



The curious case of the jury-shaped hole: A plea for real jury research

Lewis Ross 

Department of Philosophy, London School of Economics, London, UK

Abstract

Criminal juries make decisions of great importance. A key criticism of juries is that they are unreliable in a multitude of ways, from exhibiting racial or gendered biases, to misunderstanding their role, to engaging in impropriety such as internet research. Recently, some have even claimed that the use of juries creates injustice on a large scale, as a cause of low conviction rates for sexual criminality. Unfortunately, empirical research into jury deliberation is undermined by the fact that researchers are unable to study live juries. The indirect sources of evidence used by researchers suffer from various problems, the most important of which is dubious levels of ecological validity. Real jury research—studying live jury deliberation—is controversial. However, as I argue, the objections to it are unconvincing. There is in fact a moral imperative to facilitate real jury research.

Keywords

bias, criminal law, criminal procedure, jury trial, mock juries, rape conviction rates, real jury research

Introduction

Jury trials are among few situations where private citizens direct the coercive machinery of the state. Unsurprisingly then, the merits of jury trials are a topic of perennial debate. Given that most trials across the world are decided by professional judges, the right to be tried in front of a panel of lay peers can seem like a cumbersome affordance.¹ However, defenders of trial by jury claim a number of advantages, such as instantiating or symbolising democracy,² improving the character of those who

1. Here, I focus on juries in criminal trials. I won't discuss the (less common) use of juries in civil trials, although much of what I say applies.

2. E.g. Abramson (2000).

Corresponding author:

Lewis Ross, Department of Philosophy, Logic and Scientific Method, London School of Economics, Lakatos Building, London, WC2A 2AE, UK.

Email: l.ross2@lse.ac.uk

participate,³ serving as the ‘conscience of the community’,⁴ and even acting as a bulwark against state oppression.⁵

Periodically, fuel is added to the debate in the form of conjecture and empirical evidence suggesting the poor performance of trial by jury. Taking a topical and much vexed example, some suggest that jury trials are partly responsible for the execrably low conviction rate for sexual crimes, creating what has been called a ‘justice gap’ for complainants.⁶ This concern has been buttressed by evidence suggesting that jurors are influenced by a variety of ‘rape myths’ when deliberating.⁷ This is an example of an *epistemic objection* against trial by jury. Broadly, epistemic objections contend that juries are unreliable arbiters of truth or that they reach the truth in an unsafe or deficient manner. Understanding the nature and weight of epistemic challenges to jury trial is crucial for sensibly comparing juries with alternative systems, like trial by professional judge, and thus essential for deciding in whose hands criminal adjudication should lie.

But there is a glaring shortcoming that undermines debate about jury trials: namely, the paucity of evidence on real juries engaged in live deliberation. Research on jury deliberation is almost entirely indirect, typically using mock juries or various types of self-report survey. There are many problems with the current research paradigm. Among other issues: the ecological validity of current methods is extremely dubious; most research is from the United States and does not readily generalise to other jurisdictions; and there are contradictory results between high-profile research programmes that employ different methodologies. After explaining issues with the current research paradigm, I defend the obvious solution: researchers should be permitted to study real juries by allowing transcription of deliberations. The objections to this suggestion turn out to lack force. Real jury research is not only defensible and prudent: given the weighty interests at stake, it is a moral imperative.

Epistemic objections to jury trial

Trial by jury can be evaluated across two primary dimensions: (i) moral and political value, and (ii) epistemic efficacy or accuracy. How these dimensions relate is somewhat complex. The moral and political value of the jury can stand partly independent of its epistemic power. For example, the egalitarian or democratic value of juries might be decisive even if juries are somewhat less reliable than judges (just as elections might be preferable even if less reliable than having a benign dictator appoint public officials). However, whenever a jury makes a mistake, this is not just an epistemic problem. It is a moral problem whenever a jury finds an innocent person guilty, given the wrongful punishment and condemnation resulting from such errors. It is also a moral problem whenever a jury returns a false negative—when it acquits the guilty—since a common view is that states are under a moral obligation to deter crime, incapacitate the dangerous and/or offer the appropriate retributive penalty for wrongdoing. No matter what other strengths trial by jury may have, there is presumably some threshold of reliability that juries must surmount in order to be morally acceptable. The democratic value of the jury

3. This argument is an old one and often ascribed to Mill (e.g. 1975: 75–76) but is now widely championed in work on ‘deliberative democracy’.

4. E.g. see Lee (2018: Section IV) for general discussion.

5. E.g. see Huemer (2018).

6. In the UK, the recent ‘End-to-End Rape Review Report’ suggested that the complaint-to-conviction level for sexual crimes is below 2% (see: <https://www.gov.uk/government/publications/end-to-end-rape-review-report-on-findings-and-actions>). Some causes of the low conviction rate problem are found much earlier in the criminal justice process than jury trials, such as the decision not to prosecute certain types of crime. But these are intimately related: prosecutors will only proceed when they think there is a reasonable chance of a jury opting to convict. Nevertheless, the conviction rate conditional on prosecution for sexual crimes is still lower than that of other serious crimes.

7. Leverick (2020) and Chalmers et al. (2021a) discuss the current state of play, with Thomas (2020) providing a more sceptical perspective. These results will be discussed in detail below.

(for example) would not provide adequate compensation if juries were facilitating large-scale miscarriages of justice.⁸ Epistemic evaluations of trial by jury are perhaps best viewed as a *subset* of the overall moral evaluation of jury trials.

The basic contention of epistemic objections is that juries tend to make incorrect, misinformed or otherwise unsafe decisions. Before proceeding, we should acknowledge that jury critics—and indeed jury defenders—generally lack privileged access to whether a jury did return the right decision in any given case. Nevertheless, juries can make decisions that are clearly epistemically or procedurally deficient, in ways that should make us doubt the reliability of their verdicts. Here is a list of different epistemic objections to trial by jury:

1. Jury decisions are influenced by interpersonal biases, most prominently:⁹
 - Racial bias against out-groups/in favour of in-groups.
 - Gendered biases, e.g., associating gender with criminality; misogynistic myths about sexual consent.
 - Socioeconomic biases, e.g., associating ‘class’ with criminality.
 - Intra-jury bias, where interpersonal biases affect the quality of deliberation (e.g., jurors sidelining or being dominated by certain participants).
2. Jurors fail to understand their legal role or the legal parameters constraining their decision. For example, they might not understand judicial directions, the standard of proof, whether they can permissibly rely on suppositions not adduced in court, or the distinction between the *actus reus* and *mens rea*.¹⁰
3. Jurors engage in impropriety such as carrying out their own research or ‘nullifying’ trials by returning verdicts contrary to their assessment of the facts.¹¹
4. Jurors are susceptible to misunderstand the evidence presented in court, especially when it is complex (as in a fraud trial) or contains statistical components (as with DNA evidence).¹²
5. Jurors are susceptible to ‘manipulation’, e.g., by lawyerly rhetoric, gruesome evidence and other aspects of trial strategy that do not reliably uncover the truth.¹³
6. Juror decisions vary with their idiosyncratic personal characteristics, making the results of criminal trials susceptible to a degree of arbitrariness.¹⁴
7. Jury deliberation degrades the reliability of individual assessment. For example, some evidence suggests that collective discussion can create worse results (e.g., by extremifying individual viewpoints) and some important theorems on the value of collective decision-making rely on the independence of those deliberating.¹⁵

Working out the extent to which these objections afflict juror decision-making is necessary to decide the best way of structuring criminal trials.

