported.

[28] Thomas 2020: 1001. These results are worth contrasting with jury surveys in Tinsley, Baylis, and Young 2021. The UCL Jury Project engaged in quantitative analysis of jury verdicts on rape and sexual offences over a 15-year period with their data calling into question the common assertion that conviction rates for rape are uniquely low. (See footnote 22). However, while this data is suggestive and cuts against the narrative that the use of juries in assessing allegations of serious sexual criminality is problematic, it is not possible to draw robust inferences from quantitative data about whether extant tendencies to acquit are due to rape myths or not.

[29] For, example see here, here, here, or here.

[30] Leverick 2020. See Waiton (2023) for critical discussion of the evidence-base for this claim.

[31] These concerns are only deepened by the fact that social psychology has been undergoing what is often called a 'replication crisis', in which it is claimed that many results fail to replicate or are beset by other methodological problems. See Nosek et al. 2022 for an up-to-date discussion.

> asked participants to answer questions quite unlike those they would be asked at trial (e.g., being asked to return a Likert-scale response rather than being given legal directions as to whether or not they believe the actus reus and mens rea has been proven beyond a reasonable doubt). These types of study simply do not speak to whether rape myths would be undermined or even exaggerated by the real-life conditions of jury deliberation.

The same systematic review also canvasses better-quality qualitative evidence supporting the prevalence of rape myths. A number of mock jury studies canvassed included some effort towards trial re-enactment, especially a landmark series of studies carried out by Vanessa Munro and collaborators which involved acted-out trial simulations of ~75 minutes in duration plus collective deliberation.[32] These qualitative studies have been further supported by a study commissioned by the Scottish Government, which funded a mock jury study that re-enacted various elements of a real trial, shown to mock jurors in an hour-long video, who were then given time to collectively deliberate. This study also yielded qualitative evidence that jurors believe and deliberate about rape myths, contrary to the findings of the post-trial survey conducted by the influential UCL Jury Project.[33] On the basis of this evidence—the unrealistic quantitative data, along with suggestive qualitative results—the Scottish Government has (at the time of writing) introduced controversial legislation to enable a scheme pilot of judge-only trials for serious sexual offences.[34]

The overall position is striking. Two the UK's most prominent teams of jury researchers disagree about the rape myth question and there has been vigorous—and indeed heated—debate between the leaders of the different research programmes about the validity of their respective research programmes.[35] Contentious government policy is being driven by a relatively small body of contested evidence, even as the Law Commission bemoans the limited nature of our knowledge. All the while, damaging headlines about the poor performance of the UK's criminal justice systems pile up. What deepens the unsatisfactory nature of the current situation is that this disagreement about the prevalence of rape myths is in fact nothing new. Over a decade ago, Helen Reece argued in the *Oxford Journal of Legal Studies* that the growing view that rape myths explain poor conviction rates for rape is unfounded.[36] It is a telling fact about the slow and even repetitive nature of the debate that the phrase 'myths

[32] Finch and Munro 2006; Ellison and Munro 2009b; Ellison and Munro 2009a; Ellison and Munro 2010; Ellison and Munro 2015. Also see, Taylor and Joudo 2005.

[33] Chalmers, Leverick, and Munro 2021b. It is notable, however, that in this study rape myths were very often challenged by other members of the mock jury. Although the Scottish Government-commissioned mock jury study was high quality, it was also limited in scope. Only one scripted scenario concerning sexual offences was presented to the mock jury.

[34] The *Victims, Witnesses, and Justice Reform (Scotland) Bill*.

[35] See footnote 15.

[36] See contrasting views in Reece 2013 and Conaghan and Russell 2014.

about myths' appears in the title of distinct academic articles about evidence for juries believing rape myths, ten years apart.[37]

3. Mock Juries, Real Trials

The Law Commission, when acknowledging the deficits in the current state of knowledge, see the argument for further research. They write that:

> [T]here may be a case for opening up the possibility of further research with jurors, including in relation to deliberations, as there are many gaps in what we presently know. [At 2.52, p.60]

I agree that further—and crucially new types of—jury research are needed. I will now argue that there is a natural solution to many of the problems with the current mock jury research paradigm. This proposal aims to occupy an unfilled space that is an improvement on current research, yet without raising concerns that are typically levelled against the more radical call for live deliberation research. Live deliberation research attracts objections that it would: constitute undue interference against the right of the accused to a procedurally sound trial; raise safety or external influence concerns for jurors; or would undermine the privacy of jury deliberations by setting a politically dangerous precedent.[38] Therefore, I offer a proposal that raises none of these concerns.

The proposal is very simple. Namely, mock juries should be facilitated to observe and deliberate on real criminal trials. These 'shadow juries' would observe real and ordinary criminal trials, assessed as usual by a regular primary jury. Ideally, the shadows would observe these trials live and in-person just as does the real jury.[39] The participants in the study would not be the primary jury, who would be entirely unaffected by the research. Rather the experimental shadow juries would be deliberating for the purpose of furnishing policy-makers with gold-standard evidence about jury decision-making. In short, the proposal is: '*mock juries, real trials*'.

Is such research feasible? This question can be answered confidently in the affirmative, for the simple reason that there is successful precedent. Over half a century ago, the UK Home Office funded research with the characteristics I have described, and court officials cooperated with researchers to facilitate it.[40] Shadow juries were selected using the electoral roll and invited into the courtroom to observe

[37] Conaghan and Russell 2014 and Daly et al 2023.
[38] See Ross 2023a for discussion.
[39] As a second-best, trials could be live-streamed or recorded and then shown to the shadow jury. This would, in my view, be considerably suboptimal due to the reduced realism.
[40] See McCabe and Purves 1974.

real trials.[41] Thirty trials of different types were observed by the shadow juries. Mirroring the real jury, the shadow jury would elect a foreperson and then retire to deliberate after hearing the evidence, before issuing a verdict. Their deliberations were recorded then transcribed, before being anonymised and discussed in a research pamphlet published by the Oxford University Penal Research Unit (which has since been renamed the Centre for Criminology).

