Public Choice
https://doi.org/10.1007/s11127-025-01330-0

# Juries in flux

**Francesco Parisi**[1] · **Ram Singh**[2]

Received: 4 November 2024 / Accepted: 23 August 2025


**Abstract**
While the right to a trial by an impartial jury remains a cornerstone of the Anglo-American legal tradition, the modus operandi of a "trial by jury" in the United States has been in constant flux. During the last 125 years, twenty-eight states in the U.S. reduced the size of their juries, while three others allowed non-unanimous verdicts in felony and/or misdemeanor cases. Blackstonian ratios and burdens of proof exhibited similar variations across jurisdiction. In 2020, the U.S. Supreme Court cast a critical eye on non-unanimous juries and reintroduced the requirement of unanimity for all felony convictions. In 2023, jury size also received scrutiny from the U.S. Supreme Court, underscoring the enduring volatility of criminal jury practices in the United States. U.S. states currently have the authority to determine jury size, retrial rates, and the Blackstonian ratio applied within their jurisdictions. In this paper, we examine the critical interdependence of these changes in jury structure and their impact on the expected correctness of verdicts and on the decisiveness and accuracy of the jury process. We conclude by discussing how these state-level decisions are often inconsistent with many states' ability to uphold their stated policy objectives and Blackstonian ratios.

**Keywords** Jury size · Standard of proof · Blackstonian ratios · Unintended consequences

**JEL Classification** K0 · K4

An early draft of this paper was presented as the Keynote Lecture at the 10th International Workshop on the Economic Analysis of Litigation, held at University of Catania (Italy) on June 22–23, 2023. One of the authors of this article served as a Consulting Expert for the Attorney General of Florida in *Pretell v. State*, 339 So.3d 506 (Fla. Dist. Ct. App. 2022), *cert. denied*, 143 S. Ct. 1027 (2023). However, all views and analyses presented in this paper are solely those of the authors and do not reflect the positions or opinions of the Attorney General's Office or any other affiliated entities.

✉ Francesco Parisi
parisi@umn.edu

Ram Singh
ramsingh@econdse.org

[1] University of Bologna, Bologna, Italy

[2] Delhi School of Economics, University of Delhi, New Delhi, India

Published online: 18 November 2025

## 1 Introduction

Trials by jury are a fixture of the American criminal justice system. The American jury is commonly depicted as comprising of twelve jurors who deliberate and reach a decision with unanimity, convicting only when the defendant's guilt is established beyond a reasonable doubt.[1] Despite this enduring iconic image of the American jury, its structure has undergone numerous transformations over the past 125 years. From 1898 to 2020, the U.S. Supreme Court allowed criminal convictions to be imposed by juries that did not vote unanimously.[2] Additionally, many U.S. jurisdictions have reduced jury sizes over the years.[3] Great variations are also observed in the manner in which the "beyond a reasonable doubt" standard is implemented across U.S. jurisdictions.

This paper investigates how key institutional features of American criminal jury trials—the number of jurors, the "beyond a reasonable doubt" standard, and the likelihood of retrial after a hung jury—interact to shape both verdict correctness and the decisiveness and overall accuracy of the justice process. Drawing on Urken and Traflet (1983), we define (1) "decisiveness" as a jury's capacity to secure the requisite consensus for a verdict (i.e., the probability of reaching a decision of conviction or acquittal, which is not equal to one because of the requirement of unanimity); (2) "correctness" as the conditional probability that a rendered verdict is factually accurate (i.e., the sum of the probability of convicting the guilty and acquitting the innocent); and (3) "accuracy" as the unconditional probability that the trial process both produces a verdict and that this verdict is correct (i.e., the ratio of the probability of acquitting over the probability of convicting defined for either the innocent or the guilty). We show that, under a unanimity rule, increasing jury size tends to make verdicts less decisive and, overall, also less correct—larger juries reduce true convictions more than they prevent wrongful acquittals—whereas raising the burden of proof boosts decisiveness for acquittals but undermines it for convictions. Our analysis reveals the interdependence of these criteria and how the optimal jury size can vary depending on changes in the standard of proof. We further examine how the frequency of retrials (i.e., refiling hung-jury mistrials) influences the process's accuracy, showing that if hung juries lead to no retrials, mistrials become de facto acquittals, and larger juries can paradoxically weaken the system's overall ability to convict the guilty and acquit the innocent.

Synthesizing these findings, we then explore how jury design shapes the Blackstonian ratio—the ratio of the probability of acquitting a guilty defendant (Type II error) to the probability of convicting an innocent one (Type I error)—which Blackstone prescribed should be ten to one. Building on our theoretical framework, we examine contemporary U.S. state practice by comparing each jurisdiction's selected jury size with its corresponding Blackstonian ratio. Our results reveal pervasive inconsistencies: states that profess similar standards

[1] In the leading 1898 case *Thompson v. Utah*, the U.S. Supreme Court construed the Sixth Amendment to require that in all criminal cases, a jury must comprise of exactly twelve members. Thompson v. Utah, 170 U.S. 343 (1898).

[2] Only in 2020 did the Supreme Court reintroduce the unanimity requirement for jury deliberations that lead to felony convictions. Ramos v. Louisiana, 140 S. Ct. 1390 (2020).