8. This is no idle worry. As mentioned in the introduction, some have argued that juries *are* facilitating widespread miscarriages of justice in the context of sexual criminality.

9. Race is the topic where the most research has been conducted, but usually in the US context. E.g. see Mitchell et al. (2005) and Sommers (2007). For discussion of bias more generally, see Daftary-Kapur (2010). The UK’s Lammy Report did not find, using quantitative data, that BAME accused persons had much different chance of conviction than white accused. This geographical divergence in evidence is a problem I return to later on. See footnote 7 for evidence on sexual/gender biases.

10. Thomas (2010; 2013; 2020) touches on some of these issues.

11. I use the term ‘impropriety’ here in a legal rather than moral sense. In my view, sometimes jury nullification is morally permissible and even obligatory.

12. See, for example, Monaghan (2018).

13. E.g. see Grady et al. (2018).

14. E.g. Devine and Caughlin (2014).

15. Hedden (2017) provides discussion.

The jury-shaped hole

Debate about the value and reliability of juries is currently hamstrung. It is hamstrung by the lack of direct evidence about *real jury deliberation*. By this I simply mean evidence taken from actual juries making decisions in the context of genuine legal trials. Globally, there is no jurisdiction that has conducted extensive real jury research on ‘live’ deliberation. In some countries, legislation presents a roadblock to real jury research, including the United Kingdom (the 1974 Juries Act and the 1981 Contempt of Court Act)¹⁶ and Canada (s.649 of the Criminal Code). Such legislation makes it a criminal offence to expose the content of jury deliberations. But even when there are not formal legislative bars on real jury research, academics often find that the institutional hurdles in gaining access are insuperable.¹⁷ Given that juries are not required to give reasons for their decision in court, their deliberations are a black box to academic researchers.¹⁸ In short, a *jury-shaped hole* is found when searching for direct evidence on how real juries deliberate when faced with high-stakes decisions.

Due to the opacity of the real jury, researchers must instead rely on various alternative proxy methods when attempting to study juries. These methods vary greatly in the extent to which they tailor the mode of inquiry to what is distinctive about juries. Different methods can be used in conjunction with the hope—a hope I argue may be forlorn—that it enables us to paint a comprehensive and accurate picture of real jury deliberation, even without direct evidence about what goes in the jury room.

Some methods—call these *generic methods*—attempt to draw inferences about real juries by appealing to data that is not specific to the task of being a juror. Generic methods often have the advantage of being reliant on readily accessible information: they are therefore economic to rely upon and do not require special access being granted to researchers. These include the following:

General Attitude Surveys. This method attempts to derive the attitude of jurors from general attitudes prevalent in society. For example, to research the role of rape myths in jury deliberation, academics might look at how people in a given society tend to score on various ‘rape myth acceptance scales’.¹⁹

Quantitative Legal Data. Another method is to draw inferences from quantitative legal data, such as statistics about the comparative conviction rate for different crimes or conviction rates between jurisdictions that use different trial methods. For example, low conviction rates for sexual criminality relative to other crimes has motivated some to identify the jury as a roadblock to conviction.^{20,21}

Deliberation Studies. Deliberation is studied by social-scientific researchers in other contexts. One way to speculate about the efficacy of jury deliberation is to draw on pre-existing generic research into the benefits and harms of collective deliberation.

16. See section 20D of the former (England and Wales) and s. 8 of the latter (Scotland and Northern Ireland). Thomas (2013) discusses the relationship between contempt of court, jury behaviour and jury research in detail.

17. Horan and Israel (2016) provide a very helpful discussion of these hurdles (from the Australian perspective) and offer suggestions about how we might address some of these.

18. Of course, the ‘unreasoned’ nature of jury verdicts is something that could be changed, although it is bound up with the current orthodoxy that jury deliberations should remain secret. Historically, some courts and jurisdictions have asked for juries to provide reasons to accompany their decisions. For discussion and an argument for requiring juries to provide explanations for their verdicts, see, for example, Coen and Doak (2017).

19. E.g. the Illinois Rape Myth Acceptance Scale (IRMAS), the Acceptance of Modern Myths About Sexual Aggression (AMMSA) or the Subtle Rape Myth Acceptance Scale (SRMAS). Each has been subject to methodological criticism.

20. Slater (*forthcoming*) provides a nuanced discussion of the idea that we ought to remove the jury in sexual offence trials.

21. An obvious limitation here is that we lack access to the counterfactual: how would the case have been decided had it been decided by a judge alone? Comparing sexual offence conviction rates to other serious crimes is, in a way, like comparing apples and oranges. One strategy to deal with this problem concerns the use of studies into ‘split verdicts’, where judges self-report disagreeing with the verdict of a jury that they presided over. See footnote 22 for an example.

Other methods—call these *tailored methods*—are more closely related to the object of study, aiming to generate evidence about the phenomenon of being a juror in particular. The two predominant methods are the following:

Mock Juries. Researchers find participants to serve in ‘mock juries’ who are asked to read or, less frequently, watch a short trial simulation and deliver some type of assessment. These simulations come in varying shades of realism.

Post-Trial Surveys. With varying degrees of institutional support, researchers can survey real jury members *after* they have served on a jury. These surveys ask jurors to self-report on their experience as a juror. In some jurisdictions, the content of what can be asked is limited by law.

These methods for studying juries can be valuable, even offering advantages over studying live juries. Notably, the use of multiple mock juries allows researchers to engage in experiments to test specific hypotheses, to use control groups, and to isolate and change only selected variables (e.g., the gender of the jurors) in a way that would be impossible using real cases. It is for this reason that we should acknowledge that tailored methods have an important role to play in an adequate science of the jury. However, as I will go on to explain, we have decisive reason not to exclusively rely on these proxies for studying real juries.

Against the current research paradigm

This section outlines the problems with research methods that fall short of studying live jury deliberations.

Two generic methods, *deliberation studies* and *quantitative legal data*, are uninformative and should be given little weight. Quantitative data about trends in (for example) conviction rates can draw our attention to certain phenomena. But it is extremely hard to draw any particular conclusion from statistical information alone, let alone detailed conclusions about the quality of jury deliberation. For example, some studies suggest that when professional judges disagree with jury convictions, this is commonly regarding sexual assault convictions.²² But such findings are compatible with many competing hypotheses: e.g., that jurors apply lower standards of proof, that they are more or less susceptible to various biases, that they better understand the moral issues in sexual assault trials, that they misinterpret legal directions, and so on. Each of these would mean different things for the reliability of juries and would require different types of practical intervention if we wanted to change such trends, thus providing little guidance on how to improve trial design.

Similarly, deliberation studies also fail to bear on many of the specific questions we have about juries, since many of the phenomena we are interested in are domain-specific. For example, questions about legal understanding, lawyerly manipulation and juror impropriety are not ones that can be resolved by looking at generic instances of deliberation. Nor is it obvious that the strengths and weaknesses of collective deliberation in other contexts studied by social scientists—e.g., deliberation about whether to engage in a civil infrastructure project—carry over to collective deliberation about moral responsibility found in the criminal courtroom. These tasks involve different types of reasoning and there is no guarantee that strengths and weaknesses of deliberation are shared between domains. Finally, a lesson we can learn from analogous discussions of social-scientific research to evaluate democratic decision-making in the political context is that theorists often draw diverging lessons from the literature, suggesting that this work is far too contested to enable us to draw clear lessons

22. See Lundmark (2010).

about jury performance.²³ Of course, such evidence might play an ancillary role in jury science. But, by itself, it cannot be accorded a decisive role.

