Helm Rebecca (Orcid ID: 0000-0003-1429-3847)

Prevalence Estimates as Priors: Juror Characteristics, Perceived Base Rates, and Verdicts in Cases Reliant on Complainant and Defendant Testimony

Rebecca K. Helm and Bethany Growns[1]

College of Social Sciences and International Studies, University of Exeter

[1] ORCiDs 0000-0003-1429-3847 and 0000-0002-6665-8134, respectively.

Prevalence Estimates as Priors: Juror Characteristics, Perceived Base Rates, and Verdicts in Cases Reliant on Complainant and Defendant Testimony

Abstract

Jurors often have to make decisions about whether they believe a complainant's or defendant's account of an event. However, the relative ambiguity of cues in testimony creates a situation where juror evaluations can vary significantly. As a result, in cases heavily reliant on testimony there is a particular likelihood that juror characteristics will be associated with verdicts, and it is important to understand these associations. This research investigates the relationships between two juror characteristics – gender and cultural worldviews – and verdicts in two such cases, and the potential for those relationships to be explained by differences in perceived prevalence of alleged events acting as prior probability judgments. As predicted, results show significant relationships between gender and cultural worldview and verdicts and show that these relationships are mediated by differences in underlying prevalence estimates. These findings have important implications for understanding associations between juror characteristics and verdicts and related policy.

Keywords: Juror decision-making, juror characteristics, juror bias, witness evidence, psychology and law.

Prevalence Estimates as Priors: Juror Characteristics, Perceived Base Rates, and Verdicts in Cases Reliant on Complainant and Defendant Testimony

In criminal cases, testimony from complainants and defendants is often important evidence. This importance is particularly clear in cases where the complainant and the defendant are the only witnesses to an alleged crime, as in many cases of sexual assault and domestic violence (see, for e.g., Hritz et al., 2015; Menaker & Cramer, 2012). Therefore, particularly as governments around the world seek to improve the handling of cases involving allegations of sexual assault (eg, Government Author, 2021; Lord Justice Clerk's Review Group, 2021; Violence Against Women Act Reauthorization Act, 2022), it is important to understand how jurors reach verdicts in cases in which testimony from a complainant and/or defendant is the primary evidence. This understanding will be important in determining how jurors are evaluating testimony, for policy relating to juror composition and judicial instructions, and to inform debates on whether jurors are appropriate decision-makers particularly in cases involving sexual assault (an issue that is and has been the subject of much debate across time and across jurisdictions, see McDonald, 2022, Chapter 1; see also Bindel, 2018).

One matter that it is important to understand is how jurors' own characteristics are associated with the judgments that they make about testimony given by defendants and complainants and, relatedly, their individual verdicts. Juror characteristics have the potential to be particularly important in cases where defendant and complainant testimony is central. While concrete statements can be made about the probative value of some evidence (e.g., random match probabilities can be given in cases involving DNA evidence), there are no currently known cues in defendant and complainant testimony that are reliable in indicating accuracy or inaccuracy (e.g., Bernstein & Loftus, 2009; Hartwig & Bond, 2011; Nortje & Tredoux, 2019). Even inconsistencies in testimony, sometimes thought to be clear indicators of deception or other inaccuracy, are now thought to tell us "little or nothing about the

accuracy of the…witness's testimony," apart from the specific inconsistent statement (Fisher et al., 2013, p. 178). There is therefore significant room for differences in interpretations, and, relatedly for associations between juror characteristics and their interpretations of evidence. These associations could be normatively desirable or undesirable heading into jury deliberations. On the one hand, research on jury decision-making has shown benefits of diversity in jury members (or discussants more generally) in promoting high-quality deliberation (e.g. Antonio et al., 2004; Bergold & Kovera, 2021). On the other hand, it is important to maximise individual juror neutrality and to avoid the influence of irrational bias on juror and jury decisions (see Devine & Caughlin, 2014). Therefore, it is important to understand when and why jurors with different characteristics tend to make different judgments. This paper examines relationships between two juror characteristics, juror gender and juror cultural worldview, and verdicts in two cases where complainant and/or defendant testimony is central, one case involving allegations of child sexual assault and one involving a domestic violence related defence to homicide. It investigates a potential mechanism to explain those relationships – differences in estimates of the prevalence of underlying events – and discusses the normative desirability of such an influence.

Existing Research on Juror Characteristics and Verdicts

Demonstrated Associations

A significant amount of work in applied psychology has demonstrated that at least in some circumstances, legal judgments are likely to differ systematically depending on the characteristics of jurors, including their gender (e.g., Bottoms et al, 2014), political beliefs (Kahan & Braman, 2008; Salerno et al, 2015), religious characteristics (Miller, 2014), age (Ruva & Hudak, 2013), and race (Mitchell et al., 2005). Importantly, this research suggests that the impact of juror characteristics is complex and can differ depending on case type and other factors (Devine & Caughlin, 2014).

In this paper, we focus on gender and cultural worldviews as two factors that have been clearly associated with changes in assessments of complainant and/or defendant believability and related verdicts that differ by case type in existing work (and thus cannot just be the product of general punitiveness/scepticism). For example, in the case of gender, work examining juror decision making in a patricide case involving allegations of child sexual abuse found that women made more pro-defendant judgments than men (Haegerich & Bottoms, 2000), but work examining allegations of child sexual assault found that women made more pro-complainant judgments than men (Bottoms et al., 2014). In the case of political beliefs, a variety of juror political and related opinions have been shown to be associated with different patterns in verdicts. For example, work examining the impact of juror cultural worldviews (composed of hierarchy and individualism, two measures associated with political opinions, capturing people's attitudes towards socially stratified roles and prioritisation of group and individual interests respectively, see Conway et al 2020; Kahan & Braman, 2006) has shown that in a case involving self-defence to homicide where a commuter had shot an allegedly threatening teenager those with more hierarchical and individualistic beiefs were more likely to find the defendant not guilty, but in a case involving self-defence to a homicide where a woman killed an allegedly abusive partner those with more hierarchical and individualistic beliefs were more likely to find the defendant guilty (Kahan & Braman, 2008). Authoritarian values have tended to be associated with more pro-prosecution judgments, but the size of the relationship has been shown to differ by case type (Devine & Caughlin, 2014).

Understanding why these juror characteristics are associated with verdicts, and, relatedly why they are associated with different verdicts in different case types, is important in assessing the normative desirability of the associations, and where intervention to reduce the associations may be necessary.

Existing Explanations

Existing attempts to explain relationships between juror characteristics and verdicts generally have focused on the fact that (per the 'Story Model' of juror decision-making, discussed below) jurors are thought to make judgments through formulating verbal narratives to explain evidence rather than through weighing information presented in a linear fashion (Pennington & Hastie, 1992). Differences in juror characteristics can lead jurors to adopt different stories based on the same evidence, and stereotypes relating to people and events will lead to a disposition to favour a certain account (Devine & Caughlin, 2014).

Experimental work has supported the idea that the role of different beliefs and stereotypes relating to actors in a case among men and women is one source of differences in verdicts from male and female jurors. For example, differing attitudes to child believability among men and women can partially explain gender differences in a case involving child sexual assault (Bottoms et al., 2014). This research has also suggested that gender differences might be driven by differences in empathy for complainants in a case, for example, work suggests that the finding that women tend to find in favour of victims more often in cases involving child sexual assault can be partially explained by the fact that women tend to empathise more with child complainants – women feel closer to child complainants and thus blame them less and blame defendants more (Bottoms et al., 2014).

There is another, unexplored, mechanism that can predict and explain relationships between gender and cultural worldview and legal judgments (including assessments of testimony accuracy) – differences in estimates of the prevalence of alleged events acting as relevant prior probability estimates.