[3] The U.S. Supreme Court continues to allow juries to comprise of fewer than twelve members, although recent challenges before the U.S. Supreme Court signal that further judicial activity on this matter can be expected. Pretell v. State, 339 So.3d 506 (Fla. Dist. Ct. App. 2022) (rejecting the argument that the appellant was entitled to a twelve-member jury when tried for capital sexual battery), *cert. denied*, 143 S. Ct. 1027 (2023).

of proof nonetheless employ very different jury configurations, and those with the same jury size often profess wildly divergent Blackstonian ratios. These misalignments suggest that absent explicit coordination between jury size, voting rules, and burden-of-proof instructions, many states are unable to realize their own stated policy objectives of balancing wrongful convictions against wrongful acquittals.

This paper is organized as follows. Section 2 provides a historical overview of the many changes to “trials by jury” carried out in the United States. We compare the Blackstonian ratios used by U.S. jurisdictions with their respective jury structures. Section 3 presents a theoretical setup of jury design. We develop a formal model to examine the interdependent effects of jury size and standards of proof on correctness of verdicts and the accuracy and decisiveness of the jury process. Section 4 extends the model to evaluate how the optimal balance between jury size and burden of proof varies under different conditions of the criminal justice system, focusing on the relevance of retrials. In Sect. 5, we return to evaluate the Blackstonian ratios adopted by U.S. state jurisdictions, showing that states’ decisions on jury size, standards of proof, and retrial rates often contradict their ability to meet stated Blackstonian ratio goals.

## 2 U.S. juries in flux

Historically, various changes have influenced how American juries function, reflecting efforts to establish a constitutional framework for regulating jury processes in the United States while allowing states autonomy in shaping their criminal justice systems. Specifically, three key features of jury trials have undergone significant evolution and scrutiny over the past 125 years: (i) jury size, (ii) voting rules, and (iii) standards of proof.

The first notable transformation concerned the size of a jury. Whereas *Thompson* interpreted the Sixth Amendment to require juries of twelve persons in criminal cases, twenty-eight states have since departed from this interpretation and have implemented reductions in jury size. States reduced jury size for a variety of reasons. Some states reduced the size of juries during World War II due to a shortage of available jurors. Other states reduced jury sizes in later years due to different concerns, including difficulties in assembling juries in rural areas with smaller populations, as well as a desire to streamline the costly process of jury selection. Several constitutional challenges followed the wave of state-level jury-size reductions. These challenges claimed that a reduction in jury size violated the right to a jury trial guaranteed by the Sixth Amendment of the U.S. Constitution, as explicitly affirmed by the U.S. Supreme Court in *Thompson v. Utah*. During the 1970s, the Supreme Court responded to these challenges by reversing the *Thompson* decision and upholding the use of smaller juries in certain circumstances. For example, in *Williams v. Florida* (1970), the Supreme Court ruled that a verdict rendered unanimously by fewer than twelve jurors in non-capital criminal cases was not inconsistent with the U.S. Constitution.[4] Several subsequent Supreme Court decisions affirmed the constitutionality of juries with fewer than twelve members, imposing some limits.[5] Currently, approximately half of U.S. jurisdictions

[4] Williams v. Florida, 399 U.S. 78 (1970). The Court reasoned that smaller juries could still provide fair and impartial trials, and that the use of six-person juries could help reduce burdens on the court system.

[5] In Ballew v. Georgia, 435 U.S. 223 (1978), the U.S. Supreme Court struck down a Georgia law that allowed juries of only five members in non-capital cases, instead setting a lower-bound limit of six members. Any

allow juries with fewer than twelve jurors in certain types of criminal cases, but this widespread practice continues to face legal challenges, as exemplified by the 2022 case *Khorrami v. Arizona*[6] and the 2023 case *Pretell v. Florida*.[7] The U.S. Supreme Court declined to grant certiorari to these cases, though not without controversy. In a dissenting opinion from the denial in *Khorrami*, Justice Gorsuch argued that the 1970 *Williams* decision, which permitted juries with fewer than twelve members, disregarded centuries of precedent and misinterpreted the Sixth Amendment. Justice Kavanaugh also favored granting certiorari in *Khorrami*. Justice Gorsuch asserted that the right to a jury trial has traditionally entailed a trial before twelve community members and criticized *Williams* for relying on flawed social science rather than sound legal reasoning: "Williams was wrong the day it was decided, it remains wrong today, and it impairs both the integrity of the American criminal justice system and the liberties of those who come before our Nation's courts." When the Supreme Court again declined to hear cases challenging smaller juries in the *Pretell* case, Gorsuch reiterated his position, writing: "If there are not yet four votes on this court to take up the question, I can only hope someday there will be."[8]

A second, less widespread transformation that took place during the twentieth century affected the requirement for unanimity in jury deliberations. Although not addressed in the holding of *Thompson v. Utah*, the requirement that juries must reach verdicts unanimously for criminal convictions was regarded as an important characteristic of a jury trial. Scholars have argued that *Thompson* did not tackle the unanimity requirement to avoid unnecessary redundancies. Unanimity was regarded as a fundamental principle of the criminal justice system because allowing for non-unanimous verdicts could undermine public trust in the justice system and increase the risk of wrongful convictions (Coughlan 2000).[9] However, the *Thompson* decision's silence on the unanimity requirement left the door open for some jurisdictions to introduce reforms that allow convictions to occur through non-unanimous verdicts. The shift away from requiring unanimous jury decisions was motivated by the need to avoid situations where a single juror could prevent a conviction in spite of compelling evidence of a defendant's guilt. Thus, this change aimed to reduce instances of jury nullification.[10] During the last century, the use of non-unanimous juries has faced numer-

jury with fewer than six members would be unconstitutional because such a jury would be too small to be representative of the relevant community and would not provide enough diversity of opinion to ensure a fair trial. In 1979, Burch v. Louisiana, 441 U.S. 130 (1979), held that states could either reduce jury size or relax the requirement for unanimity, but a given jury could not do both at the same time.