This leaves three methods to be discussed: *general attitude surveys*, *mock juries* and *post-trial surveys*. We will treat these at some length. My general complaint is with the ecological validity of these methods. This simply means that there are strong reasons to doubt that results derived from current methods licence us to draw inferences about real juries.

General attitude surveys do not tell us anything about many of the epistemic objections relating specifically to being a juror, such as juror comprehension (factual and legal), impropriety and manipulability. Similarly, there are some things that mock jury studies cannot readily investigate. For example, three concerns we might have about juries are: (i) the prevalence of independent research, (ii) the effect of media reporting on deliberation and (iii) whether jurors are reluctant to approach judicial officials when they misunderstand the law. These issues cannot be easily studied by a mock jury, depending as they do on features of the stressful real-life context of actual criminal trials. These problems are straightforward and so I won't belabour them. Rather, in what follows, I want to claim that even when we consider questions more hospitable (in theory) to the use of attitude surveys and mock juries, current methods are in bad shape. Since it is topical, I will often use the investigation of bias as a focal example.

There are two broad categories of problem with the current research paradigm, each of which undermine its usefulness and contribute to a lack of ecological validity:

1. Flaws in current research design, scope and funding.
2. Conceptual problems with the entire project of jury science without studying live juries.

These issues are not always easy to separate. Criticisms about design are often a consequence of the lack of access to real juries. Still, as I hope to show, there are conceptual issues that would undermine even the most rigorous and best-resourced studies. The need for real jury research cannot be vitiated simply by improving our methods for studying something other than live jury deliberation. I will start by focusing primarily on the first type of issue, before moving to discuss the second issue later in the section.

One recurring flaw in the extant research paradigm is that it fails to effectively replicate the conditions under which real juries make decisions. The first failure concerns the omission of deliberation. The fact is that general attitude surveys and *most* mock jury studies do not involve a deliberative component. That is, they exclude what is among the defining characteristics of jury service—discussion and debate among a group of peers with a view to forming a consensus opinion. General attitude surveys, of course, are not designed to mimic deliberation. The task of the participants in an attitude survey and a jury are entirely different: one is answering a survey while the other is deliberating in a courtroom as part of a collective. It is one thing to exhibit a bias when answering a survey, and another for it to persist in the context of prolonged collective deliberation in a high-stakes scenario. (And vice versa! It is one thing to fail to exhibit a bias when answering a survey, and another to converse and argue in a way that is free of bias). Some high-quality mock jury studies do involve deliberation; we will return to this later. But most do not, which undercuts their value as evidence about real juries. A basic defence of deliberation is that it enables people to pool evidence and change their views in response to challenge. Indeed, this is arguably the *very point* of deliberation. By failing to model the deliberative aspect of being on a jury, many current studies simply miss the point.

A second flaw is that many studies fail to ask participants to answer the questions that they would be asked in the context of a real trial. General attitude surveys, of course, do not attempt to replicate the legal decision-making process: the question is about attitudes in general, not to the guilt or innocence of an accused person. But even many mock jury studies, where we would hope that realism is at a

23. For example, compare the assessment of the empirical evidence made by Brennan (2016) with that given by Landemore (2012).

premium, ask an array of different questions which are simply unlike those faced by real jurors. For example, studies often ask jurors to rank culpability on a Likert-scale rather than the binary guilty/not guilty verdict-structure confronted by most jurors. Another issue is that mock jurors are often asked to assess guilt in a way that fails to replicate the proof structure of real trials, where both the *actus reus* and *mens rea* must be established on the relevant standard of proof. These flaws are unfortunate because they constitute a missed opportunity to test juror comprehension of real trial instructions or standards or proof and to see how the different responses given by mock jurors would affect real trial verdicts (e.g., it is hard to convert a Likert-scale on culpability to a binary verdict-schema, since we don't know where the 'cut-off' for outright guilt is in the mind of the juror). A real problem generated by this divergence is that it becomes hard to work out how the results of a mock jury would translate into a real jury verdict, especially in systems where jury verdicts need to approach unanimity or be unanimous.

A third limitation that afflicts many studies is the absence of an attempt to simulate a real trial environment and process. While environmental differences by themselves make be significant (e.g., if we thought that the solemnity of the trial process is psychologically relevant, which is surely is), a larger worry concerns the fact that many mock jury studies lack the narrative detail that we find in modern trials, including the absence of cross-examination and painstaking consideration of the narrative from different angles by prosecution and defence lawyers. Rather, most mock jury studies rely on short videos or written vignettes that simply present a scenario. There are specific empirical reasons to doubt the validity of inferences from mock studies lacking such narrative detail to real juries. One of the most popular paradigms for thinking about jury decision-making is Pennington and Hastie's 'story model', on which jurors make decisions as an exercise in narrative construction.²⁴ Juries try to construct stories which best fit the evidence, using ease of simulation as a heuristic for deciding which of the competing narratives is more likely to be correct. Ease of simulation is a psychological phenomenon which will be influenced, plausibly, by (i) third-party prompting to consider competing narratives and (ii) the availability of narrative detail onto which to project a story. Short vignettes found in most mock jury studies lack the prompting provided by prosecution/defence lawyers and cross-examination, and they lack the narrative richness found in real multi-day trials. This means that judgements formed by participants in mock studies are responding to stimuli lacking the features hypothesised to drive the decisions of real juries.

We can illustrate these empirical lacunae with some examples. Sticking with the focal case study of sexual bias, consider the following. A recent meta-analysis by Leverick (2020) canvasses an impressive amount of evidence that there is a statistically significant connection between an adverse score on various scales used to measure susceptibility to rape myths and (i) 'victim blaming' attitudes in particular cases and (ii) reluctance to convict the accused in particular cases. 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 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.

We have discussed some reasons to doubt the validity of mock jury studies. Let's turn to a more specific issue with the attempt to draw policy-relevant inferences. Observe that the ecological validity of

24. See Pennington and Hastie (1991, 1992, 1993) for foundational work on the story model.

every jury study is geographically and temporally limited. This is for the simple reason that societies vary in the attitudes of their citizens and because societies change over time. Such variation has obvious importance for the performance of jury trials. For instance, it is particularly obvious when considering worries about bias. Consider how social attitudes towards race, gender and sexuality have changed since the landmark jury studies of the 1950s. But the issue in fact ramifies far beyond bias: for example, internet competence, legal understanding, and financial and scientific literacy (relevant for complex trials) are all characteristics that can vary between populations and over time periods. The issue is equally clear across geographical as well as temporal space, given that attitudes towards (e.g.) sexuality and race vary substantially between different jurisdictions and, to a lesser yet doubtless significant extent, between regions of the same jurisdiction. The bulk of the scientific literature on juror psychology comes from the United States.²⁵ For example, the vast majority of research on the degree to which people accept rape myths has been carried out on US populations. This does not provide a basis from which to make recommendations about legal reform in other jurisdictions, such as the implication for UK legal systems that I've been focusing on here. That different cultures differ in their attitudes towards sexuality and sexual morality is a piece of common sense. As expected, there is empirical work which finds diverging susceptibility to rape myths in different societies, work which also suggests that the prevalence of such in the UK was 'low' (Barn and Powers 2021).²⁶ I make no claim about whether the acceptance of such myths is in fact low in the UK context, but clearly the geographically limited nature of such empirical work should give us pause when attempting to draw general conclusions from country-specific data. Another limitation worth pointing out, as does Leverick (2020), is that extant studies on rape myth acceptance fail to devote any attention to cases where the complainant is a male. While I have focused on studies into sexual attitudes, these general shortcomings are merely illustrative of general concerns with extant mock jury work.