This research was methodologically pathbreaking. The foreword to the pamphlet by Roger Hood writes, in a sentiment I fully agree with, that "[w]hile no-one would claim that this modest study does more than begin to illuminate an area of decision making that is vital in the administration of justice, it does indicate that research on these lines on a more ambitious scale is likely to produce data which is essential to any discussion of the reform of the jury system."[42] Unfortunately, such ambition was not matched by further institutional support in the UK for additional studies along these lines.

Yet, in the late 1970s, just a few years after the Oxford study, a twelve-case study was published in the *Stanford Law Review,* reporting shadow jury research conducted at the United States District Court for the Northern District of Illinois.[43] The purpose was to investigate whether jurors excluded during *voir dire* return different verdicts from the empanelled jury. *Two* shadow juries observed each case, live in the courtroom, along with the primary jury. During the trial, both primary and shadow juries were treated the same, receiving the same evidence, leaving and returning to the court at the same time, and having the same access to the judge with respect to clarificatory questions. Deliberation of the shadow jury was again recorded. Another shadow jury experiment that exposed participants to a real trial has since been conducted, much more recently, in Korea.[44] In that experiment, shadow juries also watched live trials before retiring to deliberate. The shadow jury deliberation was again videotaped for research purposes. A strength of both the Illinois and Korean studies, unlike the more informal Oxford study, is that their methodologies yielded quantitative data and demonstrate how shadow jury research might provide statistically significant results.

The take-away point for empirical jury research in the UK is the feasibility of the methodology. This precedent research provides a clear proof-of-concept demonstrating the viability of the *mock jury, real trial* paradigm.[45] Indeed, in some

[41] Consultancy firms in the US offer 'shadow jury' services to high-paying clients who desire tactical advice on how to convince a real jury. These vary in their realism, with some 'shadow jury' services being mere rehearsals of trial strategy to lay panels outside of the courtroom context. Some consultants even offer to put on a mock trial with a shadow jury, along with post-trial interviews, for eye-watering sums. For example, see Vinson 1982; Patterson and Spencer 2019.

[42] McCabe and Purves 1974: 2.

[43] Zeisel and Diamond 1978.

[44] Lee et al. 2013.

[45] Studies can also blend the 'mock jury' and 'shadow jury' methods, by deriving the materials used in a regular 'mock jury' experiment from a real trial. One such study is Daftary-Kapur et al. 2014, which generated research materials from the experimenter's notes from a real trial. However, this is not a

ways the research would be easier to conduct now than half a century ago, given new technological possibilities.

Resuscitating the 'mock jury, real trial' methodology would solve or at least substantially remedy two chief problems afflicting mock jury research. Firstly, it would alleviate some economic and logistical barriers that make high-quality mock jury research prohibitive. Secondly, and more importantly, it would address basic issues concerning the realism of the participant experience. Criminal trials occur naturally, every day, ranging over a great variety of crimes, factual scenarios, and with a bewildering range of persons as accused, victims, and witnesses. By using real trials as the stimuli for mock juries, there would be no need for multiple scripts, actors, retired judges, or faux-courtrooms. The real world would furnish researchers with all the variety in case-types that they could ever need. There is no better guarantee than reality itself that the cases reflect real-world conditions. Similarly, there would no concern that suboptimalities in the acting or simulated nature of the court proceedings detract from the realism of the experimental stimulus—there would be no artificially truncated mock trials, scripted witness evidence, or attempts to mimic a genuine courtroom. Every element of the experience would be just as is experienced by the primary jury. As such, there would be no concern that jurors being aware of the fictional nature of the case would influence their deliberation or lead them to attempt to 'second guess' the experimenter's purpose in devising the mock trial. Nor would mock jurors be influenced by the fact that there is no 'right answer' about what happened, since they would be deliberating about an actual event within a real trial—one with genuine victims and real people facing serious punishment.

In addition to addressing the realism problem and (some) concerns regarding economy and logistics, the *mock jury, real trial* proposal has an important advantage over observing live jury deliberation. One of the great advantages of mock juries is that they allow precise testing of specific hypotheses. Panel composition is a good example. By using mock juries, researchers can have multiple differently composed panels consider the same experimental stimulus. To take a concrete example, suppose you were interested in what has been a common topic of empirical research—whether gender composition of juries affects trial verdicts.[46] The use of mock juries allows researchers to artificially change the gender composition of different panels who are considering the exact same material—the same simulated trial—to test for differences in their response. The mock jury, real trial proposal retains this advantage. Researchers would be free to vary the composition of different shadow panels watching the same case. Other types of intervention would also be possible. For

genuine shadow jury experiment; such online, text-based studies are crucially different from exposing participants to the stimulus of a real courtroom, even if the base materials are veridical. Another interesting example is Taylor and Joudo 2005, which created a short trial re-enactment using veridical base materials. Unfortunately, this study failed to record the deliberation of the participants. Another related effort in England and Wales involved a portion of a serious fraud trial—the initial briefing of the judge and the opening statements of defence and prosecution—being recreated for jurors in the form of a video recording, see Honess 1998.