[6] Khorrami v. Arizona, 143 S. Ct. 22 (2022), *cert. denied*, (Gorsuch, J., dissenting).

[7] Pretell v. State, 339 So.3d 506 (Fla. Dist. Ct. App. 2022), *cert. denied*, 143 S. Ct. 1027 (2023).

[8] Considering the upheaval following the declaration of non-unanimous juries as unconstitutional—and the subsequent push for retroactive unanimity—the denial of certiorari in *Khorrami* and *Pretell* is unsurprising. Despite this deferral, the Supreme Court will likely need to address the issue in a future term.

[9] See Rule 31 of the Federal Rules of Criminal Procedure. See also the interesting discussion of non-unanimous convictions by Glasser (1996). For a critical discussion of the requirement for unanimity in jury decision-making, see Bornstein and Greene (2017). The authors draw on recent research using jury simulations and analyses of court records to examine how juries reach decisions.

[10] In 1898 (the same year as the *Thomson* decision), Louisiana introduced legislation that departed from the established unanimity rule. See Ramos v. Louisiana, 140 S. Ct. 1390, 1394 (2020). Legal historians suggest that Louisiana's reform was prompted by the growing ethnic diversity in jury composition and the concern that criminal verdicts could be influenced by ethnic divisions among jurors—where minority jurors might veto proposed verdicts, potentially causing mistrials. For instance, Hannaford-Agor et al. (2002) conducted a multi-phased study through the NCSC, highlighting concerns about the high frequency of mistrials in certain jurisdictions. Similarly, upon its admission to the Union in 1907, Oklahoma relaxed its unanimity require-

ous legal and political challenges at both state and federal levels.[11] Ultimately, in 2020, the U.S. Supreme Court took on this critical issue, ruling that criminal convictions reached by non-unanimous verdicts are unconstitutional. In *Ramos v. Louisiana*, the Court overturned its previous decision in *Apodaca*, holding that non-unanimous verdicts for criminal felony convictions violate the Sixth Amendment's guarantee of a jury trial.[12] Whereas unanimous jury verdicts became necessary as of 2020 for felony convictions because of the *Ramos* decision, carveouts for allowing non-unanimity soon resurfaced in other corners of U.S. law. Most notably, in 2023 Florida modified the capital sentencing statutes to allow capital punishment when eight out of twelve jurors determine that a defendant should be sentenced to death.[13] Whether the Florida statutory amendment will face a challenge (either in state court or the U.S. Supreme Court) remains to be seen.[14]

A third area of flux in the U.S. criminal justice system relates to the burden of proof that must be met in criminal adjudication. Despite much criticism, the "beyond a reasonable doubt" (hereinafter, BARD) standard continues to stand as a cornerstone of the American legal system.[15] In 1970, the U.S. Supreme Court added its seal, raising the BARD standard to the stature of a constitutional principle under the Fifth Amendment to the U.S. Constitution.[16] However, the Court made no attempt to define what the practical application of the standard entails, or to determine how strong the weight of evidence should be in probabilistic terms. The Supreme Court's reluctance to set a consistent and tractable probabilistic value or to establish some structural guidance for States in implementing the BARD standard leaves room for heterogeneous interpretations of this standard, for courts and juries alike. In providing jury instructions, courts often refer to the Blackstonian ratios that have been adopted in their respective jurisdictions. The choice between different numerical val-

ment to permit non-unanimous convictions in misdemeanour cases. In 1934, Oregon became the third state to allow non-unanimous jury convictions, following a constitutional amendment.

[11] Duncan v. Louisiana, 391 U.S. 145 (1968); Apodaca v. Oregon, 406 U.S. 404 (1972); Johnson v. Louisiana, 406 U.S. 356 (1972).

[12] Ramos v. Louisiana, 140 S. Ct. 1390 (2020). In 2021, the U.S. Supreme Court returned to this issue to settle an important question left open by *Ramos*—whether Louisiana and Oregon were required to reconsider prior non-unanimous convictions. In Edwards v. Vannoy, 141 S. Ct. 1547 (2021), the Supreme Court resolved this question in the negative—in a split decision, the Court held that the *Ramos* decision does not automatically apply retroactively to previously decided cases: "States remain free, if they choose, to retroactively apply the jury-unanimity rules as a matter of state law in state post-conviction proceedings." In this way, the Supreme Court left the decision of how to handle prior non-unanimous convictions up to the states. The Oregon Supreme Court opted to apply the jury-unanimity rule retroactively, effectively overturning hundreds of non-unanimous jury verdicts. Watkins v. Ackley, 523 P.3d 86 (2022); Sparling (2022). Conversely, the Supreme Court of Louisiana held that the jury-unanimity ruling in *Ramos* did not apply retroactively in Louisiana. State v. Riddick, 351 So. 3d 273 (La. 2022).

[13] 2023 Fla. Sess. Law Serv. 450 (West). The statutory amendment appears to align with the Florida Supreme Court's holding in *State v. Poole*, 297 So. 3d 487 (Fla. 2020).