Prevalence Estimates as Priors

The Importance of Prior Probability Estimates

Prior probability estimates are important in the evaluation of evidence, both normatively and descriptively. Within probability theory, Bayes' rule provides a mechanism for updating prior beliefs based on new evidence (see Donovan & Mickey, 2019, for a discussion in the legal context see Devine, 2012). According to this rule, decision-makers combine the perceived prior probability of a fact being true (e.g., the likelihood of a person having a disease, or having committed a crime) with new evidence (e.g., test results or case evidence) to form a posterior probability estimate (e.g., a medical diagnosis or legal judgment, in the medical context see Donlan et al., 1986). Rationally, prior probability estimates are important in decision-making. Take the frequently-cited 'cab' example (Tversky & Kahneman, 1980), in which 85% of taxis in a particular city are green and 15% are blue. A witness identifies the cab involved in a hit and run as being blue, but evidence shows that the witness only makes correct colour identifications in 80% of cases. Determining the likelihood that the cab was blue involves not only knowing how often the witness makes correct identifications, but also the likelihood of the taxi having been blue or green in the first place (the prior probability, or base rate). Thus, the correct probability of the cab being blue is only 41% rather than 80% (provided that the accuracy of the witness is independent of the proportion of cabs of each colour, see Birnhaum, 1983).

At least in some contexts, people tend to under-weight or even ignore prior probability estimates in comparison to new information (see Pennycook & Thompson, 2017). However, this insensitivity to prior probabilities appears to be dependent on task conditions, and these probabilities do have an influence on many decisions (for a review see Koehler, 1996). Research has provided important insight into when people are likely to be more influenced by prior probability estimates and when they are likely to neglect or underweight them. People are likely to rely more on prior probability estimates where they do not have sufficiently individuating information to distinguish a case from underlying probabilities, and

to rely less on such estimates where they have strongly individuating information (Epley & Dunning, 2000). Thus, where strong evidence (e.g. CCTV showing a person committing a crime) can individuate a situation from a prior probability, that prior probability is likely to become less important. However, where evidence is more ambiguous and less able to strongly individuate a situation from underlying prior probabilities, those underlying prior probabilities are likely to have more of an influence (i.e. the evidence will not pull people away from prior probabilities as strongly) (see also Shah et al., 2016). This relative difference in reliance on prior probabilities is rational. For example, imagine a patient presenting with symptoms of a very rare disease. If a diagnostic test with very high sensitivity and specificity indicates that they have the disease then they are likely to have the disease despite it being rare. However, if a diagnostic test with very weak sensitivity and specificity suggests the person has the disease it remains unlikely they have the disease, due to the rarity of the disease in the general population. The weak diagnosticity of the test means that its result should not pull decision-makers strongly away from underlying averages. Therefore, put simply, both prior probability estimates and new evidence should and often do inform beliefs, but their relative impact is determined by a range of factors including the extent to which the new evidence can individuate a case at hand from those prior probability estimates.

Existing Research on Prior Probability Estimates in Legal Contexts

Research seeking to descriptively model juror and jury decision-making suggests that jurors are not making decisions in line with a strict Bayesian framework (which should be viewed more as prescriptive than descriptive, see Devine, 2012). Rather, the leading model of jury decision-making, the 'Story Model', suggests that jurors reach decisions by imposing a narrative organisation on trial evidence (Pennington & Hastie, 1992). Jurors are thought to work through evidence, focusing on particular meaningful elements, and draw on their own knowledge as well as evidence to form and evaluate potential narratives (Pennington &

Hastie, 1992). Thus, they are not necessarily updating beliefs in light of each individual 'piece' of evidence, and are likely to neglect some pieces of evidence entirely and focus more on others in making their assessments of guilt.

However, prior probabilities can influence lay evaluations of evidence. This influence has been shown specifically in the case of evaluating the accuracy of the accounts of others. This work has shown that giving participants information about prior probabilities (e.g. telling them that of the statements they will hear 50% will be true and the other 50% will be false) does change their evaluations of the statements of others, specifically by decreasing the presence of 'truth bias,' where people tend to believe others are telling the truth more often than they actually are, despite not increasing the accuracy of evaluations (Domagalski, 2020; Kassin et al, 2005). Thus, prior probabilities may be playing a role in interpretations of evidence, although their influence is unlikely to be statistically correct in a Bayesian sense (see also Smith et al., 1996 on juror use of probabilistic evidence).

This influence of prior probabilities, consistent with the work on prior probabilities described above, can be accounted for within the framework of the Story Model. According to the model, a story's acceptance (and relatedly its role in informing verdict) is predicated on its coverage, coherence, and uniqueness (Pennington & Hastie, 1986; 1992). Coherence is composed of completeness, plausibility and consistency (Pennington & Hastie, 1992). It is important to note that the plausibility of a story, according to the model, is evaluated with general knowledge rather than with case-specific evidence (Majeed & Mahmood, 2021; Pennington & Hastie, 1992). A story is plausible where it adheres with general knowledge about how events tend to happen, and should accurately portray states and occurrences from the world (Bex, 2016). Although the story model itself does not mention prior probability estimates specifically, such estimates constitute knowledge about how events tend to happen

and thus are clearly relevant to plausibility. The more likely something is prior to hearing evidence (the higher the prior probability estimate), the more plausible it is that it happened.

Prior Probability Estimates, Testimony, and Verdicts

When interpreting testimonial evidence given by a witness and a defendant, there is no clear probabilistic evidence to rely on. Of course in the real world, jurors are not provided with a percentage probability that a defendant is lying. However, there are underlying base rates that have the potential to act as prior probability estimates – perceptions of how rare or common alleged events are (which we will refer to as prevalence estimates). If it is very common for people to make false allegations, for example, the prior probability of a particular allegation being false is higher. These estimates of commonality (acting as prior probabilities) naturally feed into plausibility judgments. The more common something is, the more plausible it is that it has happened. In the medical context, for example, it is more plausible that a person presenting to a hospital in the UK with a sore throat, fever, and nausea has a bacterial infection than it is that they have Polio (this sentiment is sometimes captured in medical training by the phrase "When you hear hooves, think horse not zebra", see Goldstein, 2017). That is because while bacterial infections are very common, Polio infections are very rare. In this way, the influence of prior probabilities described above is likely to feed into the decision-making process posited by the Story Model. Juries are likely to more readily accept stories incorporating events perceived as relatively common than stories incorporating events that are perceived as relatively rare.

Importantly, because of the lack of clear diagnostic cues in testimony to indicate (in)accuracy or (dis)honesty this influence of underlying prevalence estimates has the potential to be particularly important (since as noted above decision-makers will be less drawn away from underlying averages where they do not have sufficiently robust information to individuate cases at hand from those underlying averages). Thus, jurors might be more

willing to accept oral testimony from a complainant or defendant as convincing and sufficient to substantiate (or prevent) a conviction when it relates to an event that they perceive as being more common (and therefore more plausible). In this context, differences in underlying prevalence estimates have the potential to explain associations between juror characteristics and verdicts.

Prevalence Estimates as a Mechanism to Explain Gender and Worldview Differences

If prevalence estimates are important in influencing evaluations of testimony, differences in these estimates among jurors from different groups have the potential to predict and explain group differences in evaluations of testimony and related verdicts. In this paper, as discussed above, we will focus on associations between gender and verdicts and cultural worldview and verdicts, since these associations have been found by previous research to differ by case type, particularly in case types we are interested in where complainant and/or defendant testimony is important. Importantly, types of case where complainant and/or defendant testimony is often central, notably cases involving sexual assault and domestic violence, are also cases in which we expect to see clear differences in relevant underlying prevalence estimates. This overlap occurs because debates about these issues tend to be polarised by both political beliefs and gender (e.g. research has found that belief in 'sexual assault myths' is higher among Republicans than Democrats and among men than women, in the United States, see Ortiz & Smith, 2021).