[46] E.g. see Thomas 2007, 2010 for research relating to jurors in England and Wales.

instance, some panels could remain in situ and others not for certain judicial instructions or legal discussions or cross-examinations, some could receive materials on avoiding rape myths and others not, and so forth, to see how the presence or absence of these affects the eventual deliberation of the panel. This would be highly useful for the ongoing debate on sexual prosecutions. Thus, one of the key benefits of mock jury research—the targeted testing of certain hypotheses—could be retained alongside a vastly superior experimental stimulus.[47]

Finally, another advantage of my proposal is that it expands the range of issues that a mock jury study can speak to. A perennial question about jury performance is how jurors are affected by media reporting of criminal suspects and (to a lesser extent) of victims. This cannot be tested for in a regular mock jury study, since the stimulus is a piece of fiction—there is simply no media reporting to influence the deliberation. By contrast, if we have mock juries discuss real cases, it becomes possible to gauge their susceptibility to media reporting happening in the real-world.

4. Five Criticisms and Replies: Elaborating the Proposal

4.1 Self-selection Bias

A criticism of mock jury studies is that they are unrepresentative. This was certainly true in many early mock jury studies, which often used exclusively student populations. Most mock studies now go to much greater lengths to ensure a demographically representative panel. However, even demographically representative panels can be criticised concerning validity. Cheryl Thomas has forcefully made the argument that all mock juries fall prey to a 'self-selection' bias—by their nature, they only include the sorts of people with a prior interest in taking part in jury-like trials. Thomas claims that this makes participants in mock juries unlike real jurors, as she adduces evidence that 87% of real jurors—prior to the experience of serving on a real jury—would rather not have been called on to do jury service.[48]

[47] One criticism of the proposed research is that it would lack statistical power. If only one or two mock panels observe a given live trial, this limits the sample size from which to infer. This is an area in which the proposed research might fare disfavourably compared to regular mock jury studies, where any number of participants can be asked to consider the same artificial stimulus by iterating the number of mock trials. Some may regard this as an important argument against the proposed research being conducive to yielding decisive evidential advances. But while it may be impossible to get statistically robust results concerning juror reactions to individual trials, one may draw inferences about the propensity of jurors to rely on rape myths in rape trials generally. One way to test the effect of deliberation specifically would be to canvass mock juror opinions of the case—including the applicability of rape myths—pre-deliberation and then compare the same juror's views post-deliberation. I leave further details to empirical researchers. Notwithstanding these methodological questions, many highly cited and influential studies on rape myths have focused only—or chiefly—on qualitative data. In this respect, the proposed research would be on par with other influential types of research that have motivated the ongoing debate and superior to them with respect to the issues of realism extensively discussed. More broadly, the proposed research will provide just one plank of multi-method work used to address issues surrounding juror evaluation of sexual assault trials.

[48] Thomas 2020: 1006.

Even if this criticism is forceful, various versions of my proposal avoid it. Namely, jurors who are called for regular jury service could be asked to take part in the proposed research. Participation could occur either before, instead of, or after serving on their first (regular) trial. This would ensure a representative panel, immune from self-selection bias. One concrete possibility would be to use jurors who are called to serve but subsequently dismissed. Exactly this approach has been successfully adopted in a range of studies, with high voluntary participation rates.[49] Another possibility is to have jurors serve on multiple cases, once as a primary juror and once as a shadow. The provisions for jury service in England and Wales already allow for jurors serving on two trials, with the current instructions stating that: "if the trial is shorter than 10 days, you may be asked to be a juror on other trials." As Michael Zander's Crown Court study indicated, jurors being asked to serve on multiple cases is not especially uncommon.[50] Serving on *multiple* trials may only work if the proposed research was only restricted to short (and hence less serious) cases. In any case, there are different options here. The basic point is that issues of selection bias can be readily overcome by sourcing participants from real jury pools.[51]

4.2 Disagreement between Primary and Experimental Juries

Every jury is idiosyncratic. It is unreasonable to suppose that every logically possible panel of jurors would return the very same decision. Indeed, it is always possible that a given criminal jury returns a verdict that departs from what most other panels would have said after considering the same evidence. The acceptance of this eventuality is contained within the wide discretion allotted to criminal fact-finders and the reluctance of appellate courts to overturn decision simply on the grounds of disagreement.[52] Of course, when there is only one panel deliberating, the idiosyncrasy of each jury is maintained as an 'opaque' truth. The spectre of disagreement between merely possible juries is, at present, rarely salient. But simply because the idiosyncrasy of juries is hidden from view does not mean that we do not already accept it as a routine

[49] Exactly this approach was taken in the proof-of-concept studies mentioned earlier; both the Illinois and Korean study use jurors who had been called for service using the normal procedure but had not been selected to serve. The use of dismissed jurors has also been successfully used in mock jury research by Cheryl Thomas on the impact of race on jury verdicts, e.g. Thomas 2007; 2010.

[50] Zander and Henderson 1993: 226-228.

[51] If the proposed research were to have institutional support, it may be hoped that the same system for compensating primary jurors could be applied to shadow jurors. It would, of course, be a much more radical suggestion to argue that those who have been called but dismissed from primary jury service could be mandated to take part in shadow jury research.

[52] Counterfactual disagreement between appellate panels is even more obvious, since cases can be determined by the agreement or dissent of a single tie-breaking judge. It could have been the case that a different judge with different views was assigned to hear the case. In this sense, even decisions in highly important appellate cases are sometimes not modally robust. In this vein, Sunstein (2003) provides empirical evidence (from the US) that appellate court decisions are influenced by panel composition, notably concerning political affiliation of the judge and their appointees. See Hanretty (2020) for a less pessimistic take on judicial behaviour in the UK Supreme Court.

part of criminal justice. We do not, after all, try to cross-reference the decision of one jury with other juries.