[14] *See* News Serv. of Fla., *Florida Death Penalty Changes Causing 'Chaos,' Attorneys Say*, Tampa Bay Times (Aug. 23, 2023), https://www.tampabay.com/news/florida/2023/08/23/florida-death-penalty-changes-causing-chaos-attorneys-say.

[15] The BARD standard is a legal standard that applies to criminal cases, and it is used to assess whether the prosecution has met its burden of proving the defendant's guilt. Its origins trace back to the English common law tradition, which held that a criminal defendant could only be found guilty if the evidence brought by the prosecutor established guilt "beyond a reasonable doubt." In the United States, all jurisdictions have uniformly recognized this aspirational standard since the late 19th and early twentieth centuries.

[16] *In re* Winship, 397 U.S. 358, 364 (1970). The Court held that this standard of proof should be considered as embedded in the "due process" clause.

ues of the Blackstonian ratio necessarily implies that different jurisdictions are committed to different standards of proof (Pi et al. 2020). Courts' emphasis on the desire to provide a high level of accuracy in criminal adjudication captures the competing policy objectives of protecting innocent defendants and holding guilty individuals accountable for their actions. The Blackstonian ratio balances these two competing objectives when determining the choice between different BARD standards.

As recently shown in the empirical study carried out by Pi et al. (2020), behind the nominal uniformity of the BARD standard, this evidentiary pillar is at risk of being implemented inconsistently across U.S. jurisdictions. Courts invoke the concept of the Blackstonian ratio to convey the general idea that if there is any doubt in the case, it is better to err toward leniency than toward punishment. Volokh (1997) and Pi et al. (2020) researched the official proclamations of Blackstonian ratios across U.S. jurisdictions, unveiling substantial differences between states. Judge Blackstone's dictum is "it is better that ten guilty persons escape, than that one innocent suffers" (Blackstone 1769). Blackstone's ten-to-one ratio is just one—albeit the most common—of a set of possible measures that can guide jury considerations on the right standard of proof. Other scholars have echoed Blackstone's principle while embracing different numerical values for the ratio. Most notably, in a letter sent to Benjamin Vaughan on March 14 1785, Benjamin Franklin endorsed a much higher standard for the safeguard of defendants: "It is better 100 guilty persons should escape than that one innocent person should suffer" (Franklin 1785 [1785], p. 293).[17]

We suggest that decades of jury structure reforms have introduced inconsistencies in criminal justice systems across states. Table 1 presents current data on jury sizes used in each state for felony (F) and misdemeanor (M) cases,[18] matched with the (updated) data collected by Pi et al. (2020), reporting Blackstonian ratios, listed by U.S. jurisdictions.[19]

Blackstonian ratios reflect the level of protection granted to defendants by defining the acceptable trade-off between wrongful convictions and wrongful acquittals. A jurisdiction with a ratio of 1:10, for instance, would require that the likelihood of convicting an innocent person be ten times lower than the likelihood of acquitting a guilty one. A jurisdiction with a ratio of 1:20 would place an even greater emphasis on avoiding wrongful convictions, tolerating a higher probability of acquitting guilty individuals.

In the upcoming sections, we introduce a theoretical model that examines how changes in jury structure—specifically jury size, BARD standards, and retrial rates—impact the frequency of wrongful convictions and acquittals. These factors are closely linked to the Blackstonian ratios adopted in different jurisdictions. Our model's findings provide fresh insights into the data presented in Table 1, revealing significant inconsistencies within the criminal justice systems of several U.S. states.

[17] For a literature review on Blackstonian ratio problems, see also Rizzolli and Saraceno (2013).

[18] Table 1 summarizes the jury sizes used across U.S. jurisdictions. See, Bureau of Justice Statistics (2004). While most states set jury size based on offense severity—distinguishing between felonies and misdemeanors—some use alternative criteria. For instance, Indiana differentiates by court level, and Louisiana by minimum sentence severity. For clarity and ease of comparison, we have standardized the coding, distinguishing jury sizes for the broad categories of felony (F) and misdemeanor (M) cases. Appendix 2 contains the list of state regulations that reduced jury sizes for both felonies and misdemeanors. For a list of states that reduced jury sizes only for misdemeanors or for certain categories of lesser offenses, and the specific criteria utilized for each jurisdiction, see Bureau of Justice Statistics (2004) and Luppi and Parisi (2013).

[19] The Blackstonian ratios are based on the most recent values adopted by state Supreme Courts. Cases and citations can be found in Pi, et al. (2020). For a previous important research surveying the official proclamations of Blackstonian ratios across U.S. jurisdictions, see also Volokh (1997).