Existing research provides insight into specific ways in which prevalence estimates relevant in these types of case may differ by gender and cultural worldview. First, estimates may differ as a result of differential experiences. Women are significantly more likely than men to have experienced sexual assault themselves (e.g., Elliott et al., 2004) and to have been victims of domestic abuse and controlling and coercive behaviours themselves (e.g., Walby & Towers, 2017), as a result, they may perceive these experiences as more common than men

do. Similarly, some work suggests people with more conservative ideology (ideology typically associated with hierarchical worldview, see e.g., Conway et al. 2020) are less likely to report sexual assault (Jose et al., 2021) suggesting that they have either experienced sexual assault less, or discuss it less, which could contribute to sexual assault appearing less prevalent to them and their peers. In addition, cultural cognition, or identity protective cognition, could result in men and those with hierarchical worldview, in particular, being more dismissive of information and news suggesting high levels of sexual assault and violence against women and being more receptive to information and news suggesting false allegations by women than women and those with greater beliefs in equality (for more information on identity protective cognition see Kahan, 2017, see also Kahan et al., 2011). These suggestions (at least in sexual assault cases) are supported by research in the United States context showing that men and Republicans (a political affiliation linked to more hierarchical and individualistic beliefs, Conway et al., 2020) are less likely than women and Democrats to view sexual assault as a serious issue in need of addressing (Ortiz & Smith, 2021).

Differences in prevalence estimates, if they exist, have the potential to explain relationships between gender and cultural worldviews and verdicts, particularly in cases involving sexual assault and domestic violence (including domestic violence related defences to homicide which depend partly on believing domestic violence took place).

Hypotheses

In this paper, we will examine associations between gender, cultural worldview, underlying prevalence estimates, and verdicts in two cases centred around complainant and/or defendant testimony (one involving child sexual assault and the other involving a domestic violence related defence to homicide). Each participant will make judgments relating to both cases. Based on the rationale outlined above, we expect that men and those

with more hierarchical and individualistic beliefs will rate the prevalence of sexual assault and harm from domestic abuse as lower and the prevalence of false allegations and fabrication of harm as higher than women and those with less hierarchical and individualistic beliefs will rate them, and that this will contribute to men and those with more hierarchical and individualistic beliefs making more pro-defendant judgments in a case of alleged sexual assault, and less pro-defendant judgments in a case of a domestic violence related defence to homicide. These predictions will be tested by looking at associations between gender, cultural worldview and prevalence estimates, and gender, cultural worldviews and verdicts, and by testing whether identified differences in verdicts are mediated by differences in prevalence estimates. We have conceptualised gender and cultural worldviews as mediators rather than moderators in this work, since predictions relate to the processes through which those variables are related to verdicts, rather than suggesting that the strength or direction of the relationship between a specific prevalence estimate and verdict would change based on gender or cultural worldview (e.g. a view that an event is prevalent, if held, would be expected to increase belief that that event occurred for both men and women rather than to have a different effect on men than on women). These predictions were all pre-registered on OSF (https://osf.io/ayp6k).

Method

Participants, Approvals, and Pre-Registration

Participants ($N = 539$; mean age $= 37.61$ years; $SD = 13.60$, 72.5% female, 27.5% male) were recruited through Prolific (a platform shown to produce high-quality data and filter out bots and inattentive participants, see Peer et al., 2017). Power analyses using WebPower (Zhang & Yuan, 2018) showed that this sample size would be sufficient for mediation analyses with path coefficients of .2 (previous work suggests a medium effect size when examining relationships between demographic variables and verdicts, see Kahan &

Brahman, 2008; Bottoms et al., 2014), variance values of 1, and power of 80% (which would require a sample of 393; we planned to recruit 500 to account for potential gender differences in the sample). The survey included 7 attention check questions. All participants correctly passed our pre-registered inclusion criteria of answering at least 5 out of the 7 attention check questions correctly and were all included in analyses. The study received all necessary ethical approvals. The experiment was preregistered, and the preregistration, data, and materials are available at https://osf.io/ayp6k.[2]

Materials and Procedure

Each participant completed three sets of tasks, described below, with the order of the first two tasks counterbalanced (so that half of the participants gave prevalence estimates before the applied judgment task, and half of the participants gave prevalence estimates after completing the applied judgment task).[3] Participants who received the prevalence estimates first completed the prevalence estimates then made judgments in the two cases (either child sexual assault followed by homicide or homicide followed by child sexual assault) and then answered questions on cultural worldview and other demographics. Participants who received the applied judgment task first made judgments in the two cases (either child sexual assault followed by homicide or homicide followed by child sexual assault), then completed the prevalence estimates, and then answered questions on cultural worldview and other

---

[2] Note that our analyses deviated from the pre-registered plan in that we used linear and logistic regression and generalized estimating equation models in analyses, rather than mixed effects regression models.

[3] Results did indicate an effect of order on prevalence estimates but not on verdicts, detailed in the accompanying supplemental materials. Where participants saw cases first they gave significantly higher estimates of the prevalence of some events. Case verdicts did not differ significantly based on when they were delivered before or after prevalence estimates. To account for this potential order effect and to ensure it was not driving our results we did two things. First, we ran all analyses controlling for order. With the exception of some small differences in the significance of values in our analyses relating gender and hierarchy and prevalence estimates all significant results (including the results of our mediations) remained the same. Second, we conducted all analyses with only participants who gave prevalence estimates before verdict and therefore were not influenced by identified order effects. Despite reduced power, all significant results were replicated in this analysis. Full details and results of these analyses is provided in the accompanying supplemental materials.

demographics. A summary of conditions with accompanying sample sizes is presented in Table 1.

Table 1: *Sample split by gender and condition.*

Gender	Order	Case seen first	Initial impression question	N in condition
Female	Cases first	Sexual assault	Yes	46
			No	57
		Homicide	Yes	53
			No	48
	Estimates first	Sexual assault	Yes	44
			No	59
		Homicide	Yes	53
			No	31
Male	Cases first	Sexual assault	Yes	17
			No	15
		Homicide	Yes	20
			No	19
	Estimates first	Sexual assault	Yes	17
			No	17
		Homicide	Yes	17
			No	26

Applied Judgment Task

Participants read the testimony of a complainant and/or defendant in two vignettes involving accusations of criminal behaviour. One vignette involved a girl accusing her teacher of sexual assault (the child sexual assault case), and the other vignette involved a woman who killed her husband now trying to use a defence based on being a victim of domestic abuse (the homicide case). The vignettes were presented as testimony in legal cases. A brief summary of each side's position relating to what had happened was given, participants then read testimony in each case, indicated a case verdict, and provided ratings of the testimony.

In the child sexual assault case, participants read testimony from a complainant (the person making the accusation) and a defendant (the person the allegation was made against).

The complainant claimed that she had been inappropriately touched by the defendant. The defendant claimed that he had never touched the girl inappropriately and had acted professionally at all times. In the homicide case, participants read testimony from the defendant only (since the only other person with direct experience of the event had been killed). The defendant claimed that she had killed her husband as a result of a recognised psychological disorder brought on by many years of mental and emotional abuse from him. The prosecution claimed that she was fabricating her psychological illness as an excuse for killing her husband, which she did out of anger and a desire for revenge. Note, the prosecution argument was not presented as testimony, since the prosecution lawyer did not view the event or have direct knowledge of what happened.