However, the possibility of disagreement between the 'primary' jury and the experimental shadow juries could be used to object to the proposed research.[53] There are different versions of this objection. First, there is a general worry about a negative effect on public confidence if it were to become known that primary juries return decisions at odds with shadow juries evaluating the case in tandem. A second objection is more specific and concerns the fact that the proposed mock jury research might identify possible miscarriages of justice—for example, if all shadow jurors voted to acquit a person who was convicted by the primary jury. These types of objections are familiar, with similar ideas being raised as refutations of live deliberation research.[54]

The most combative response to the public confidence issue is that uncovering the extent of variability between different jury panels is deeply important for understanding criminal justice. The value of this finding, concerning the fundamentals of jury decision-making, is large. And so, we might think that the importance of such research outweighs countervailing concerns about how it is perceived. This simple response may not convince. So, I will engage with worries about public confidence on their own terms. A first response is that research already exists speaking to discrepancies between the decision-making of fact-finders: namely, juries versus judges.[55] This research has been assimilated and has not led to large-scale loss of faith in criminal justice. My second response is to highlight the elitism in the idea that public confidence in the criminal justice system should be secured by maintaining ignorance about the nature of the system.[56] Faith in public institutions should be based on a realistic assessment of their performance, not misconceptions. My third response is that the status quo is in fact already undermining confidence in criminal justice, with the media regularly running stories about the failures of the courts to deliver justice in sexual offence cases, and practitioners reporting that this lack of confidence is now regularly cited by complainants and putative complainants when deciding on the extent to which they want to commit to and maintain cooperation with investigations. When it comes to maintaining public confidence in criminal justice, all options, including doing nothing, come with risks. There are also threats to public confidence that come from being too proactive without proper evidence. For example, the Scottish Government's proposed 'judge only' pilot for sexual offences has—for better or worse—been criticised by lawyers, judges and media commentators, with one concern being that it represents a tacit commitment to a conviction target for rape.[57] More generally, the idea of using different procedures for some serious cases (i.e. some with juries,

[53] In the Oxford Penal Research Unit Study, the real jury and mock jury disagreed in a quarter of cases.
[54] E.g. see Lee et al. 2013. Also see Zander 2013.
[55] E.g. see Robbennolot 2005 or Lundmark 2010.
[56] Of course, there are more sophisticated defences of the importance of 'ritual' in criminal adjudication that have been long-discussed, see for example Tribe 1971.
[57] For summary, see: https://www.ft.com/content/47cc8c1b-cca3-423b-96a9-e54e26c3a072

some without) risks creating a 'class system' of criminal protections. The proposed research is certainly no more controversial than these other proposals.

The final point is whether variability between primary and shadow panels might be problematic insofar as it could constitute awkward evidence of miscarriage of justice. This may occur if a primary jury opted to convict but there was wide agreement among shadow jurors that the evidence supports acquittal. Legally, it seems doubtful that the disagreement of a shadow panel (or even two) would constitute sufficient reason to regard a conviction as unsafe. Already, majority verdicts—verdicts based on non-unanimous agreement among jurors—are regularly used to underpin criminal convictions notwithstanding the fact that some jurors wish to acquit.[58] Such disagreement among jurors is not a legal basis for overturning a conviction, so one may legitimately argue that disagreement between primary and shadow jury panels should not provide such a basis either. There is, in England and Wales, ground for overturning a conviction when the appeal court holds there to be a 'lurking doubt' about the safety of the conviction.[59] Lurking doubt cases are those in which no new legal argument or evidence is adduced, but the appellant merely claims that the fact-finder was mistaken in their assessment of the evidence at the original trial. However, the burden for establishing this doubt has been described as 'formidable' and it is only "in the most exceptional circumstances that a conviction will be quashed on this ground alone."[60] Again, disagreement among jurors in cases involving majority verdicts is not by itself one of these exceptional circumstances, and nor are disagreements between appellant and the trial jury, so it would seem logical to suppose that disagreement between primary and shadow jury panels would not overcome the formidable bar to establishing a lurking doubt to the legally required standard. Indeed, so high is the bar that even in cases where there are allegations of irregularity among deliberating jurors, there is a general reluctance to quash a conviction on this basis without stronger evidence of impropriety.[61] While such avenues for appeal exist, the legal test for quashing a conviction is demanding. Speaking more generally, those who observe trials—and even those who are convicted—are free to criticise the verdict, maintain the innocence of the accused, and to publish their critical views. In this respect, the status of shadow jurors (or writings about them) would be no different from any other trial observer. It is equally doubtful whether evidence of shadow/primary panel disagreement would concern the Criminal Cases Review Commission (the statutory body concerned with miscarriages of justice) in the event that an accused person made an application to them. CCRC referrals of cases back to the appeal courts are usually restricted to instances where some new evidence or legal

[58] In Scotland, criminal verdicts require only a simple majority of 8 out of 15 jurors. In England and Wales, 10 out of 12 jurors can suffice for a conviction following the Criminal Justice Act 1967, S. 13. Of course, there is room for disagreement about the wisdom of majority verdicts. My point is just that the by-now familiar existence of non-unanimity about verdicts does not currently create a crisis of confidence in criminal adjudication.

[59] *R v Cooper* [1969] 1 All ER 32.

[60] *R v Pope* [2013] 1 Cr. App. R. 14, quote at para 14.