**Table 1** Jury Size and Blackstonian Ratios in U.S. Courts

| Black-stonian ratio | Jury size | List of states |
|---|---|---|
| 1:1 | 12 | Alaska (F); Colorado (F); Georgia (F); Hawaii (F); Illinois (M and F); Iowa (F); Kentucky (F); Louisiana (F); Massachusetts (F); Minnesota (F); Montana (F); Nebraska (F); Oregon (F); Pennsylvania (F); Tennessee (M and F); Texas (F); Virginia (F); U.S. (M and F); Washington (M and F, or less at defendant's choice) |
| | 8 | Arizona (M and F) |
| | 7 | Virginia (M) |
| | 6 | Alaska (M); Colorado (M); Connecticut (M and F); Georgia (M); Hawaii (M); Illinois (M and F); Indiana (M and F); Iowa (M); Kentucky (M); Louisiana (M); Massachusetts (M); Minnesota (M); Montana (M); Nebraska (M); New Jersey (M and F, by agreement); Oregon (M); Pennsylvania (M, by agreement); Tennessee (M and F); Texas (M); Virginia (M) |
| 1:5 | 12 | Alabama (M and F); Arkansas (F); New York (F) |
| | 6 | Arkansas (F, by agreement); New York (M) |
| 1: Some | 12 | Mississippi (F); Missouri (F) |
| | 6 | Mississippi (M); Missouri (M, by agreement) |
| 1:10 | 12 | California (F); Maine (M and F); Michigan (F); New Hampshire (F); North Carolina (M and F); Oklahoma (F); Rodhe Island (M and F) |
| | 8 | Utah (M and F) |
| | 6 | California (M, by agreement); Florida (M and F); Michigan (M); New Hampshire (M); Oklahoma (M); Wisconsin (M and F, by agreement) |
| 1: Many | 12 | Idaho (F); South Carolina (F); South Dakota (M and F) |
| | 6 | Idaho (M); South Carolina (M) |
| 1:99 | 12 | New Mexico (F); Ohio (F); West Virginia (F) |
| | 8 | Ohio (M) |
| | 6 | New Mexico (M); West Virginia (M) |
| 1:100 | 12 | Wyoming (F) |
| | 6 | Wyoming (M) |
| None | 12 | Delaware (M and F); Maryland (M and F); Nevada (F); North Dakota (F); Vermont (M and F) |
| | 6 | Nevada (M); Kansas (M and F, by agreement); North Dakota (M) |

# 3 A theoretical model

## 3.1 Related literature

Most of the prior law and economics literature examined specific dimensions of the jury process in isolation, focusing on how large a jury should be. For example, Ladha (1995) showed that Condorcet's jury theorem still holds under certain conditions, even when votes are correlated. Paroush (1997) offered insights into the limitations of the original Condorcet jury theorem, challenging the conclusion that larger juries necessarily yield more accurate decisions. Ben-Yashar and Paroush (2000) explored the robustness of the broader literature on jury theorems, showing that when jurors exhibit heterogeneous levels of competence, jury size becomes a critical optimization factor. Dharmapala and McAdams (2003) extended the idea of expressive law (i.e., how laws can generate compliance through their expressive function, independent of deterrence) to jury decisions, suggesting that larger juries could enhance the expressive impact on society's perception of justice. Helland and Raviv (2008) challenged the consensus on optimal jury size by showing that the probability of Type I and Type II errors (false convictions and false acquittals) is not sensitive to jury size when jury deliberation follows a random walk model.

Other scholars concentrated their analyses on how juries should vote to convict. For example, Klevorick and Rothschild (1979) focused on how jury size and voting rules influence verdict outcomes and deliberation time. Klevorick et al. (1984) built on Klevorick and Rothschild (1979) to examine how jurors integrate and process information presented during a trial. Yet other scholars focused on how to optimally set the standard of proof. For example, Demougin and Fluet (2005) carried out a comparative analysis exploring how different standards of proof balance deterrence and error minimization, providing a theoretical framework for understanding the implications of the "beyond a reasonable doubt" standard. Ognedal (2005) explored how changes in the standard of proof affect deterrence across different types of crimes, suggesting that lowering the standard of proof may reduce deterrence for more serious offences due to limited prison capacity. Finally, Lando (2009) challenged the traditional view that the determination of the beyond a reasonable doubt standards should be based solely on the balancing of wrongful conviction and acquittal costs, instead examining the impact of the proof standard on deterrence of crime.

The optimal size for a jury and the associated impact on verdict accuracy have been specific objects of theoretical and experimental studies. After the Supreme Court established the constitutionality of jury sizes smaller than the traditional twelve-member configuration, empirical research examined the practical effects of changes in jury size on verdict accuracy and deliberation dynamics. A notable study by Kerr and MacCoun (1979) evaluated the ability of mock juries to process and retain critical evidence from the trial. The recall of facts was used as a proxy for the verdict accuracy. The study showed that smaller juries—composed of at least six members—can adequately process and retain the information needed for accurate decision-making. Informational recall accuracy was consistent across jury sizes. Once a minimum threshold of around six members was reached, the addition of more jurors did not substantially enhance the jury's ability to retain information. Spencer (2007) reached similar conclusions using judge-jury agreement rates as proxies for verdict correctness. The results suggested that smaller juries with at least six members can achieve comparable accuracy to larger juries, raising questions about the necessity of the traditional twelve-member jury size.[20]

More recent contributions have produced results that diverge from the earlier literature. Suzuki (2012) presented evidence that smaller juries are more likely to convict when a defendant's guilt is less than overwhelmingly certain. His findings reveal that when the evidence strongly points to guilt, all jury sizes are highly likely to convict. However, for

[20] Although Spencer (2007) did not directly measure jury correctness, the study estimated that juries reached accurate verdicts in approximately 87% of cases, with an estimated probability of incorrect conviction in the presence of actual innocence at 0.25 (standard error 0.07).

cases where guilt is less certain, notable differences emerge.[21] In direct contrast to earlier studies, Watanabe (2020) demonstrated the clear superiority of the twelve-member jury over smaller configurations. The author focused on two critical factors driving consensus within the jury—opinion homogeneity and anti-conformity tendencies. Based on the author's calibration of these forces, the study concluded that juries of approximately twelve members strike an optimal balance.[22] Watanabe's numerical simulation assumes a majority voting rule, overlooking the fact that historically, most juries could only convict with a unanimous vote. This omission misses a critical aspect of the jury process. With a majority rule and tie-breaking mechanism, juries are guaranteed to reach a verdict, thereby eliminating the chance of a hung-jury. In contrast, when unanimity is required, the jury may be unable to reach a verdict, leading to a potential mistrial. Our study utilizes a more structured economic model which focuses on unanimous decision rules and includes decisiveness as a critical ingredient of accuracy.