At this juncture one might ask: *Can these limitations be ameliorated simply by conducting more rigorous, better-supported or better-resourced mock jury research?* I now argue that there would remain powerful reasons for scepticism, even regarding better-resourced and more faithful studies. Even if mock jury studies did replicate criminal trials perfectly, they would still only be replications. Mock jurors do not confront a decision which holds any real-life importance (nor do those who complete general attitude surveys like a Rape Myth Acceptance Scale). Real-life juries, in contrast, are deliberating about something that will have life-changing effects for those subject to their decision. This adverts to a classic philosophical distinction, between practical and theoretical reasoning. One popular view is that deliberation about a decision with practical consequences (as faced by a real jury) is an entirely different cognitive task than a decision that is merely theoretical (as is faced by a mock jury). Without taking a stance on the underlying psychology behind these processes, it is at least reasonable to suspect that people might adopt different deliberative approaches when faced with a decision of very high stakes (e.g., a real jury knowing that a false positive is a 20-year jail term) compared to cases where there are no consequences (and where it may be known that there is not even a 'real' answer). Indeed, a natural view is that people require more evidence in order to be convinced of a judgement when the costs of error are higher.²⁷ (For example, someone will need more evidence to be confident that a meal is free of peanuts when their guest has a fatal nut allergy than they would otherwise). The costs of error, plausibly, determine the evidential standards we bring to bear when engaging in inquiry.²⁸

25. See footnote 9 for discussion of divergent evidence about racial bias and juries between the UK and the US.

26. The studies discussed suffer from the substantial limitation of being conducted on a student population. As noted, we might expect such a sample to be unrepresentative of the broader population when it comes to attitudes about sexuality.

27. In epistemology, this view is often called 'pragmatic encroachment'. For recent work, see Kim (2017).

28. Indeed, as I have argued elsewhere (see Ross *forthcoming*) 2020, it is attractive to suppose that responsible legal fact-finders will use higher evidential standards when confronted with cases where the costs for the accused are higher. For example, they will require stronger evidence before regarding a murder conviction as permissible than they would a petty theft conviction.

Indeed, the jury instruction used in criminal cases in England and Wales is to ask whether they ‘are sure’ of the guilt of the accused. ‘Being sure’—supposedly equivalent to beyond a reasonable doubt²⁹—seems like a strong candidate for a mental state that is sensitive to the consequences of the decision. This is not merely philosophical speculation; some empirical evidence in fact supports the idea that the severity of punishment at stake influences jury verdicts, with jurors being more reluctant to convict when their verdict will lead to serious costs for the accused.³⁰ Mock juries face decisions with no real cost of error (e.g., no punishment in the offing) while real juries can literally be faced with life-or-death decisions. If the costs of error do determine the standards and methods we use when inquiring, then we have a clear argument for doubting the ecological validity of many mock jury studies—even if the study was impeccable in addressing other methodological concerns outlined in this section.³¹ This provides a theoretical basis for a suggestion made by Cheryl Thomas (2020: 1010) in her criticism of mock jury work, where she supposes that serving on a real jury might have a ‘transformational impact’ on jurors compared to participants in a psychological trial. The gravity of being entrusted responsibility over the life and liberty of the accused is one example of a transformational feature of jury service, a feature not replicable by mock studies.

Beyond the criticisms already noted, there are a host of familiar ecological validity criticisms that afflict any artificial psychological study into juries. For example, mock jurors are a self-selecting sample which may not represent the broader range of citizens who are compelled by law to serve on a jury. A long-standing criticism of many psychological studies is that undergraduate students are not representative of the viewpoints of the broader population.³² This can be remedied in part by seeking more representative samples, but Thomas’s (2020: 1006) post-trial surveys of real jurors found that 87% of jurors canvassed would not have participated in jury service had it been optional or voluntary. This casts doubt on whether those who participate in mock jury trials on a voluntary basis can ever provide a representative sample.³³

The presence of the experimenter influencing responses and behaviour is another obvious concern. Participants in mock juries are keenly aware that they are taking part in a study—it is the salient feature of their task. This worry extends to post-trial surveys. As Daly et al. (2022: 5) point out, by way of illustration, a recent programme of post-trial survey conducted by Cheryl Thomas’s Jury Project at University College London had a near 100% participation rate. This is of course unheard of among voluntary studies, where a small sample is used following an invitation extended to a large pool of participants. This raises the concern that participants felt that participation was somehow required or desired by the court itself, which raises obvious concerns about the veracity of responses.³⁴ The same critics note that ‘social desirability bias’ might be especially acute if participants believed that researchers were operating with the imprimatur of the court. Of course, the same worries about social desirability bias can be raised against the mock jury studies discussed above. While some of these problems might be dealt with through better design, it is extremely difficult to ‘design away’ the presence of the experimenter in an ethical way when it comes to any simulation of a trial.

29. See the *Crown Court Compendium* (5–3).

30. E.g. see Kerr (1978), Kaplan and Krupa (1986) or Bindler and Hjalmarsson (2018).

31. Of course, sensitivity to the stakes may be more impactful when it comes to certain activities of the jury (e.g. interpreting the standard of proof or deliberating about sexual consent) than others (e.g. understanding complicated statistical or forensic evidence). The weight of my conceptual objection to the current paradigm, like all of the issues raised, must be evaluated against the specific research that we have in mind.

32. However, see Bornstein et al. (2017) for a defence of mock jury research on this front.

33. Other (remediable) worries include whether participants take their task seriously, and the degree to which they understand what is being asked of them. For example, Cullen and Monds (2020) report wide inconsistencies in whether jury studies exclude those with poor comprehension and, if they do, the basis for the exclusion.

34. For example, responses which pertain to juror impropriety or incompetence are likely to be questionable in cases where respondents believe they are speaking to an official of the court.

Finally, we must also consider the credentials of mock jury research against the current state of play in social psychology generally. There are real concerns about historic and current research in social psychology. Psychology is currently in the midst of what is popularly known as a ‘replication crisis’—although the issues go beyond the replicability of results and refer to a broader suite of challenges that includes publishing practices, researcher misconduct, academic priority-setting and criticisms of how we understand statistical significance.³⁵ The challenges suggested by the replication crisis are striking. For example, large collaborative investigations suggest that we are not warranted in believing that even *half* of psychological studies would replicate (see Open Science Collaboration 2015). Even in the most prestigious and selective journals, results often turn out not to replicate. Moreover, results that fail to replicate are not reliably cited any less than results that do replicate, suggesting that studies with methodological flaws are taking root within the scientific literature. These worries extend to create doubt about some of the most widely discussed theories and research paradigms in modern psychology: there have been serious challenges to the ‘science’ of implicit bias, priming and nudging to take some high-profile examples.³⁶ A further methodological point should make us hesitant. The well-known ‘file drawer effect’ is a type of publication bias whereby researchers are much more likely to publish studies—and journals more likely to accept studies—establishing a statistically significant relationship between the hypothesis and the study variable. This effect also extends to the likelihood of studies being selected for inclusion in subsequent meta-analyses. Null results relating to hypothesised jury failings is a likely candidate for susceptibility to this publication bias.³⁷ Given these worries, we have generically strong reasons to doubt the reliability of many of the studies in the jury literature, apiece with reasons to doubt the reliability of many studies in the psychological literature more generally.