Participants read each vignette (with order counterbalanced), they then indicated how believable, consistent, confident, reliable, honest, and accurate they thought each protagonist was on an eleven-point Likert scale from 0 (not at all) to 10 (extremely), made a legal judgment of guilt (verdict) and answered some other questions about their perceptions of the case. Half of the participants were also randomly assigned to give their impression of who they believed before seeing the testimony, based only on the initial brief summary of each side's position.[4] For the purposes of this paper, the relationship between prevalence estimates and verdicts is examined. Results for other dependent measures overlapped significantly with verdict, and are presented in the accompanying supplemental materials.

Prevalence Estimates

Participants answered questions about their perceptions of how common events of the type alleged by each protagonist in our vignettes were. Specifically, they rated how often

---

[4] This question was included to allow exploratory analyses of the point in the process at which characteristics were (or were not) associated with verdicts. Responses are not analysed for the purposes of this paper but supplementary analyses were run replicating our main analyses and controlling for whether this question was present (alongside order). These analyses suggest that the presence or absence of this question did not influence our significant results. Full analyses and results are provided in the accompanying supplemental materials.

girls are sexually abused by people in positions of trust (relevant to the child sexual assault vignette, abbreviated to 'prevalence of child sexual assault'), how often girls make false allegations of sexual assault (relevant to the child sexual assault vignette, abbreviated to 'prevalence of false allegations'), how often women suffer psychological harm as a result of abuse by their partners (relevant to the homicide vignette, abbreviated to 'prevalence of psychological harm'),[5] and how often women fabricate psychological harm as an excuse for bad judgment (relevant to the homicide vignette, abbreviated to 'prevalence of fabrication of psychological harm').

Participants gave an evaluation of prevalence, rating how often each event occurs on an eleven-point scale from 0 (this never happens) to 10 (this happens extremely often). This scale was used in all analyses reported below. Participants were also asked to indicate which specific prevalence most closely reflected their belief on an 8-point scale from the event happening to (or being done by) more than one in every ten people to less than one in every million people (e.g., more than one in every ten girls, about one in every ten girls, about one in every hundred girls, about one in every thousand girls, about one in every ten thousand girls, about one in every hundred thousand girls, about one in every million girls, less than one in every million girls). Responses on this scale correlated strongly with responses on our 11-point scale, showing that responses on our scale reflect differences in numerical estimates not just differences in what is considered low or high (child sexual assault $rho(53)$=-.523, p<.001; false allegations $rho(53)$=-.525, p<.001, psychological harm $rho(53)$=-.469, p<.001, fabricated harm $rho(53)$=-.451, p<.001).

After giving these estimates, participants were asked to provide a brief description (25 – 50 words) of how they had reached them. These descriptions demonstrated participants had

---

[5] We also asked participants to estimate the prevalence how often women are abused by their partners generally, but analyses focused on the more specific estimate as it was more directly applicable to the relevant homicide vignette.

engaged with and undersood the questions correctly (although a small number of participants did note difficulty understanding the question relating to fabricating psychological harm), and are available to review on OSF.

Cultural Worldview and Demographics

After giving prevalence estimates and applied judgments, participants completed the Cultural Worldview Scale (Kahan & Braman, 2008). The scale contains two subscales – Hierarchy and Individualism. The Individualism scale contains 17 items and measures the relative priority that participants assign to group and individual interests as opposed to communitarianism (e.g., 'The government should do more to advance society's goals, even if that means limiting the freedom and choice of individuals'; $\alpha = .85$). The Hierarchy scale consists of 13 items assessing participants' attitudes towards socially stratified roles (hierarchy vs. egalitarianism; e.g., 'We have gone too far in pushing equal rights in this country'; $\alpha = .91$). Responses are on a seven-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). These scales represent two dimensions of cultural worldview and have been verified as highly reliable measures of the latent dispositions of subjects towards those types of ideology (Kahan & Braman, 2008). Scores on the scales were highly correlated – $r(539)=.62, p<.001$ – and showed many of the same relationships with other variables. Primary analyses in this paper, therefore, focus on the hierarchy scale (as the more reliable scale) to avoid duplication, but analyses involving the individualism scale are reported in accompanying supplemental materials. Finally, participants answered some demographic questions, including indicating their gender, and some questions on other individual differences.

Results

Relationships Between Gender, Hierarchy, and Prevalence Estimates

Prevalence Estimates Descriptive Statistics

Specific prevalence estimates of participants were examined. Ratings for each occurrence are displayed in Table 2, from what was rated as most common to what was rated as least common.

Table 2: *Descriptive for prevalence estimates.*

Occurrence	*M* (min-max)	*SD*
Prevalence of psychological harm	6.88 (1-10)	2.13
Prevalence of child sexual assault	5.00 (0-10)	2.51
Prevalence of false allegations	3.91 (1-10)	2.31
Prevalence of fabrication of psychological harm	3.41 (0-10)	2.04

Note. Prevalence estimates were given on an 11-point scale ranging from 0 (this never happens) to 10 (this happens extremely often).

Gender, Hierarchy, and Prevalence Estimates

A series of four linear regressions were used to examine the associations between gender and hierarchy and each of our four prevalence estimates. Specifically, gender and hierarchy were entered as predictors of each of the prevalence estimates. In line with predictions, in the regressions predicting psychological harm prevalence ($R^2 = .03$) and child sexual assault prevalence ($R^2 = .02$), both hierarchy and gender were significant predictors of estimates. As hierarchical beliefs increased, ratings of the prevalence of psychological harm decreased ($B = -.18$, $SE = .09$, $\beta = -.09$, $t(536) = -1.99$, $p = .048$) (for every one unit increase in hierarchy, average prevalence ratings decreased by .18 units) and ratings of the prevalence of child sexual assault decreased ($B = -.21$, $SE = .11$, $\beta = -.09$, $t(536) = -1.98$, $p = .049$) (for every one unit increase in hierarchy, average prevalence ratings decreased by .21 units). Women rated the prevalence of psychological harm resulting from abuse ($B = .61$, $SE = .21$, $\beta = .13$, $t(536) = 2.95$, $p=.003$) and the prevalence of child sexual assault ($B = .61$, $SE = .21$, β

= .13, t (536) = 2.14, p=.003) as higher than men did. Moving from women (the reference group) to men corresponded to an increase of .61 units in ratings of both the prevalence of both psychological harm, and the prevalence of child sexual assault.

Also in line with predictions, in the regressions predicting the prevalence of false allegations (R^2 = .16) and the prevalence of fabrication of psychological harm (R^2 = .13), hierarchy was a significant predictor. As hierarchical beliefs increased, ratings of the prevalence of false allegations increased (B = .92, SE = .09, β = .40, t (536) = 9.96, $p < .001$) (for every one unit increase in hierarchy, average prevalence ratings increased by .92 units), as did ratings of the prevalence of harm fabrication (B = .73, SE = .08, β = .36, t (536) = 8.89, $p < .001$) (for every one unit increase in hierarchy, average prevalences ratings increased by .73 units). No significant associations between gender and these two prevalence ratings were identified.

Summary

All of our predictions relating to hierarchy were supported. Participants with more hierarchical beliefs rated the prevalence of psychological harm and child sexual abuse as less prevalent and false allegations and fabrication of psychological harm as more prevalent than participants with less hierarchical beliefs did. Two of our predictions relating to gender were supported. Female participants rated the prevalence of psychological harm and child sexual assault as more prevalent than male participants did. We did not see significant gender differences in the perceptions of the prevalence of false allegations or fabrication of psychological harm).

Relationships Between Gender and Hierarchy and Verdicts

Gender, Hierarchy, and Verdicts

A generalized estimating equation logit model (GEE) was used to analyze the relationships between gender, hierarchy, and verdicts, and how those relationships differed

by case. The model used case type (within-subjects factor), participant gender (between-subjects factor), participant hierarchy (covariate) and the interaction between case type and participant gender and case type and participant hierarchy score to predict case verdict (guilty or not guilty).