[61] For example, see *R v Mirza* [2004] UKHL 2 or *R v Thompson and Others* [2010] EWCA Crim 1623 Although, see *R v Smith* [2005] UKHL 12.

argument comes to light that was not raised in the original trial, although they have a standing ability to refer 'in exceptional circumstances' any case they see fit.[62] To summarise, there are existing mechanisms for revisiting cases involving disagreement between primary and experimental juries; the proposed research would not require the creation of any new legal mechanism to deal with this eventuality.

The current state of the law imposes a high test on the overturning of convictions. I cannot settle here whether it *would* be appropriate to revisit cases simply if there were disagreement in the manner outlined. Certainly, there is a tension between supposing that guilt has been proven beyond a reasonable doubt and the fact that a shadow jury (or two) has unanimously voted to acquit. It would take us too far afield to try to settle the jurisprudential question of whether this tension is ineradicable. But suppose you took the view that panel disagreement *would* constitute clear evidence that a conviction was unsafe. In this case, the objection to the proposed research falls away. For, it would be an untenable—extremely cynical—argument to say that uncovering potential miscarriages of justice is a defect (rather than a strength) of an otherwise sound proposal. Revealing miscarriages of justice is not a fatal objection to good research.

4.3 Lack of Realism

There is a fundamental objection that can be levelled against any way of studying jury deliberation that does not involve studying the live deliberation of the primary jury. The challenge can be introduced by posing the following question: 'Do people tend to think and deliberate differently when making decisions with serious and far-reaching consequences compared to less serious and impactful decisions?'[63] To the extent that the answer is 'Yes', any information gleaned from shadow juries may not generalise to primary juries making decisions that determine the result of real trials. In a real trial, the decision made by a jury can lead to someone being imprisoned for many years, or to the potential erroneous release of a serious criminal. No such consequence is at stake in a mock jury study, with the weightiest possible result being that one's deliberation is a small difference-maker in the evidence used to inform research and policy.

I entirely agree that this can be problem with any research that falls short of studying live jury deliberation. However, it is not as I see it an objection to other types of research. Rather, it is only an objection to the *exclusive* reliance on other types of research. Using multiple methods to gather evidence—and to calibrate the accuracy of the different methods—is good practice in social-scientific research. The ideal epistemic situation is one in which multiple types of jury research are used in tandem to produce maximally-informed, evidentially robust policy. These points

[62] Criminal Appeal Act 1995, s.13(2).

[63] For example, consider the idea that severity of punishment affects the strength of evidence needed to convict. This idea is supported by both empirical and conceptual arguments: see Bindler and Hjalmarsson 2018 and Ross 2023b, respectively.

notwithstanding, I do believe that we should explore ways to conduct research into live (primary) jury deliberation. But, given the difficulties in conducting such research—and in persuading stakeholders that it should be conducted—we should not allow the best to be an enemy of the good. The current proposal advances jury science and uses an experimental stimulus much closer to the real thing than used by large swathes of previous research. I do not claim that the proposed research solves every problem with jury science, nor that other ways of studying the jury are redundant.

4.4 Permission & Fairness to Complainants

Although trials are by default public and open to observers, the proposed research would ideally operate with the approval and assistance of the trial judge and other relevant judicial authorities. Indeed, overt cooperation would be required for crucial aspects of the realism that made the proof-of-concept studies so effective, such as the ability to submit questions to the judge during shadow deliberation. This raises questions about the extent to which giving approval to the presence of shadow juries might disrupt trial proceedings. There was no evidence from the proof-of-concept studies that the presence of the shadow jury was regarded as disruptive—and various stakeholders in these trials actively assented to their presence. However, these studies did not canvass the most serious criminal trials. Given that the motivating debate for this article is evidence in serious sexual assault trials, it is necessary to discuss the hard case of whether the research would amount to unwanted or deleterious interference in such trials.

Avoiding 'visibility' is often important for complainants giving evidence in sexual assault trials. The most common concern is avoiding being observed by the defendant when giving evidence, rather than other people who may be observing the trial.[64] Measures which protect the complainant from being observed by the defendant but maintain visibility for the shadow jury are entirely feasible (the use of screens is a now familiar part of sexual offence trials). Nevertheless, it would be entirely reasonable for a trial judge to consult with complainants before allowing the presence of a shadow jury while the complainant is giving evidence. Suppose (for sake of argument) that complainants giving evidence in person had a de facto veto power over the presence of a shadow jury. The proposed research would certainly be unfeasible in some sexual assault trials due to the unwillingness of the complainant. If the shadow jury was excluded from the court during crucial complainant evidence, the research would be non-viable. However, for the proposed research to work, only *some* cases need to be found where the research is not unwelcome. So long as this is possible, the research could proceed. There are many sexual assault trials and complainants differ greatly in their characteristics; especially given that one purpose of the research is to see how we can improve the justice system for complainants, I think it quite likely that some will

[64] Majeed-Ariss et al. 2021.

not object to the presence of a shadow jury.[65] Another reason to think we could identify trials in which the presence of a shadow jury is innocuous is the following. There are invariably cases in which permission is given for complainants to give evidence via 'special measures'.[66] Most relevant for the current discussion are cases where evidence-in-chief and cross-examination is pre-recorded or displayed in court by video. Complainants who opt to give evidence videographically, since they bypass many stresses concerning 'visibility', may well find the presence of a shadow jury in their trial less objectionable. There is also no reason to think that evidence given in such a private way would be disrupted by the presence of a shadow jury.[67]

4.5 Cost

A final objection is the proposed research is simply too costly and that compensating shadow jurors (as we surely must) to observe lengthy criminal trials is a bad use of public funds. It is true that to gather evidence about jury deliberation on real trials concerning sexual criminality, a substantial investment of time may be needed on behalf of participants. A recent estimate of the median hearing time for cases at the Crown Court was 12.4 hours, although this figure includes time for preliminary and sentencing hearings.[68] Many serious sexual criminality trials will take longer.