Post-trial surveys may avoid some issues raised in this section, since they notionally involve canvassing the opinions of real jurors reflecting on their real experience of a trial. But post-trial surveys are often limited in what researchers are permitted to ask—often by virtue of the same legislation that criminalises exposing the content of jury deliberations—and questions are not always tailored to the details of the case that they served on. These are particular flaws that might be remedied by removing certain restrictions, improving methodology and having a greater degree of institutional support. However, this would not solve important issues with such research. As even those who engage in post-trial surveys acknowledge, they are necessarily limited. Suppose that post-trial survey revealed that jurors endorse some biased line of thought, for example by accepting a myth about sexual consent. As Tinsley et al. (2021: 479) state, this may not tell us much about how important a role such reasoning played in the collective deliberation of the jury:

[W]e usually cannot know what would have happened had that reasoning not been employed. That is partly because the collective decision-making process may over-ride or neutralise flaws in individual juror decision-making, and there may have been a collective view that there was reasonable doubt anyway.

Moreover, post-trial surveys inherit the problems faced by all methods that rely on self-report of past experience. These are well-known by researchers in social science and include concerns about the unverifiable accuracy of recall (especially salient when reflecting on stressful experiences), the effect of social desirability bias whereby participants provide respectable rather than fully truthful answers (highly relevant to questions about prejudice and propriety) and the tendency for participants to overestimate their

35. See Wiggins and Christopherson (2019) for general discussion.

36. E.g. see Machery (2022), Schimmack, Heene and Kesavan (2017) and Hummel and Maedche (2019) for challenges to these respective research programmes.

37. Of course, given that the very nature of the issue concerns what evidence is made publicly available, estimating the effect of the bias is a very tricky business. See Kühberger et al. (2014) and van Aert et al. (2019) for differing perspectives on the extent of the problem.

own performance (relevant for questions about comprehension of legal directions, of evidence, of susceptibility to manipulation). There are also the previously-mentioned issues about the experimenter effect for such data, where we might worry about truthfulness with respect to jurors self-reporting various types of impropriety or undesirable attitudes. Such issues with self-report data are familiar so I won't belabour them. But these issues are especially stark for the focal example of the paper, namely bias when evaluating sexual complaints. Social desirability effect and experimenter presence is likely to exert a strong influence when querying jurors about their acceptance or reliance on 'rape myth' propositions, given that participants will know that such propositions are regarded as verboten by swathes of the population. Regardless of whether jurors claim not to be influenced by such myths, without access to the 'raw data' of deliberation transcripts there is no way to effectively cross-check whether this supposed rejection of such myths holds up in the secrecy of the jury room, or whether myths that are rejected explicitly were implicitly relied upon in deliberation. While post-trial research may involve real juries, it should not be described as 'real jury research'.

Experimental data that is limited or questionable in all of the ways outlined would be more reassuring if the evidence pointed unequivocally in the same direction. After all, we may view individual pieces of proxy jury evidence as pieces of a larger puzzle, where the truth only comes into view once pieces are assembled from different methods. However, the evidence derived from these indirect methods of studying juries can pull in different directions. And it does so concerning deeply important questions about the performance of juries. To illustrate, focusing on the UK context, consider again the debate on one of the most high-profile questions of legal policy, low conviction rates for sexual offences and the role of the jury as a potential cause. One influential scholar of evidence law, a champion of mock jury studies, argues that there is 'overwhelming evidence' that rape myths influence jurors in sexual assault cases.³⁸ But another leading light in jury research, director of the UCL Judicial Institute, in her post-trial survey research, suggests that 'hardly any jurors believe widespread myths and stereotypes about rape and sexual assault.'³⁹ This is not a comfortable situation if we want to use current methods to solve pressing policy questions. Commenting on these diverging findings, Chalmers et al. (2021a: 755), who champion mock jury methods, observe that it is:

[I]mpossible to know whether reported differences in results were genuinely due to jurors at court being somehow 'different' [than mock jurors], or were instead the result of different research design giving rise to unreliable results.

This, to me, represents the nub of the point. We lack confidence in the validity of current methods, and we lack ways to adjudicate between inconsistencies in the fruits of these methods. This is an unsatisfactory position to be in. But there is an alternative: real jury research.

A defence of real jury research

Given the limitations of current methods for studying juries, we should provide researchers with access to real-jury criminal trial deliberations. Such access could be as unobtrusive as audio recording deliberation, anonymising and transcribing it, then providing the transcriptions for research purposes only after some suitable time—for example, a number of years—has elapsed.⁴⁰ Of course, more involved methods

38. Leverick (2020: 255). See, also, the results of the large-scale mock jury project sponsored by the Scottish Government and run by many of the scholars supportive of mock jury science and critical of Thomas's post-trial surveys: <https://www.gov.scot/publications/scottish-jury-research-findings-large-mock-jury-study-2/documents/>.

39. Thomas (2020: 1001).

40. This suggestion is, naturally, not new. It has been suggested as a way to safeguard against racially prejudiced verdicts by Daly and Pattenden (2005). It is worth pointing out that audio recording may have problems of intelligibility.

involving videographic recording or live researcher presence are possible.⁴¹ But the ‘bare bones’ proposal would still constitute a considerable advance on current research methods into criminal juries in many respects. This access would need to be accompanied by efficient and reasonable procedures for providing researchers with access to the relevant material.⁴² Nonetheless, it is likely that RJR would be more economically viable than methodologically defensible mock jury study. Ecologically valid mock jury studies will often necessitate replicating long trials, with cross-examination, involving professional legal participants or paid actors, using real courtrooms or large venues, and compensating study subject for their considerable expenditure in time. RJR records something that is happening anyway rather than attempting to recreate it.

The basic argument for real jury research is simple: it would provide us with high-quality information about juries and how they respond to different pressures when deliberating about high-stakes and morally charged questions. A programme of real jury research would also help us better understand the extent to which current methods are probative, by providing empirical confirmation or disconfirmation for existing research. This information would be of inestimable assistance in improving the trial process: by helping us see the path forwards in making the legal system more hospitable for victims of sexual assault, to see the extent to which juries are influenced by media reporting, to see whether jurors generally understand key legal concepts and where they struggle if they do not, to see how juries handle complex evidence, to better understand the extent of juror impropriety, to see whether jurors discriminate against defendants, complainants and each other, and so on for every epistemic objection against the use of juries. And while this paper has focused on the epistemic credentials of the jury, we should remember that juries are also said to have moral and political value. I should emphasise: *many of the criticisms of current methods in testing the epistemic credentials of the jury also apply to our ability to assess the moral and political value of the jury more broadly*. If current methods lack ecological validity, they cannot be used to tell us (for example) whether serving on a jury has a character-improving effect, whether juries use their power to nullify unjust laws, or whether jurors make decisions with reference to moral convictions in such a way as to make them a democratic or representative body. This is yet another argument for RJR. In short, RJR would leave us better informed about the benefits and disadvantages of the jury system overall—and, if we decide to retain jury trials, how to make them work best for society.⁴³

The basic case for real jury research is clear—the burden is on the opponent to explain why we should *not* engage in such research. Of course, not all useful research should be conducted. Some research violates the rights of those involved, might lead to practical dangers, or might degrade social harmony. Real jury research is not on par with testing dangerous medicines on human subjects, studying how to construct dangerous biological weapons, or studying group differences in intelligence. Nevertheless, the idea of providing academic researchers access to real juries (or more accurately, transcriptions of their deliberations) is surprisingly controversial. No country routinely allows such access. And jurors themselves seem to agree that their own deliberations should be kept secret.⁴⁴ So, I want to consider a range of objections to real jury research that I have encountered when discussing this proposal with legal academics and practitioners.