Other scholars have explored the optimal combination of voting requirements and jury size (e.g., Urken and Traflet 1983; Neilson and Winter 2005; Luppi and Parisi 2013; Guerra et al. 2020).[23] Studies covering similar ground to our paper include Urken and Traflet (1983), Neilson and Winter (2000), and Luppi and Parisi (2013). Urken and Traflet identified two jury performance evaluation criteria that are considered in our paper. The first criterion pertains to a jury's ability to make a correct decision. The second criterion relates to a jury's ability to reach the level of consensus necessary for a verdict. We refer to these two criteria as "correctness" and "decisiveness," respectively. These two criteria should be considered in conjunction with one another when evaluating jury design. From this perspective, non-unanimous jury verdicts can be viewed as means for increasing the decisiveness of trials by reducing mistrial rates. In analyzing voting requirements, Urken and Traflet adhered to the assumptions that a twelve-member jury made decisions according to the conventional standard of proof. Luppi and Parisi (2013) explored the effect of changes in jury size on decisiveness (i.e., on the probability of reaching a verdict) under a unanimity rule.

Whereas much of the literature on jury design has investigated the effect of jury size and voting requirements on the decisiveness of juries, most articles neglect the critical interdependence of these factors, and how that interdependence maximizes the correctness of jury verdicts and the accuracy of trial outcomes. In this paper, along with jury size and voting requirements, we analyze the role of standards of proof in jury design. The BARD standard requires the factfinder to attain a high degree of confidence in the defendant's guilt to convict, but neither the standard's quantitative interpretation nor its theoretical justifications are clear or commonly agreed upon. Unlike the preponderance standard, which is understood to

[21] According to Suzuki's (2012) results, in scenarios where there is an estimated 80 percent probability of guilt, a twelve-member jury would unanimously vote to convict less than 10 percent of the time. In contrast, a six-member jury would be expected to unanimously convict over 25 percent of the time. Furthermore, if juries were allowed to convict on a non-unanimous basis, with a nine out of twelve votes, the jury would reach a guilty verdict in approximately 60 percent of similar cases.

[22] The optimality value in Watanabe (2020) is identified by balancing the trade-offs between accuracy (correct decision-making) and efficiency (minimization of time) of deliberations. The study posits that the commonly used twelve-member jury size may reflect an evolved balance among these two arguments of the objective function. On this point, Suzuki (2012) provides a more sceptical view of the origins of the twelve-member jury configuration, dating its origins back to 725 A.D., when Welsh King Morgan instituted jury trials and chose the number twelve, symbolically connecting the judge and jury to Jesus and his twelve Apostles.

[23] See also King and Nesbit (2009) for an empirical study on jury size and voting requirements in civil jury trials.

refer to a precise, greater than 50% measure (e.g., Clermont and Sherwin 2002; Demougin and Fluet 2006), the BARD standard has hitherto resisted reduction to a numerical value (e.g., Kaplow 2012). The rationales for courts' reluctance to quantify this standard are theoretically weak.[24] That said, if a quantitative measure of the standard of proof is deemed desirable for guiding jury decisions, determining an optimal measure presents a challenge from the theoretical point of view. Kaplow (2011) addressed this issue through a normative analysis on how the burden of proof should be optimally set in the presence of a trade-off between deterrence benefits and chilling costs.

In the following section, we develop a formal model to consider the interdependent effects of jury size and standards of proof on the (i) correctness of verdicts and (ii) decisiveness of the jury process to evaluate the trade-offs that emerge under different conditions of the criminal justice system. Our analysis shows how variations in the standard of proof influence the optimal jury size.

### 3.2 Basic setup

In this section we present a basic theoretical setup to discuss how variations in jury size affect the optimal choice of the standard of proof with respect to different expected trial outcomes. Our setup relies upon the criminal trial models in Neilson and Winter (2005) and Guerra et al. (2020), examining how the optimal choice of jury size varies with the standard of proof for criminal convictions when juries are required to deliberate unanimously.

The setup of our model has the following features: (i) evidence is presented to $N$ jurors; (ii) jurors are heterogeneous in the interpretation and assessment of the evidence presented to them; (iii) jurors compare their assessment of the evidence to the BARD standard, $e_I$; and (iv) a verdict is reached by the jury according to a unanimity rule.

We consider a criminal trial where nature chooses both whether an individual committed a crime, as well as the amount of incriminating evidence that is found against that individual. In our setup, the prosecution brings charges against individuals who have allegedly committed crimes (hereinafter, we shall refer to individuals charged with a crime as "defendants"). Defendants may have either committed the crime (guilty, or G-defendants) or not committed the crime (innocent, or I-defendants). Denote the number of guilty defendants as $G$ and the number of innocent defendants as $I$.

During trial, the prosecution and defense present evidence supporting their respective positions. Let, $e$ denote the balance of incriminating evidence presented through the adversarial discovery from both sides. A higher value of $e$ entails a greater level of incriminating evidence, which correlates with a higher probability of guilt. Depending on the context, in some cases the evidence incriminating a G-defendant can take on low values of $e$ (e.g., the prosecution did not find much evidence to prove a G-defendant's guilt). On the other hand, $e$ might take high values for I-defendants. Therefore, it is not possible to infer the defendant's guilt with certainty based on any given value of $e$.