I do not think that it can be seriously maintained that the proposed research would represent a poor use of public funds. Reforms to criminal justice affect our fundamental rights and liberties. Moreover, such reforms tend to stay in place for a

[65] Another possibility to minimise the intrusiveness would be to have the shadow panels be somewhat smaller than the regular primary jury. Having a handful of additional persons in the public gallery need not be especially conspicuous. Of course, this would detract from the validity of the experiment—but, if the aim is to study deliberation about evidence in such cases, it would remain a valuable exercise. This raises further practical questions. For example, where exactly should the shadow jury be seated in the court? The proof-of-concept studies mentioned already discuss such issues, alongside myriad other practical matters. For example, Zeisel and Diamond 1978 seated the shadow in the first row of the public gallery. In any case, it surely cannot be suggested that one of the great debates about criminal justice must go unresolved due to a lack of chairs.

[66] In England and Wales, see ss. 23-30 of The Youth Justice and Criminal Evidence Act 1999 and parts of the Criminal Justice Act 2003. Giving evidence in this way could become the norm in Scotland if the Victims, Witnesses and Justice Reform (Scotland) Bill is adopted, see in particular S.59.

[67] While the proposed research is for shadow juries to observe trials live, there are of course myriad other—much less optimal—possibilities involving trials being live-streamed or recorded for consumption by the shadow jury. These alternatives may prove less disruptive or stressful for complainants in sensitive trials. Indeed, the broadcasting and recording of legal proceedings is now a familiar and uncontroversial practice, for much wider and less restricted circumstances than the proposed research. Taking just a handful of examples, appeals to the UK Supreme Court, public inquiries, cases at the Court of Session, and sentencing remarks at the High Court of Justiciary are routinely either live-streamed or recorded for wide public consumption—even when this might involve the disclosure of information that some participants would rather remained private. Any videographic version of the proposed research would be limited in audience and subject to strict rules about confidentiality and use of data. It remains an option if the preferable live version of the research was regarded to be untenable.

[68] https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-april-to-june-2023/criminal-court-statistics-quarterly-april-to-june-2023

long time, affecting thousands and thousands of criminal trials. Mistakes in both the direction of over-correcting *and* under-correcting in response to the threat of the alleged 'justice gap' for complainants are of grave importance. In the direction of over-correction, we risk unnecessarily weakening safeguards against false conviction, leaving citizens susceptible to wrongful punishment. In the direction of under-correction, we risk allowing widespread sexual criminality to continue with impunity, leaving victims without recourse. There is a moral imperative, especially when research is not univocal, to gather the best evidence that we can before committing to fundamental reforms of criminal justice.[69] In any case, using real trials as a stimulus is less costly than simulating multi-day trials using talented actors and paying participants to deliberate about those. The only types of research on jury deliberation that are much cheaper are types of research that are open to serious methodological criticism.

5. Conclusion

It is unsatisfactory in the modern era that consultations on legal reform feel the temptation to note that we currently lack basic knowledge about the jury system—a fundamental and ancient feature of criminal justice. A primary reason for this epistemic deficit is the legal and institutional hurdles preventing research on live jury deliberation. Nevertheless, live jury research remains extremely controversial and there are persistent concerns with the idea of interfering with live criminal proceedings for the purposes of research.

What this paper has done is propose another way to study jury deliberation that gets very close to the real thing, but without invoking the same concerns. This proposal is to use real criminal trials as the stimulus for high quality mock jury studies. Crucially, this proposal ameliorates important methodological deficits with the current research paradigm. Especially at a time when leading jury researchers disagree on basic points about the performance of juries, facilitating this research should be a welcome way to advance the debate. Reform of criminal procedure is a slow business, and any reforms are likely to remain in place for many decades. Given that this is so, it is incumbent on governments, the research community, and the legal profession to ensure that any such reform is predicated on gold-standard evidence, utilising as many different feasible methods as possible. In synthesising the findings from different research methodologies, we can advance the debate on jury performance. I hope to here have

[69] The moral imperative to gather best evidence applies equally as a response to other cost-related objections to the proposed research, for example concerning the effect that it might have on 'backlogs' of cases currently faced by certain jurisdictions (chiefly owing to the COVID pandemic). Simply, the importance of the issue warrants additional resourcing so that any research would not impede the rate at which such backlogs are cleared. In any case, the argument developed in this paper will remain relevant for future periods after which current backlogs have diminished. Thanks to an anonymous referee for pressing this issue.

made the case for a useful, underutilised and intellectually robust way to gather new and much-needed evidence.[70]

References

Anwar, S., P. Bayer, and R. Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127, no. 2: 1017–55.

Bindler, Anna, and Randi Hjalmarsson. 2018. "How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments." *American Economic Journal: Economic Policy* 10, no. 4: 36–78.

Bornstein, Brian. 'Jury Simulation Research: Pros, Cons, Trends, and Alternatives.' In *The Psychology of Juries.*, edited by Margaret Bull Kovera, 207–26. Washington: American Psychological Association, 2017.

Bornstein, Brian H., Jonathan M. Golding, Jeffrey Neuschatz, Christopher Kimbrough, Krystia Reed, Casey Magyarics, and Katherine Luecht. 2017. "Mock Juror Sampling Issues in Jury Simulation Research: A Meta-Analysis." *Law and Human Behavior* 41, no. 1: 13–28.

Brown, Jennifer, and Sandra Walklate, eds. 2012. *Handbook on Sexual Violence*. London ; New York: Routledge.