41. This is not unheard of. Permission was once granted to videotape 50 *civil* juries in Arizona for a targeted investigation of how one procedural change might affect the quality of jury deliberation. See Diamond et al. (2003).

42. See Horan and Israel (2016) for more detailed discussion of informal hurdles for researchers in the criminal justice system.

43. Such work would also be of wide academic interest: it would contribute to the burgeoning study of collective deliberation, a topic of intrinsic interest and instrumental importance for the evaluation of democracy more generally.

44. Thomas (2010: 40) suggests, on the basis of a post-trial survey of real jurors, that 82% agreed with the proposition that jury deliberations should be secret. However, the significance of this finding is somewhat tempered by later findings by Thomas (2020) which suggest that juror comprehension of the disclosure rules is generally abysmal.

I will start with what I take to be the main objection.

Objection one: RJR would cause injustice

Justice to the accused and the putative victim (or the appearance thereof) requires prohibiting extraneous influences. Recording jury deliberations could affect the way that the jury deliberates and, therefore, influence the outcome of the trial.

I agree that any interference with the jury must pass a high justificatory bar, due to the weighty interests at stake in a criminal trial. For example, it would certainly be unjustified to test hallucinogens on jurors to advance hallucinogen research, given (i) the availability of other ways of advancing such research, (ii) the fact that conducting such research is not a pressing moral priority and (iii) the likely deleterious effect that hallucinogens would have on the quality of deliberation. But real jury research, I suggest, passes the justificatory bar on each count.

Firstly, as we have discussed already, other ways of prosecuting jury research—for example mock juries—have real question marks concerning ecological validity. Studying real juries is not an indulgence when it comes to working out how real juries deliberate but rather a necessary means to the end. This seems especially compelling in the context of the current research environment, where mock jury studies and post-trial surveys point in different directions, and there is a lack of higher-order confidence about the appropriateness of the methods used.

Secondly, there *is* a pressing moral imperative to advance our understanding of how juries operate. In this sense, we need to attend to the risks of *not* engaging in such research. Epistemic objections to jury trials cast them as unreliable, claiming that we have reason to think that juries often make mistakes that lead to innocent people being convicted or guilty people going free. Moreover, according to some epistemic objections, the inaccuracy of jury decision-making is particularly onerous for groups that are already subject to bias, like certain ethnic groups or victims of sexual assault. These concerns about accuracy are of tremendous moral weight, when multiplied over the many thousands of criminal trials that occur annually and will continue to occur in the future. The state has a duty to structure criminal trials so that we eliminate as far as possible the harms of false positives and false negatives—or at least strike the right balance between the two. Real jury research will help the state discharge this duty and so there is a strong moral argument for permitting it.

One might respond by insisting that the prohibition on RJR is simply deontic, in virtue of some *right* that trial participants have not to be subject to even modest extraneous influence. But at best (for the opponent of RJR) this style of response will lead to dialectical deadlock, since there are also rights that we can ascribe to participants in trials that demand that they are tried in a way that meets certain standards of deliberative competence. Consider Brennan's 'Competence Principle', taken from philosophical work in deliberative democracy:

Competence Principle: It is unjust to deprive citizens of life, liberty or property, or to alter their life prospects significantly, by force and threats of force as a result of decisions made by an incompetent or morally unreasonable deliberative body, or as a result of decisions made in an incompetent and morally unreasonable way. [Brennan 2011: 704]

It seems to me that the right to have weighty decisions decided in a competent manner is just as, if not more, compelling than any right that would advocate against transcription of deliberation. Given that questions about the competence of juries is precisely what is at issue in motivating RJR, there are deontic considerations in favour of real jury research—considerations that very arguably trump those against it.

Thirdly, I concede that the moral case for real jury research would be undermined if we had strong reason to suppose that the extraneous influence of transcribing jury deliberation would create widespread

miscarriages of justice. By this I mean, the moral argument would be undermined if we had decisive reason to suppose that real jury research would cause inaccurate verdicts to be returned where, counterfactually without transcription, the jury would have gotten things right. However, I do not see what reason we have to accept such a suggestion, and certainly not to think that we have such decisive reasons to believe it. Even if we supposed that RJR might influence the content of some jurors' assertions in the jury room, it is far from obvious that this would have a net deleterious effect. Earlier, we discussed worries that the presence of experimenters might cause the suppression biased assertions on behalf of participants. In a more attenuated way, we might have a similar worry about RJR. This may amount to an argument against the ecological validity even of RJR (I return to this shortly in Objection 2). But it is not a clear *moral* argument against RJR, for the reason that the suppression of biased assertions is not obviously an accuracy-inhibiting phenomenon. The standard assumption after all is that the expression of (e.g.) rape myths in deliberation is something that makes juries *less* accurate. So, one might doubt that the suppression of biases would itself foster inaccuracy or widespread injustice.⁴⁵ After all, I take it that there is no moral argument against real jury research on the grounds that it would *improve* the quality of deliberations.⁴⁶ In short, there is no obvious moral argument against RJR once we consider the types of influence that it might plausibly have. And the long-term project of designing a defensible approach to criminal adjudication is of such deep importance that the expected negative influence of RJR would have to be very great to overturn the moral argument for RJR.

Objection two: RJR would itself lack ecological validity

Transcribing real jury deliberation will change the way that jurors deliberate. For example, jurors will show less candour if they are being observed. So, the data will not generalise to real juries deliberating in a non-observed scenario. Real jury research would therefore be self-defeating and ecologically invalid.

As discussed above, one way to illustrate this objection might be to consider concerns relating to bias. Perhaps jurors with racist or sexist views would be less likely to advance these views in deliberation if they knew that their comments were being transcribed. This would mean that the evidence gained from real jury research could mislead, e.g., by making us underestimate the prevalence of bias in deliberation and fostering a false sense of security about the epistemic competence of the jury.

The first practical response is that the method for recording jury deliberation could be very unobtrusive, such as audio recording. Jurors are already in an unfamiliar situation when entering the jury chamber and most will be serving as jurors for the first time. Introducing audio recording is not a distortion of the status quo for these jurors, as they have no prior experience of being in a jury room.

Secondly, in mock jury studies, it has not been shown that jurors refrain from making assertions with biased contents. Chalmers et al. (2021b) document rape myths at play in mock jury deliberation. This suggests that observation is not entirely destructive of candour in the way that this objection supposes. (What *is* unclear is whether these biased assertions would be dealt with differently in the context of deliberation if the jurors were facing the task of delivering a verdict that would have a significant real-life consequence.) Of course, accepting this rebuttal does not require rowing back on our earlier scepticism about the ecological validity of mock jury studies. It seems sensible to expect that any suppressive effect of observation would be *weaker* in the case of discreet recording of deliberation in the unfamiliar context

45. One issue may be: if biased jurors were inhibited (by the prospect of recording) from airing their biases, which in turn prevents these biases from being challenged and refuted by fellow jurors, causing a material difference in the eventual verdict. Some of what I say about *Objection Two* speaks against this worry, but I do not dismiss it out of hand.