The GEE revealed two significant main effects. First, it revealed a significant main effect of case type, such that the odds of a participant finding the defendant guilty in the child sexual assault case were about 80% lower than the odds of a participant finding the defendant guilty in the homicide case (B = -1.57, SE = .34, OR = .21, 95% CI [.11, .40], p < .001). Second, it revealed a significant main effect of hierarchy, such that for each unit increase in hierarchy score, the odds of finding the defendant guilty decreased by approximately 40% (B = -.50, SE = .09, OR = .61, 95% CI [.51, .73], p < .001).

The GEE also revealed the predicted significant interactions between gender and case type (B = .78, SE = .28, OR = 2.19, 95% CI [1.26, 3.79], p = .005) and hierarchy and case type (B = .71, SE = .13, OR = 2.03, 95% CI [1.59, 2.59], p < .001). In terms of gender, follow-up logistic regression analyses using gender to predict verdict found that in the child sexual assault case (Nagelkerke R^2 = .02) the odds of a female participant finding the defendant guilty were about 77% higher than the odds of a male participant finding the defendant guilty (B = .58, SE = .20, OR = 1.77, p = .005), but in the homicide case (Nagelkerke R^2 = .01) the odds of a female participant finding the defendant guilty were about 30% lower than the odds of a male participant finding the defendant guilty, although note that this effect just missed significance (B = -.38, SE = .20, OR = .69, p = .053).

In terms of hierarchy, follow-up logistic regression analyses using hierarchy to predict verdict found that in the child sexual assault case (Nagelkerke R^2 = .02) participants with higher levels of hierarchy were less likely to find the defendant guilty (B = -.23, SE = .09, OR = .79, p = .009; for every one unit increase in hierarchy, the odds of a participant finding the

defendant guilty decreased by approximately 20%) and in the homicide case (Nagelkerke R^2 = .08) participants with higher levels of hierarchy were more likely to find the defendant guilty ($B = .51$, $SE = .09$, $OR = 1.67$, $p < .001$; for every one unit increase in hierarchy, the odds of a participant finding the defendant guilty increased by approximately 67%).

Summary

Our regression analyses, therefore, showed a number of significant relationships between gender and hierarchy, and verdicts. Specifically: (1) in the child sexual assault case, women were more likely to find the defendant guilty than men, (2) in the child sexual assault case, participants scoring higher on the hierarchy scale were less likely to find the defendant guilty, (3) in the homicide case, participants scoring higher on the hierarchy scale were more likely to find the defendant guilty.

Examining Relationships: Mediation Analyses

Next, we examined whether relationships between gender and hierarchy and prevalence estimates would mediate relationships between gender and hierarchy and verdict. Analyses were conducted for all significant relationships between gender and verdict and hierarchy and verdict. Hypothesised mediations were tested using a bootstrapping approach using the Process macro in SPSS (model 4) (see Hayes, 2022). Unstandardized indirect effects were computed for each of 5,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. Regression coefficients are unstandardized and expressed in a log-odds metric. Mediation figures were all created using semdiag (Mai et al., 2018).

Gender and Child Sexual Assault Verdicts

The relationship between gender and verdicts in the child sexual assault case was mediated by rated prevalence of sexual assault of girls (see Figure 1). We tested the significance of the indirect effect using bootstrapping procedures. The bootstrapped

unstandardized indirect effect was .09 (95% CI [.01, .19]). Thus, the indirect effect was statistically significant. This analysis was re-run with hierarchy as a covariate and all significant and non-significant results remained the same.
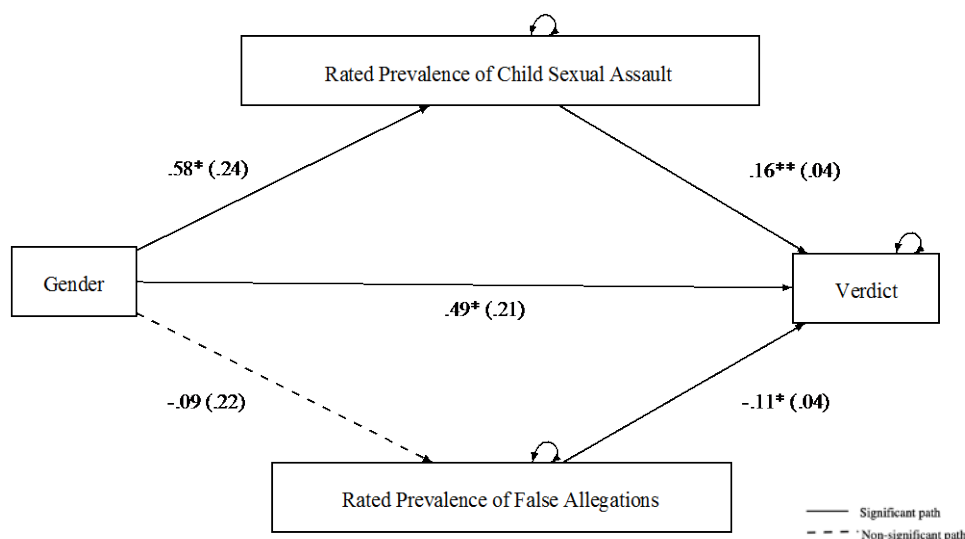

Figure 1: Mediation of the relationship between gender and verdict by rated prevalence of sexual assault of girls. Coefficients are unstandardized, *p < .05, **p < .001. Circular arrors represent sources of variation not included in the model (i.e. error terms).

Hierarchy and Child Sexual Assault Verdicts

The relationship between hierarchy and verdicts in the child sexual assault case was mediated by rated prevalence of sexual assault of girls and rated prevalence of false allegations made by girls (see Figure 2).

We tested the significance of indirect effects using bootstrapping procedures. The bootstrapped unstandardized indirect effect of hierarchy on verdict through rated prevalence of child sexual assault was -.04 (95% CI [-.08, -.01]). Thus, this indirect effect was statistically significant. The bootstrapped unstandardized indirect effect of hierarchy on verdict through rated prevalence of false allegations was -.09 (95% CI [-.17, -.01]). Thus, this indirect effect was also statistically significant. These analysis were re-run with gender as a covariate and significant results remained the same.
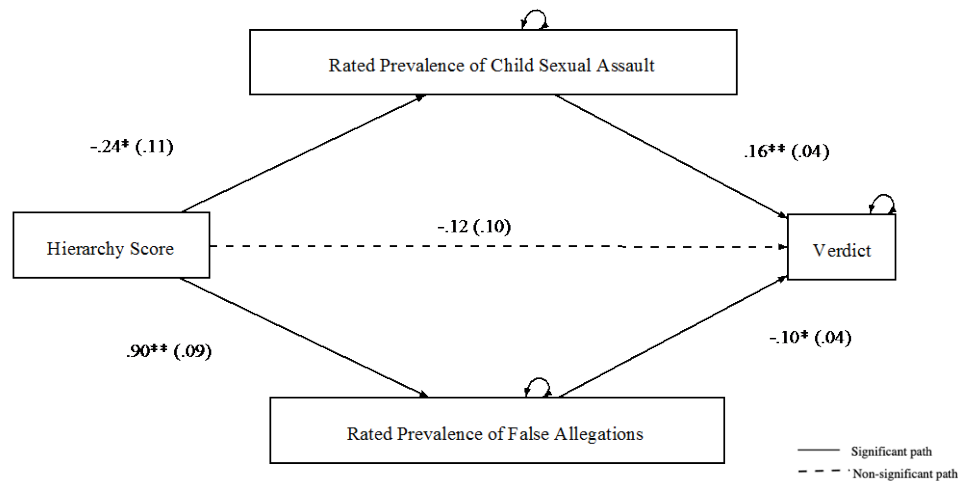
Figure 2: Mediation of the relationship between hierarchy and verdict by relevant prevaence ratings. Coefficients are unstandardized, *p < .05, **p < .001. Circular arrors represent sources of variation not included in the model (i.e. error terms).