Burt, Martha R. 1980. "Cultural Myths and Supports for Rape." *Journal of Personality and Social Psychology* 38, no. 2: 217–30.

Chalmers, James, and Fiona Leverick. 2016. "How Should We Go About Jury Research in Scotland?" no. 10.

Chalmers, James, Fiona Leverick, and Vanessa Munro. 2021a. "Why the Jury Is, and Should Still Be, out on Rape Deliberation." *Criminal Law Review* 9 Sweet & Maxwell: 753–71.

Chalmers, James, Fiona Leverick, and Vanessa E. Munro. 2021b. "The Provenance of What Is Proven: Exploring (Mock) Jury Deliberation in Scottish Rape Trials." *Journal of Law and Society* 48, no. 2: 226–49.

Conaghan, Joanne, and Yvette Russell. 2014. "Rape Myths, Law, and Feminist Research: 'Myths About Myths'?" *Feminist Legal Studies* 22, no. 1: 25–48.

———. 2023. *Sexual History Evidence and the Rape Trial*. Bristol University Press.

Daftary-Kapur, Tarika, Steven D. Penrod, Maureen O'Connor, and Brian Wallace. 2014. "Examining Pretrial Publicity in a Shadow Jury Paradigm: Issues of Slant, Quantity, Persistence and Generalizability." *Law and Human Behavior* 38, no. 5: 462–77.

Daly, Ellen, Olivia Smith, Hannah Bows, Jennifer Brown, James Chalmers, Sharon Cowan, Miranda Horvath, et al. 2023. "Myths about Myths? A Commentary on Thomas (2020) and the Question of Jury Rape Myth Acceptance." *Journal of Gender-Based Violence* 7, no. 1 Policy Press, sec. Journal of Gender-Based Violence: 189–200.

Ellison, Louise, and Vanessa E. Munro. 2009a. "REACTING TO RAPE: Exploring Mock Jurors' Assessments of Complainant Credibility." *The British Journal of Criminology* 49, no. 2 Oxford University Press: 202–19.

———. 2009b. "Of 'Normal Sex' and 'Real Rape': Exploring The Use of Socio-Sexual Scripts in (Mock) Jury Deliberation." *Social & Legal Studies* 18, no. 3: 291–312.

Ellison, Louise, and Vanessa E Munro. 2010. "Getting to (Not) Guilty: Examining Jurors' Deliberative Processes in, and beyond, the Context of a Mock Rape Trial." *Legal Studies* 30, no. 1: 74–97.

———. 2015. "'Telling Tales': Exploring Narratives of Life and Law within the (Mock) Jury Room." *Legal Studies* 35, no. 2: 201–25.

Finch, Emily, and Vanessa E Munro. 2006. "Breaking Boundaries? Sexual Consent in the Jury Room." *Legal Studies* 26, no. 3: 303–20.

Gerger, Heike, Hanna Kley, Gerd Bohner, and Frank Siebler. 2007. "The Acceptance of Modern Myths about Sexual Aggression Scale: Development and Validation in German and English." *Aggressive Behavior* 33, no. 5: 422–40.

[70] Thanks to an audience at the University of Warwick for discussion, with special mention to Victor Tadros for a conversation that sparked interest in this idea. I'd also like to particularly thank the anonymous referees for their thoroughness and care.