46. Similarly, if we thought that observation might lead to jurors not sharing the fruits of illicit personal research, it is hard to see that there is any moral argument against an influence that would improve the procedural fairness of jury deliberation.

of jury service compared to people who are overtly participating in a psychological study such as mock jury experiment.

Thirdly, and most importantly, this objection proves too much to be convincing. Accepting it would engender a much broader scepticism about social-scientific research. Consider that the ‘observation destroys candour’ would equally apply to a large swathe of the methods we currently use to draw inferences about real juries. For instance, mock jury studies involve observation in an even more salient way than would audio recording of real juries (participants are signing up for a psychological study after all). If observation undermines the validity of data, then many current methods are in equally bad if not worse shape.

Objection three: Jury privacy is politically/morally important

The secrecy of jury deliberations is of political value. One argument for trial by jury is enabling ‘jury nullification’. This is where juries return not guilty verdicts based on moral conviction (because they disagree with the law or its application) rather than because they think that the accused did not break the law. Transcribing jury deliberations erodes the right to nullification by introducing oversight into the jury chamber.

Relying on nullification to reject real jury research is a somewhat awkward argument for a defender of the status quo to make, as the right of nullification is not officially recognised by most legal systems, even though it is something that juries can do.⁴⁷ Nevertheless, I do believe that the power of jury nullification can be an important strength of jury trials. In my view, real jury research should be conducted as part of an extended yet time-limited research programme. While such research ought to be repeated periodically (due to the temporal fragility of jury findings) it need not become the norm that jury deliberations are routinely transcribed. It would be possible to conduct such research while maintaining the general expectation of privacy for most jury deliberations. There’s no reason to accept slippery-slope reasoning from RJR to comprehensive state surveillance on legal trials.

More generally, whether the power of nullification is indeed a strength of jury trials is a delicate issue. It is one thing for an occasional drugs trial to be nullified due to jurors (sensibly) disagreeing with punitive narcotics laws, but quite another if juries are nullifying based on prejudice or general anti-authoritarian sentiment. For example, Jim Crow-era trials often found white juries nullifying attempts to bring perpetrators of racially-motivated crimes to justice. This is not a practice that we ought to be defending. Whether or not jury nullification is a sensible power to afford juries is, I think, not something that can be defended *a priori*—the answer will depend, to some extent, on how this power is in fact used by real juries. Therefore, to properly assess the merits of jury nullification, we require more information about its prevalence and motivations.

Beyond specific issue of nullification, and the relationship between privacy and candour discussed above, I don’t see any general argument for privacy as a weighty reason to prevent RJR. Data should be anonymised before being made accessible for academic study and we already accept that open justice sometimes trumps the interests of privacy (people can view trials in-person, journalists can write about them and case reports are made available for legal practitioners and academics to consult). Indeed, we often think that there is political value (namely democratic value) in publicising the grounds on which decisions are made and that, all else being equal, it is a problem when people are subject to important decisions without there being any way to scrutinise the reasons for the decision.

47. For this reason, Brooks (2004) refers to the ability to nullify as a ‘power’ rather than a right of the jury.

Objection four—RJR would reveal injustice

Keeping transcripts of juror deliberations raises the possibility that miscarriages of justice could be identified. This could: (i) undermine public confidence in the legal system, or (ii) generate disruptive appeals or compensation claims.

On its face, this strikes me as a peculiar type of objection. Uncovering otherwise hidden miscarriages of justice is a feature of a proposal that—all else equal—should be regarded as a strength rather than a weakness. If the criminal justice system is making mistakes or unsafe decisions as often as this objection presupposes, we should not take succour in the fact that such unsafe decisions or mistakes are currently hidden from view. Rather, we should be committed to reforming the system if mistakes are common—if this involves allowing additional appeals or compensation claims, then so be it. Rectifying otherwise hidden injustice should be viewed as an advantage rather than a weakness of the proposal. After all, if juries turned out to be so inaccurate as to lose public confidence or be patently unjust, it is not as if there is no alternative. Adjudication by professional judge and mixed judge/jury systems are viable models.

A different way to press the objection is put by Michael Zander (2013), who suggests that research about juries could give a *misleading* impression that the jury system was not performing well, even though it was generally in good shape (slips and howlers tend to make headlines above routine instances of conscientious deliberation). This objection could be dealt with by ensuring that researchers are bound by strict confidentiality agreements which would preclude them from revealing transcripts and information that could generate adverse publicity (and, indeed, from generating appeals or compensation claims). Moreover, it is already the case that some mock jury studies are undermining confidence in the jury system.

Objection five—RJR puts jurors at risk

Retaining transcripts of juror deliberations increases the chance that information will fall into the wrong hands and leave jurors open to reprisals.

First of all, we should note that only certain trials—prominently those involving organised crime or ‘crime families’—come with any non-trivial risk of reprisals. And we already sometimes treat such trials differently in various ways, e.g., hearing them without a jury or with special protection measures in place. We could simply exclude cases involving risk of reprisal from the scope of real jury research. Moreover, my proposal is for juror deliberation to be anonymised and only made available for academic scrutiny after a considerable time period (e.g., two years) has elapsed. This significantly reduces the chance that any particular juror could be identified from the transcript and that parties involved would retain an incentive for reprisals. Finally, securing sensitive data is certainly not a problem unique to jury research and not a problem that typically leads to an outright prohibition on scientific research. Rather, measures are adopted to safeguard information from falling into the wrong hands. For example, many academic institutions have training programmes and secure environments for accessing sensitive data that prevent the information from being transmitted to outside parties. Many university libraries have secure stations from which sensitive data can be accessed.⁴⁸

Serving on a jury already comes with the risk of reprisals, especially given that many jurisdictions require unanimity to convict (and so it will be apparent in a conviction case that a given juror voted to convict) and adhere to the principles of open justice. These risks are part and parcel of the current system. The added risk that would stem from permitting real jury research is negligible and there are practical steps that could be taken to mitigate this risk, just as we do when handling other types of sensitive information.

48. <https://safepodnetwork.ac.uk/>

Conclusion

Accurate and safe adjudication in criminal trials is of great moral importance. Every accuracy-based flaw in the trial process gives rise to an ethical complaint. To decide whether criminal justice systems that use juries are morally appropriate, we need to know how juries fare in generating accurate and safe results.

One article in the popular media describes the use of juries as a ‘long, messy experiment’.⁴⁹ This is inaccurate: we are not gathering the data required for it to count as an experiment. Rather, the use of juries is currently a black box. Current indirect methods for studying juries—mock juries, attitude surveys and post-trial survey—have their place in a mature jury science. But exclusive reliance on such indirect methods is entirely unsatisfactory: they allow us to investigate only a limited number of issues, there is conflict between extant results, and there are many reasons to worry about the ecological validity of results generated from these methods.

There is a moral imperative to engage in *real jury research*—allowing researchers to access the contents of real jury deliberations in order to assess concerns posed against jury trials. As this paper argued, the objections that have been raised against RJR turn out to lack force. This recommendation is especially pressing, since there has been much debate about reforming or even abolishing trial by jury—notably, in response to the problem of low conviction rates for sexual offences. Remedying our ignorance about the internal workings of jury trials *before* eliminating or reforming trial by jury is an eminently sensible thing to do. This can be achieved with a programme of real jury research.