Hierarchy and Homicide Verdict

The relationship between hierarchy and verdicts in the child sexual assault case was mediated by rated prevalence of fabrication of psychological harm (see Figure 3).
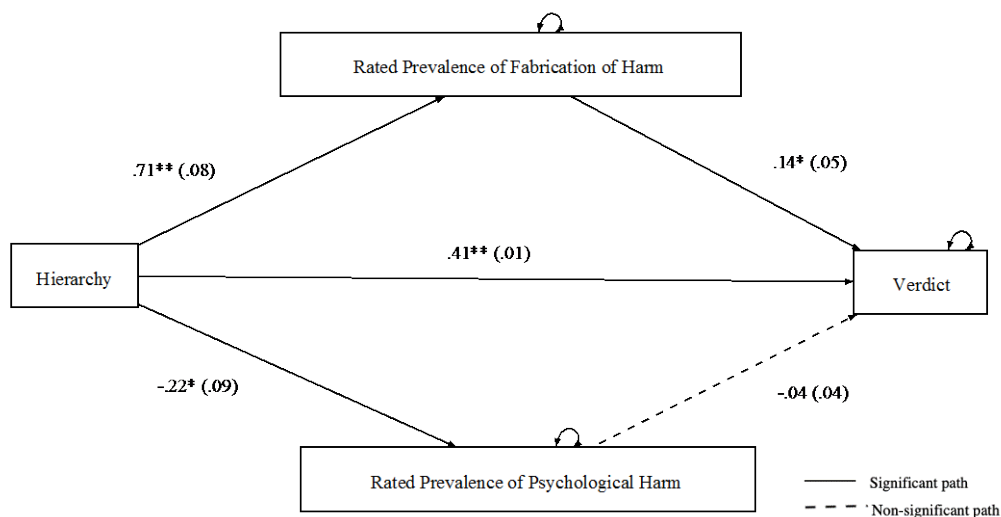

Figure 3: Mediation of the relationship between hierarchy and verdict by relevant prevaence ratings. Coefficients are unstandardized, *p < .05, **p < .001. Circular arrors represent sources of variation not included in the model (i.e. error terms).

We tested the significance of indirect effects using bootstrapping procedures. The bootstrapped unstandardized indirect effect of hierarchy on verdict through rated prevalence

of the fabrication of harm was .10 (95% CI [.03, .18]). Thus, this indirect effect was statistically significant. The bootstrapped unstandardized indirect effect of hierarchy on verdict through rated prevalence of psychological harm was .01 (95% CI [-.01, .04]). Thus, this indirect effect was not statistically significant. These analyses were re-run with gender as a covariate and all significant and non-significant results remained the same.

Mediation Models Summary

Our mediation analyses therefore found that the significant association between gender and verdict and both significant associations between hierarchy and verdict were mediated by relevant prevalence estimates. The relationship between gender and verdict in the child sexual assault case was mediated by ratings of the prevalence of child sexual assault, the relationship between hierarchy and verdict in the child sexual assault case was mediated by ratings of the prevalence of child sexual assault and false allegations, and the relationship between hierarchy and verdict in the homicide case was mediated by ratings of the prevalence of fabrication of psychological harm.

Discussion

The results of this study provide insight into relationships between demographic variables (specifically gender and cultural worldview) and a potential causal mechanism underlying those relationships – differences in perceptions of prevalence of relevant events. Results show that ratings of the prevalence of legally important events significantly differ based on gender and cultural worldviews (including hierarchy, reported in our main analysis, and individualism, reported in the supplementary materials) and that these differences mediate relationships between gender and cultural worldviews and verdicts in cases reliant on testimony from a complainant and / or defendant. Results of the study, summarised in more detail below, provide insight in three key areas: (1) how demographic variables might contribute to differences in the perceived prevalence of legally important events, (2) the

relationships between juror characteristics and verdicts, and (3) how differences in legally important prevalence estimates can mediate relationships between juror characteristics and verdicts.

Prevalence Estimates

This research is the first study, to our knowledge, to examine demographic differences in estimates of the prevalence of legally relevant events. Women and those with more hierarchical beliefs rated the prevalence of child sexual assault and prevalence of psychological harm from abuse as higher than men and those with less hierarchical beliefs did, and those with more hierarchical beliefs and those with more individualistic beliefs rated the prevalence of false allegations and the fabrication of psychological harm as higher than those with lower hierarchical and individualistic beliefs did. In some ways, these results are unsurprising due to the sometimes polarised nature of discourse surrounding these issues (e.g., Ortiz & Smith, 2021).

On the one hand, since ground-truth prevalence is impossible to know, it might be argued that it is reasonable for different people to reach different estimates relating to prevalence, and that jury deliberations will ensure all viewpoints are reflected to reach the most reasonable conclusions. However, it is not entirely clear that this is the case, and the risk that biases resulting from personal experience and information consumption (which may be a systematic problem influencing more than one juror, for example resulting from high profile media coverage of rare events) is influencing these estimates should be noted. Potentially high-quality information and education can help in reducing such biases, and ensuring appropriately informed estimates. Organisations such as Rape Crisis and the National Association for the Prevention of Cruelty to Children in the UK already produce materials attempting to educate people on the best current information about prevalence (Rape Crisis,

nd; NSPCC, 2021). More appropriately informed estimates in this area are important politically as well as legally.

Juror Characteristics and Verdicts

This research also contributes to a body of evidence showing that juror characteristics, specifically gender and cultural worldview, are associated with verdicts and that this association can vary by case type. In our child sexual assault case, men and those with more hierarchical beliefs were less likely to find the defendant guilty than women and those with more hierarchical beliefs, and in our homicide case, those with more hierarchical and individualistic views were more likely to find the defendant guilty than those with less hierarchical and individualistic views (it should be noted that qualitatively women were less likely to find the defendant guilty than men, but this effect just missed statistical significance).

These findings replicate and extend previous work examining relationships between gender and cultural worldview and verdicts. Specifically, they demonstrate relationships between these characteristics and verdicts that are in line with those demonstrated in previous work (specifically men making more pro-defendant determinations in cases involving child sexual assault; Bottoms et al., 2015, and men and those with more hierarchical and individualistic worldviews reaching more pro-prosecution determinations in cases involving a domestic violence related defence to homicide; Kahan & Braman, 2008). They also support the prediction that those with more hierarchical and individualistic views are likely, at least in ambiguous cases, reach more prosecution-friendly judgments in cases involving child sexual assault. The work therefore contributes to the literature examining associations between juror characteristics and verdicts, discussed above.

Results also contribute to policy-debates surrounding gender balance on juries, by highlighting the importance of this balance. In the case of gender in particular, an imbalance

is not uncommon on modern juries. For example, in 2017 a study in Ireland showed that men dominated 57% of juries (with women dominating only 17%) (Gallagher, 2017). This number rose to 61% in rape trials (Gallagher, 2017). Such imbalances have the potential to be harmful in allowing an influence of bias or even just a perspective that may be present in one group and over-represented when that group dominates on a jury, and should be considered particularly important where prevalence estimates may be polarised and where complainant and / or defendant testimony is central.