Formally speaking, ex ante (before the trial), $e$ is a continuous random variable. Priors about $e$ depend on whether the defendant has committed the crime or not. The evidence

[24] Caselaw articulates rationales for the lack of a quantitative specification of the standard of proof. *See, e.g.*, McCullough v. State, 657 P.2d 1157, 1159 (Nev. 1983), stating: "The concept of reasonable doubt is inherently qualitative. Any attempt to quantify it may impermissibly lower the prosecution's burden of proof, and is likely to confuse rather than clarify."

| | |
|---|---|
| t = 1 | • BARD standard of proof $e_I$ is determined.<br>• The defendant's type is determined: G-defendant or I-defendant. |
| t = 2 | • Prosecution and defense present evidence, $e$.<br>• Jurors interpret $e$ (signals $s_i$ are generated). |
| t = 3 | • Jurors deliberate and render a verdict. |

**Fig. 1** Timeline

is more likely to be incriminating if the defendant committed the crime, as compared to instances in which he did not commit the crime. Formally, for G-defendants, $e$ follows a conditional density function $f\,(e|G)$; otherwise, $e$ follows a conditional distribution with associated density function $f\,(e|I)$. Let $F\,(e|G)$ and F $(e|I)$ be the distribution functions.

The assumption that evidence is more likely to be incriminating when the defendant committed the crime—i.e., a G-defendant generates stronger evidence (in expected terms) than an I-defendant—means that $F\,(e|G)$ first order stochastically dominates F $(e|I)$, which is a standard assumption in criminal models (e.g., Rubinfeld and Sappington 1987; Miceli 1990; 2009, p.125; Feess and Wohlschlegel 2009; Rizzolli and Saraceno 2013; Guerra et al. 2020).

Jurors are untrained ordinary people, with heterogeneous skills for evaluating the facts of the case. Even though the same evidence is presented to all jurors at trial, individual jurors may interpret evidence differently and may at times give too much or too little weight to the evidence presented. In other words, jurors read $e$ with error (Fig. 1).

Formally speaking, like Neilson and Winter (2005) and Guerra et al. (2020), we assume that jurors do not directly observe the true strength of the evidence presented, but rather they observe signals of varying strength related to the evidence.[25] This idea can be formalized: after looking at $e$, juror $i$ receives a signal $s_i$ indicating the strength of the case against the defendant. The higher the value of $s_i$, the more persuasive is the case against the defendant. Let,

$$s_i = e + \varepsilon_i,$$

where $\varepsilon_i$ is the error term for juror for $i$. For any given $e$, each juror receives a different signal depending the realization of the error term. A juror receiving a strong signal is more likely to believe that the defendant is guilty than a juror receiving a weak signal. For sim-

[25] The difference in strength can result from a variety of reasons. Jurors might observe $s$ with error, as in Neilson and Winter (2000); jurors might disagree about the BARD standard, as in Neilson and Winter (2005); or jurors might have different perceptions about the strength of the evidence, as in Feddersen and Pesendorfer (1998) and Guarnaschelli et al. (2000). For similar formulations, see Guerra et al. (2020). See also Arce et al. (1996) and Alpern and Chen (2017), considering other forms of juror heterogeneity.

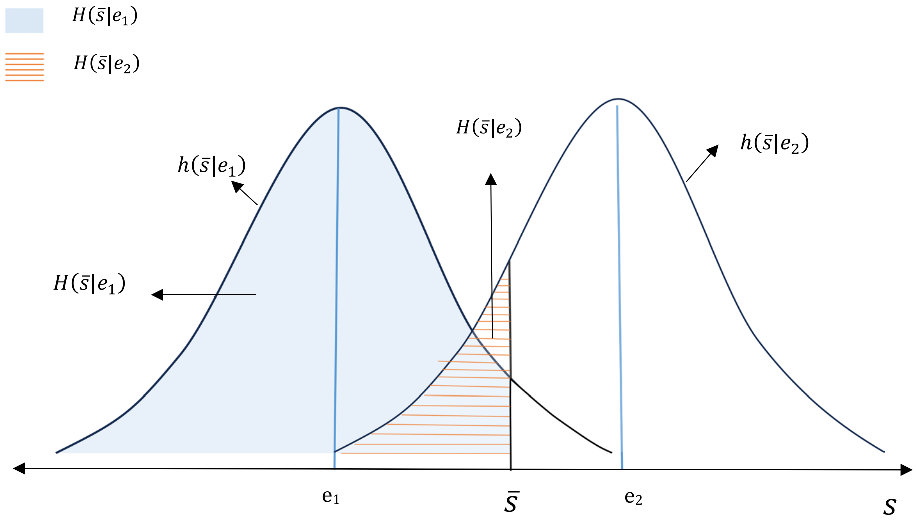


**Fig. 2** Probability signal meets BARD, given evidence levels $e_1$ and $e_2$

plicity, assume that for any given $e$, $\varepsilon_i$ are independently and identically distributed such that $\varepsilon_i \sim N\left(0, \sigma^2\right)$ for all $i \in \{1, \ldots, N\}$.[26]

In view of the above, the distribution of $s_i$ is essentially derived from the underlying distribution of $\varepsilon_i$. Also, $s_i\left(e\right) \sim N\left(e, \sigma^2\right)$ for all $i \in \{1, \ldots, N\}$. Let $h_i\left(.\right)$ and $H_i\left(.\right)$ denote the density and distribution functions, respectively, for $s_i$. In view of the assumption that $\varepsilon_i$ are independently and identically distributed, $h_i\left(.\right)$ and $H_i\left(.\right)$ are the same for all $s_i$, $i \in \{1, \ldots, N\}$. So, we let $h\left(.\right)$ and $H\left(.\right)$ denote the density and distribution function for $s_i$ for all $i$. Clearly, $h\left(.\right)$ and $H\left(.\right)$ are functions of $e$: $h\left(.|e\right)$ and $H\left(.|e\right)$.