Hanretty, Chris. 2020. *A Court of Specialists*. Oxford University Press.
Hedden, Brian R. 2017. "Should Juries Deliberate?" *Social Epistemology* 31, no. 4: 368–86.
Honess, T M. 1998. "Juror Competence in Processing Complex Information: Implications from a Simulation of the Maxwell Trial."
Horan, Jacqueline, and Mark Israel. 2016. "Beyond the Legal Barriers: Institutional Gatekeeping and Real Jury Research." *Australian & New Zealand Journal of Criminology* 49, no. 3: 422–36.
Jenkins, Katharine. 2021. "Rape Myths: What Are They and What Can We Do About Them?" *Royal Institute of Philosophy Supplements* 89 Cambridge University Press: 37–49.
Lee, Jae-Hyup, Jisuk Woo, June Woong Rhee, Jeong Min Choi, and Hyunki Shin. 2013. "What's Happening in the Jury Room? - Analyzing Shadow Jury Deliberations in Korea." 13, no. 1.
Leverick, Fiona. 2020. "What Do We Know about Rape Myths and Juror Decision Making?" *The International Journal of Evidence & Proof* 24, no. 3: 255–79.
Lonsway, Kimberly A., and Louise F. Fitzgerald. 1994. "Rape Myths: In Review." *Psychology of Women Quarterly* 18, no. 2: 133–64.
Lundmark, Thomas. 2010. "'Split Verdicts' in Scotland: A Judicial Survey." *Edinburgh Law Review* 14, no. 2: 225–35.
Majeed-Ariss, Rabiya, Alice Brockway, Kate Cook, and Catherine White. 2021. "'Could Do Better': Report on the Use of Special Measures in Sexual Offences Cases." *Criminology & Criminal Justice* 21, no. 1 SAGE Publications: 89–106.
Mott, Nicole. 2003. "The Current Debate on Juror Questions: 'To Ask or Not to Ask, That Is the Question.'" *Chicago-Kent Law Review* 78, no. 3: 1099.
Nosek, Brian A., Tom E. Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S. Corker, Anna Dreber, Fiona Fidler, et al. 2022. "Replicability, Robustness, and Reproducibility in Psychological Science." *Annual Review of Psychology* 73, no. 1: 719–48.
Ormston, Rachel, James Chalmers, Fiona Leverick, Vanessa Munro, and Lorraine Murray. 2019. "Scottish Jury Research: Findings from a Large Scale Mock Jury Study." Research Reports or Papers Edinburgh: Scottish Government. At https://www.gov.scot/publications/scottish-jury-research-fingings-large-mock-jury-study-2/, accessed April 27, 2023.
Patterson, Jeff, and Luke Spencer. 2019. "Lessons Learned in Jury Research: Mock Jurors, Shadow Jurors, Post-Truth Jurors, and Jury Consultants Trial Tips." *Brief* 49, no. 1: [iii]-64.
Payne, Diana L., Kimberly A. Lonsway, and Louise F. Fitzgerald. 1999. "Rape Myth Acceptance: Exploration of Its Structure and Its Measurement Using theIllinois Rape Myth Acceptance Scale." *Journal of Research in Personality* 33, no. 1: 27–68.
Reece, Helen. 2013. "Rape Myths: Is Elite Opinion Right and Popular Opinion Wrong?" *Oxford Journal of Legal Studies* 33, no. 3: 445–73.
Robbennolot, Jennifer. 2005. "Evaluating Juries by Comparison to Judges: A Benchmark for Judging?" *Florida State University Law Review* 32, no. 2: 469–509.
Ross, Lewis. 2023a. "The Curious Case of the Jury-Shaped Hole: A Plea for Real Jury Research." *The International Journal of Evidence & Proof* SAGE Publications Ltd. At https://doi.org/10.1177/13657127221150451, accessed January 5, 2023.
———. 2023b. "Criminal Proof: Fixed or Flexible?" *The Philosophical Quarterly*.
Saks, Michael J. 1997. "What Do Jury Experiments Tell Us about How Juries (Should) Make Decisions." *Southern California Interdisciplinary Law Journal* 6, no. 1: 1–54.
Slater, Joe. 2023. "Just Judge: The Jury on Trial." *American Philosophical Quarterly* 60, no. 2: 169–86.
Sommers, Samuel R. 2007. "Race and the Decision Making of Juries." *Legal and Criminological Psychology* 12, no. 2: 171–87.
Sunstein, Cass. 2003. *Why Societies Need Dissent*. Harvard University Press.
Taylor, Natalie, and Jacqueline Joudo. 2005. "The Impact of Pre-Recorded Video and Closed Circuit Television Testimony by Adult Sexual Assault Complainants on Jury Decision-Making: An Experimental Study: (583042012-001)." At http://doi.apa.org/get-pe-doi.cfm?doi=10.1037/e583042012-001, accessed November 20, 2023.
Temkin, Jennifer, and Barbara Krahé. 2008. *Sexual Assault and the Justice Gap: A Question of Attitude*. Hart Publishing.

Thomas, Cheryl. 2007. “Diversity and Fairness in the Jury System.” 2/07 Ministry of Justice Research Series.
———. 2010. “Are Juries Fair?” Ministry of Justice Research Series 1/10.
———. 2013. “Avoiding the Perfect Storm of Juror Contempt.” *Criminal Law Review* no. 6.
———. 2017. “Ethnicity and the Fairness of Jury Trials in England and Wales 2006-2014.” *Criminal Law Review* 11: 860–76.
———. 2020. “The 21st Century Jury: Contempt, Bias and the Impact of Jury Service.” *Criminal Law Review* no. 11 Thomson Reuters: 987–1011.
———. 2021. “A Response to ‘The Jury Is Still Out.’” *Criminal Law Review* 9: 772–73.
———. 2023. “Juries, Rape and Sexual Offences in the Crown Court 2007-21.” *Criminal Law Review* 3: 200–225.
Tinsley, Yvette, Claire Baylis, and Warren Young. 2021. “‘I Think She’s Learnt Her Lesson’: Juror Use of Cultural Misconceptions in Sexual Violence Trials.” *Victoria University of Wellington Law Review* 52, no. 2: 463–86.
Tribe, Laurence H. 1971. “Trial by Mathematics: Precision and Ritual in the Legal Process.” *Harvard Law Review* 84, no. 6: 1329.
Vidmar, Neil, Shari Diamond, Mary Rose, and René Ellis. 2003. “Juror Discussions During Civil Trials: Studying an Arizona Innovation.” *Arizona Law Review* 45: 1–83.
Vinson, Donald E. 1982. “The Shadow Jury: An Experiment in Litigation Science.” *American Bar Association Journal* 68, no. 10: 1242–47.
Waiton, Stuart. 2023. " What do we know about ‘rape myth’ research and the claim that there is ‘overwhelming evidence’ that juries are prejudiced in rape trials?" *The International Journal of Evidence & Proof*: 1-24. DOI: 10.1177/13657127231217510.
Wiener, Richard L., Daniel A. Krauss, and Joel D. Lieberman. 2011. “Mock Jury Research: Where Do We Go from Here?” *Behavioral Sciences & the Law* 29, no. 3: 467–79.
Wiggins, Bradford J., and Cody D. Christopherson. 2019. “The Replication Crisis in Psychology: An Overview for Theoretical and Philosophical Psychology.” *Journal of Theoretical and Philosophical Psychology* 39, no. 4: 202–17.
Zander, Michael and Paul Henderson. 1993. *Crown Court Study*. Research Study No: 19. The Royal Commission on Criminal Justice. HM Stationery Office.
Zander, Michael. 2013. “Research Should Not Be Permitted in the Jury Room.” *Criminal Law & Justice Weekly* 177.
Zeisel, Hans, and Shari Seidman Diamond. 1978. “The Effect of Peremptory Challenges on Jury and Verdict: An Experiment in a Federal District Court.” *Stanford Law Review* 30, no. 3: 491.