Prevalence Estimates as Mediators of Associations

Importantly, results also show that differences in underlying prevalence estimates significantly mediate all demonstrated differences in verdict. All differences in verdict by gender and hierarchy and individualism were mediated by underlying prevalence estimates (although direct effects also remained in the majority of cases, and future work should continue to examine other factors important in understanding these associations). These findings are important in providing new understanding relating to associations between juror characteristics and verdicts in this area. The findings make sense within the context of the Story Model, where juror perceptions of the plausibility of events in competing narratives is thought to contribute to evaluations of competing narratives (with more plausibility indicating greater coherence, see Pennington & Hastie, 1992). However, the fit of prevalence estimates within the Story Model framework was not considered by this work, and should be examined more precisely in the future.

These findings suggest that to some extent associations between gender and cultural worldview and verdicts may be the result of rational thought processes, based on underlying prevalence estimates. These estimates may also be important in understanding other demonstrated associations between verdicts and characteristics, discussed above, including those involving race (Mitchell et al., 2005), age (Ruva & Hudak, 2013), and religious

characteristics (Miller, 2014). An important question in this context is the extent to which an influence of prevalence estimates, which vary as a result of juror characteristics, is normatively desirable. Variations in prevalence estimates could be desirable – when jurors with different perspectives come together different perspectives based on different perceptions of commonality are reflected and debated in the deliberation process to reach reasonable conclusions. However, the extent to which the influences are normatively acceptable depends on what informs underlying prevalence estimates and whether those estimates are sufficiently and appropriately individuated to a particular case (e.g., Welsh & Navarro, 2012). While the jury system specifically relies on jurors drawing on personal experience, this may be inappropriate where a juror's experience base is drastically different from that of the people they are making judgments about.

Unreasonable and uninformed views about relevant prevalence, for example continuing views that any kind of sexual assault is extremely uncommon, should be viewed in the same way as problematic beliefs such as rape myths, as damaging misconceptions to be corrected. Our results show huge variations in estimates of prevalence, indicating that at least some jurors opinions are likely to be problematic, although these problematic views may be tempered by jury deliberation. Better education surrounding prevalence may help to reduce the risk of problematic views being reflected in jury decisionmaking, and to result in more informed decision making as a result. Education may be delivered in the form of judicial instructions where there is a risk of significant error, however the precise figures to give in such instructions would have to be given significant consideration particularly since ground truth in this area is difficult to determine (although see attempts discussed above). If figures were given, these figures would need to be developed independently and their influence tested empirically prior to introduction. Additionally, an important remaining question for the legal system is the extent to which prevalence estimates should be influencing juries at

all. Any influence of such biases is problematic for those who experience relatively rare events (the "zebras" in the legal system).

Limitations and Conclusions.

These results should be interpreted in light of some important limitations, particularly due to the experimental nature of the work. First, the study has limitations in terms of ecological validity. Participants saw relatively brief case materials that were presented to them in written format, and each made decisions in two cases, one after the other (something that would not happen in practice). In addition, participants gave individual decisions and thus the impact of deliberations on verdicts was not accounted for. It should also be noted that our homicide case only had a single witness, which limits ecological validity, and that our homicide case and child sexual assault case differed in several ways that could limit generalisability – as well as only having a single witness, our homicide case contained arguments from the prosecution in place of a prosecution witness. No comparable arguments were included in our child sexual assault case. Future research should examine whether the relationships between prevalence estimates and evaluations remain constant when testimony is given orally and in a more realistic case context, and how individual prevalence estimates are translated into verdicts in deliberating and diverse groups. While our work taps into phenomena that are cognitively important, the specific sizes of these effects could be larger or smaller in a more realistic context. The impact of deliberations is particularly important in this area since juries are likely to consist of a mixture of genders and cultural worldviews.

Another limitation of the study was that the link between prevalence estimates and applied decisions could have been enhanced by our experimental design since the same participants were asked for both judgments. This possibility seems unlikely since we counterbalanced these judgments to control for order and to minimize priming effects, and replicated all analyses in our condition in which people gave prevalence estimates prior to

seeing the cases where we had no significant order effects. Nevertheless, future research might consider embedding prevalence estimates in other tasks to make them less obvious to participants and further reduce this possibility. Finally, our study tested the relationship between juror characteristics and verdicts in two specific cases where an effect of prevalence estimates was expected. Future research should examine which other associations might be mediated by these estimates, and which associations may have an entirely independent cause. Research should also consider examining the ways that gender and hierarchy interact with each other in predicting prevalence estimates and verdicts, which was not explored here.

Despite these limitations, the findings have important implications for future research and, ultimately, policy, for the reasons discussed above. Careful consideration needs to be given to whether providing prevalence information to jurors could be helpful in some cases, and the role that education might have in improving decision-making, particularly where jurors are making judgments about people who are different from themselves in important ways. In these cases, providing education directly in the jury room could be beneficial (subject to careful development and testing of instructions, discussed above). There is some precedent for courts accepting the use of base rates in other contexts, particularly where it is difficult or impossible to obtain evidence of a more individuating sort (see Koehler, 2002). This precedent can be drawn on in designing effective and normatively desirable procedural interventions in this area. Finally, it should be noted that perhaps the diversity of juries and the combining of different estimates (when done appropriately) could potentially give juries an advantage over judges in this area, who lack this diversity and are also known to be susceptible to decision-making biases (e.g., Guthrie et al., 2007). This possibility should be borne in mind in evaluating suggestions to replace juries with judges in certain classes of case (e.g., Lord Justice Clerk's Review Group, 2021).

REFERENCES

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science, 15*(8), 507-510. https://doi.org/10.1111/j.0956-7976.2004.00710.x.

Bergold, A. N., & Bull Kovera, M. (2021). Diversity's Impact on the Quality of Deliberations. *Personality and Social Psychology Bulletin*, 01461672211040960. https://doi.org/10.1177/01461672211040960.

Bernstein, D. M., & Loftus, E. F. (2009). How to tell if a particular memory is true or false. *Perspectives on Psychological Science, 4*(4), 370-374. https://doi.org/10.1111/j.1745-6924.2009.01140.x.

Bex, F. (2016). Analysing stories using schemes. In H. P. Hendrik Kaptein and Bart Vergeij (Eds.). *Legal Evidence and Proof: Statistics, Stories, Logic.* Routledge.

Bindel, J. (2018, November 21). 'Juries have no place in rape trials. They simply can't be trusted' *The Guardian.* https://www.theguardian.com/commentisfree/2018/nov/21/juries-rape-trials-myths-justice.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *The American Journal of Psychology*, 85-94. https://doi.org/10.2307/1422211.

Bottoms, B. L., Peter-Hagene, L. C., Stevenson, M. C., Wiley, T. R., Mitchell, T. S., & Goodman, G. S. (2014). Explaining gender differences in jurors' reactions to child sexual assault cases. *Behavioral Sciences & the Law, 32*(6), 789-812. https://doi.org/10.1002/bsl.2147.

Conway III, L. G., Chan, L., & Woodard, S. R. (2020). Socio-ecological influences on political ideology. *Current Opinion in Psychology*, *32*, 76-80. https://doi.org/10.1016/j.copsyc.2019.06.034.

Devine, D. J. (2012). *Jury Decision Making*. New York University Press.

Devine, D. J., & Caughlin, D. E. (2014). Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. *Psychology, Public Policy, and Law*, *20*(2), 109-134. https://doi.org/10.1037/law0000006.

Dolan, J. G., Bordley, D. R., & Mushlin, A. I. (1986). An eualuation of clinicians' subjective prior probability estimates. *Medical Decision Making*, *6*(4), 216-223. https://doi.org/10.1177/0272989X8600600406.

Domagalski, K., Gongola, J., Lyon, T. D., Clark, S. E., & Quas, J. A. (2020). Detecting children's true and false denials of wrongdoing: Effects of question type and base rate knowledge. *Behavioral Sciences & the Law, 38*(6), 612-629. https://doi.org/10.1002/bsl.2487.