Acknowledgements

For encouragement and very helpful feedback, I would like to thank audiences at a workshop on juries at the University of Glasgow and the CELPA group at the University of Warwick, along with various people who took the time to discuss this paper with me informally. The anonymous review was also valuable and improved the manuscript, especially regarding the breadth of empirical evidence.


Declaration of conflicting interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Lewis Ross  <https://orcid.org/0000-0001-9683-9924>

References

- Abramson J (2000) *We, the Jury. The Jury System and the Ideal of Democracy*. Massachusetts: Harvard University Press.
- Bindler A and Hjalmarsson R (2018) How punishment severity affects jury verdicts: Evidence from two natural experiments. *American Economic Journal: Economic Policy* 10(4): 36–78.
- Bornstein BH, Golding JM, Neuschatz J, et al. (2017) Juror sampling issues in jury simulation research: A meta-analysis. *Law and Human Behavior* 41(1): 13–28.
- Brennan J (2011) The right to a competent electorate. *Philosophical Quarterly* 61(245): 700–724.

49. <https://blogs.scientificamerican.com/news-blog/the-all-white-jury-v-the-diverse-ev/>

- Brennan J (2016) *Against Democracy*. New Jersey: Princeton University Press.
- Brooks T (2004) A defence of jury nullification. *Res Publica (Liverpool, England)* 10(4): 401–423.
- Chalmers J, Leverick F and Munro V (2021a) Why the jury is, and should still be, out on rape deliberation. *Criminal Law Review* 9: 753–771.
- Chalmers J, Leverick F and Munro VE (2021b) The provenance of what is proven: Exploring (mock) jury deliberation in Scottish rape trials. *Journal of Law and Society* 48(2): 226–249.
- Coen M and Doak J (2017) Embedding explained jury verdicts in the English criminal trial. *Legal Studies* 37(4): 786–806.
- Cullen HJ and Monds LA (2020) Jury simulation studies: To exclude or not to exclude participants based on a lack of comprehension of the case? *Applied Cognitive Psychology* 34(5): 1224–1233.
- Daftary–Kapur T, Dumas R and Penrod SD (2010) Jury decision–making biases and methods to counter them. *Legal and Criminological Psychology* 15(1): 133–154.
- Daly E, Smith O, Bows H, et al. (2022) Myths about myths? A commentary on Thomas (2020) and the question of jury rape myth acceptance. Epub ahead of print 6 January 2022. *Journal of Gender-Based Violence*. DOI: 10.1332/239868021X16371459419254.
- Daly G and Pattenden R (2005) Racial bias and the English criminal trial jury. *The Cambridge Law Journal* 64(3): 678–710.
- Devine DJ and Caughlin DE (2014) Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law* 20(2): 109–134.
- Diamond SS, Vidmar N, Rose M, et al. (2003) Juror discussions during civil trials: Studying an Arizona innovation. *Ariz. L. Rev* 45: 1.
- Grady RH, Reiser L, Garcia RJ, et al. (2018) Impact of gruesome photographic evidence on legal decisions: A meta-analysis. *Psychiatry, Psychology and Law* 25(4): 503–521.
- Hedden BR (2017) Should juries deliberate? *Social Epistemology* 31(4): 368–386.
- Horan J and Israel M (2016) Beyond the legal barriers: Institutional gatekeeping and real jury research. *Australian & New Zealand Journal of Criminology* 49(3): 422–436.
- Huemer M (2018) A defense of jury nullification. In: *The Palgrave Handbook of Philosophy and Public Policy*. Springer Verlag, 39–50.
- Hummel D and Maedche A (2019) How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics* 80: 47–58.
- Kaplan MF and Krupa S (1986) Severe penalties under the control of others can reduce guilt verdicts. *Law & Psychology Review* 10: 1–18.
- Kerr NL (1978) Severity of prescribed penalty and mock jurors' verdicts. *Journal of Personality and Social Psychology* 36(12): 1431–1442.
- Kim B (2017) Pragmatic encroachment in epistemology. *Philosophy Compass* 12(5): e12415.
- Kühberger A, Fritz A and Scherndl T (2014) Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE* 9(9): e105825.
- Landemore H (2012) *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. New Jersey: Princeton University Press.
- Lee Y (2018) The criminal jury, moral judgments, and political representation. *University of Illinois Law Review* 2018(4): 1255–1291.
- Leverick F (2020) What do we know about rape myths and juror decision making? *The International Journal of Evidence & Proof* 24(3): 255–279.
- Lundmark T (2010) 'Split verdicts' in Scotland: A judicial survey. *Edinburgh Law Review* 14(2): 225.

- Machery E (2022) Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews: Cognitive Science* 13(1): e1569.
- Mill JS (1975) *Three Essays: 'On Liberty,' 'Representative Government,' and 'The Subjection of Women'*. Edited by Richard Wollheim. New York: Oxford University Press.
- Mitchell TL, Haw RM, Pfeifer JE, et al. (2005) Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behaviour* 29(6): 621–637.
- Monaghan N (2018) Who should try 'complex fraud trials'? Reconsidering the composition of the tribunal of fact 30 years after roskill. In: *Financial Crime and Corporate Misconduct*. Abingdon, UK: Routledge, 176–191.
- Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science (New York, N.Y.)* 349: 6251.
- Pennington N and Hastie R (1991) A cognitive theory of juror decision making: The story model. *Cardozo L. Rev* 13: 519.
- Pennington N and Hastie R (1992) Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology* 62(2): 189–206.
- Pennington N and Hastie R (1993) The story model for juror decision making. In: Hastie R (ed) *Inside the Juror: The Psychology of Juror Decision Making*. New York: Cambridge University Press, 192–221.
- Ross Lewis (forthcoming) Criminal Proof: Fixed or Flexible. *The Philosophical Quarterly*.
- Schimmack U, Heene M and Kesavan K (2017) Reconstruction of a train wreck: How priming research went off the rails. Accessed: <https://replicationindex.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-off-the-rails/>.
- Slater J (forthcoming) Judge judge: The jury on trial. In (Ross, Egler and Bastian eds.) 'Philosophical dimensions of the trial' symposium for *American Philosophical Quarterly*.
- Sommers SR (2007) Race and the decision making of juries. *Legal and Criminological Psychology* 12(2): 171–187.
- Thomas C (2010) *Are Juries Fair?* Ministry of Justice Research Series 1/10.
- Thomas C (2013) Avoiding the perfect storm of juror contempt. *Criminal Law Review* 6: 483–503.
- Thomas C (2020) The 21st century jury: Contempt, bias and the impact of jury service. *Criminal Law Review* 2020(11): 987–1011.
- Tinsley Y, Baylis C and Young W (2021) 'I think she's learnt her lesson': Juror use of cultural misconceptions in sexual violence trials. *Victoria University of Wellington Law Review* 52(2): 463–486.
- van Aert RCM, Wicherts JM and van Assen MALM (2019) Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLOS ONE* 14(4): e0215052.
- Wiggins BJ and Christopherson CD (2019) The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology* 39(4): 202–217.
- Zander M (2013) Research should not be permitted in the jury room. *Criminal Law & Justice Weekly* 77(13/14): 215–216.