Donovan, T. & Mickey, R. M. (2019). *Bayesian Statistics for Beginners.* Oxford University Press. https://doi.org/10.1093/oso/9780198841296.001.0001.

Elliott, D. M., Mok, D. S., & Briere, J. (2004). Adult sexual assault: Prevalence, symptomatology, and sex differences in the general population. *Journal of Traumatic Stress: Official Publication of the International Society for Traumatic Stress Studies*, *17*(3), 203-211. https://doi.org/10.1023/B:JOTS.0000029263.11104.23.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology, 79*(6), 861–875. https://doi.org/10.1037/0022-3514.79.6.861.

Fisher, R. P., Vrij, A., & Leins, D. A. (2013). Does testimonial inconsistency indicate memory inaccuracy and deception? Beliefs, empirical research, and theory. In *Applied*

issues in investigative interviewing, eyewitness memory, and credibility assessment (pp. 173-189). Springer, New York, NY.

Gallagher, C. (2017, July 17). Women under-represented on juries in serious criminal trials. *Irish Times.* https://www.irishtimes.com/news/crime-and-law/women-under-represented-on-juries-in-serious-criminal-trials-1.3156886.

Goldstein, B. (2017, July 17). *When you hear hooves, think horses not zebra.* https://datasmart.ash.harvard.edu/news/article/when-you-hear-hooves-think-horse-not-zebra-1079#:~:text=One%20of%20the%20things%20that,than%20a%20rare%2C%20improbable%20one.

Government Author (2021). *The end-to-end rape review report on findings and actions.* HM Government.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review*, *93*, 1-44.

Haegerich, T. M., & Bottoms, B. L. (2000). Empathy and jurors' decisions in patricide trials involving child sexual assault allegations. *Law and human behavior*, *24*(4), 421-448. https://doi.org/10.1023/A:1005592213294.

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643. https://doi.org/10.1037/a0023589.

Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd ed.). Guilford Press.

Hritz, A. C., Royer, C. E., Helm, R. K., Burd, K. A., Ojeda, K., & Ceci, S. J. (2015). Children's suggestibility research: Things to know before interviewing a

child. *Anuario de Psicología Jurídica*, *25*(1), 3-12.

https://doi.org/10.1016/j.apj.2014.09.002.

Jose, R., Fowler, J. H., & Raj, A. (2021). Political differences in American reports of sexual

harassment and assault. *Journal of interpersonal violence*, *36*(15-16), 7695-7721.

https://doi.org/10.1177/0886260519835003.

Kahan, D. M., & Braman, D. (2006). Cultural cognition and public policy. *Yale Law &

Policy Review*, *24*, 149-172.

Kahan, D. M., & Braman, D. (2008). The self-defensive cognition of self-defense. *American

Criminal Law Review*, *45*, 1-66.

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective

cognition. SSRN. https://doi.org/10.2139/ ssrn.2973067.

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw

one": A comparative study of college students and police investigators. Law and

Human Behavior, 29, 211-227. https://doi.org/10.1007/s10979-005-2416-9.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and

methodological challenges. *Behavioral and brain sciences*, *19*(1), 1-17.

https://doi.org/10.1017/S0140525X00041157.

Koehler, J. J. (2002). When do courts think base rate statistics are relevant? Jurimetrics, 373-

402.

Lord Justice Clerk's Review Group (2021), *Improving the Management of Sexual Offence

Cases*. https://www.scotcourts.gov.uk/docs/default-source/default-document-

library/reports-and-data/Improving-the-management-of-Sexual-Offence-

Cases.pdf?sfvrsn=6.

Mai, Y., Zhang, Z., & Yuan, K. (2022). *An Open Source WYSIWYG Web Application for Drawing Path Diagrams of Structural Equation Models.* https://semdiag.psychstat.org.

Majeed, N., & Mahmoon, A. K. (2021). The story model of judicial decision-making and reasoning with evidence. *Global Legal Studies Review, 6(2),* 7-13. http://dx.doi.org/10.31703/glsr.2021(VI-II).02.

McDonald, E. (2022). *In the absence of a jury: Examining judge-alone rape trials.* Canterbury University Press.

Menaker, T. A., & Cramer, R. J. (2012). The victim as witness: Strategies for increasing credibility among rape victim-witnesses in court. *Journal of Forensic Psychology Practice*, *12*(5), 424-438. https://doi.org/10.1080/15228932.2012.713829.

Miller, M. K., Maskaly, J., Peoples, C. D., & Sigillo, A. E. (2014). The relationship between mock jurors' religious characteristics and their verdicts and sentencing decisions. *Psychology of Religion and Spirituality*, 6(3), 188-197. https://doi.org/10.1037/a0036344.

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior, 29*(6), 621-637. https://doi.org/10.1007/s10979-005-8122-9.

Nortje, A., & Tredoux, C. (2019). How good are we at detecting deception? A review of current techniques and theories. *South African Journal of Psychology*, *49*(4), 491-504. https://doi.org/10.1177/0081246318822953.

NSPCC (2021). *Child Sexual Abuse: Statistics Briefing.* *https://learning.nspcc.org.uk/research-resources/statistics-briefings/child-sexual-abuse.*

Ortiz, R. R., & Smith, A. M. (2022). A social identity threat perspective on why partisans may engage in greater victim blaming and sexual assault myth acceptance in the# MeToo era. *Violence Against Women*, *28*(5), 1302-1325. https://doi.org/10.1177/10778012211014554.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163. https://doi.org/10.1016/j.jesp.2017.01.006.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*(2), 242-258. https://doi.org/10.1037/0022-3514.51.2.242.

Pennington, N., & 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. https://doi.org/10.1037/0022-3514.62.2.189.

Pennycook, G., & Thompson, V. A. (2017). Base-rate neglect. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment and Memory* (pp. 44–61). Routledge/Taylor & Francis Group.

Rape Crisis (nd). *Statistics – Sexual Violence*. https://rapecrisis.org.uk/get-informed/about-sexual-violence/statistics-sexual-violence/.

Ruva, C. L., & Hudak, E. M. (2013). Pretrial publicity and juror age affect mock-juror decision making. *Psychology, Crime & Law*, 19(2), 179-202. https://doi.org/10.1080/1068316X.2011.616509.

Salerno, J. M., Najdowski, C. J., Bottoms, B. L., Harrington, E., Kemner, G., & Dave, R. (2015). Excusing murder? Conservative jurors' acceptance of the gay-panic defense. *Psychology, Public Policy, and Law, 21*(1), 24. https://doi.org/10.1037/law0000024.

Shah, P., Harris, A., Bird, G., Catmur, C. & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology, 90,* 71-127. https://doi.org/10.1016/j.cogpsych.2016.05.004.

Smith, B. C., Penrod, S. D., Otto, A. L., & Park, R. C. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior*, *20*(1), 49-82. https://doi.org/10.1007/BF01499132.

Violence Against Women Reauthorization Act (2022) U. S. §3623. *https://www.congress.gov/bill/117th-congress/house-bill/1620/text*.

Walby, S., & Towers, J. S. (2017). Measuring violence to end violence: mainstreaming gender. *Journal of Gender-Based Violence*, *1*(1), 11-31. https://doi.org/10.1332/239868017X14913081639155.

Welsh, M. B., & Navarro, D. J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119*(1), 1-14. https://doi.org/10.1016/j.obhdp.2012.04.001.

Zhang, Z., & Yuan, K.-H. (2018). *Practical Statistical Power Analysis Using Webpower and R*. ISDSA Press. https://webpower.psychstat.org.