Put differently, in our set up, jurors are ex-ante identical and neutral (i.e., they do not suffer from bias in favor or against the defendant). A priori, each juror is equally likely to receive a strong or weak signal, and their reading of the evidence is on average correct. For each juror, the expected strength of the signal is positively correlated to the strength of the case, $e$. However, jurors are ex-post heterogeneous in that they receive signals of different strength—some receive strong signals whereas other receive weak signals. Therefore, their assessments of the case are heterogeneous and independent from one another.

Let $e_I$ be the legal standard of proof that the prosecution must meet to prove the defendant's guilt. Hereinafter, we shall refer to this standard as the BARD standard. For any given $e_I$, higher levels of incriminating evidence, $e$, increase the likelihood that the signal $s_i$ will satisfy the BARD standard, $e_I$. Formally, $H\left(.|e\right)$ is a decreasing function of $e$. To illustrate, we use Fig. 2. Let $e_1$ and $e_2$ denote the two possible levels of incriminating evidence presented during trial, where $e_2 > e_1$. The figure shows the probability distributions of signals generated by evidence levels $e_1$ and $e_2$. Note that $h\left(\bar{s}|e_1\right)$ and $h\left(\bar{s}|e_2\right)$ are the density plots corresponding to evidence levels $e_1$ and $e_2$, respectively. Therefore, $H\left(\bar{s}|e_1\right)$ is the probability that the signal value is less than or equal to a chosen level, say $\bar{s}$, for evidence level $e_1$. $H\left(\bar{s}|e_1\right)$ is measured as the area under the function $h\left(\bar{s}|e_1\right)$ to the left of the

[26] Assuming a normal distribution for $\varepsilon^i$ simplifies the formal analysis. Our results are qualitatively robust to other distributions.

BARD standard, $\overline{s}$. Similarly, $H(\overline{s}|e_2)$ is the area under the function $h(\overline{s}|e_2)$ to the left of $\overline{s}$. Clearly, $H(\overline{s}|e_1) > H(\overline{s}|e_2)$. In other words, when the incriminating evidence is weak, a juror is more likely to find the defendant not guilty. Conversely, stronger incriminating evidence strengthens the signal received by the juror, making it more likely for the defendant to be found guilty.

As in Neilson and Winter (2005) and Guerra et al. (2020), we assume that jurors do not vote strategically and that they make decisions independently from one another, deliberating on the defendant's guilt based on the signal they received.[27] So, a juror $i$ finds the defendant guilty if the strength of the incriminating evidence meets the required BARD standard, $s_i > e_I$. If $s_i \leq e_I$, the juror has doubts about the guilt of the defendant and declares the defendant 'not guilty', as instructed by the judge. This means that for any given $e$ presented at trial at $t = 2$, each juror will find the defendant not guilty with probability $H(e_I|e)$, and guilty with probability $1 - H(e_I|e)$. For a given $e$, $H(e_I|e)$ increases in $e_I$.

In our setup, the jury must decide unanimously. So, the defendant will be found guilty if $s_i > e_I$ for all $i \in \{1, \ldots, N\}$. If $s_i \leq e_I$ for all $i \in \{1, \ldots, N\}$, the verdict will be not-guilty. So, for a given $e$ presented to the jury, the probability of a guilty verdict is $[1 - H(e_I|e)]^N$. The probability of a not-guilty verdict is $[H(e_I|e)]^N$. The probability of mistrial is $1 - [H(e_I|e)]^N - [1 - H(e_I|e)]^N$.

Taking the jury size ($N$) as given, for any given $e$ presented to the jury, the probability of a guilty verdict is decreasing in the standard $e_I$, whereas the probability of a not-guilty verdict is increasing in $e_I$. For any given BARD standard $e_I$ and for any given $e$, the probability of a not-guilty verdict decreases with jury size ($N$); and the probability of a guilty verdict also decreases with jury size. So, decisiveness decreases with jury size (i.e., the probability of a mistrial increases with jury size).[28] However, the effect of $e_I$ on decisiveness is ambiguous.

We can now derive the probabilities of a wrongful conviction, a wrongful acquittal, and a hung jury in a single trial, as estimated at time $t = 1$.[29]

Consider the case where the defendant did not commit the crime (i.e., at $t = 1$, the defendant is an I-defendant). In this case, the evidence will be generated following distribution function $F(e|I)$. This means that a juror will find the defendant not-guilty with probability $\int H(e_I|e)\, dF(e|I) = \int H(e_I|e) f(e|I)\, de$, and guilty with probability $1 - \int H(e_I|e)\, dF(e|I)$. Let,

$$\tilde{H}(e_I|I) = \int H(e_I|e)\, dF(e|I).$$

[27] This assumption—which is the behavior assumed by Condorcet—allows us to isolate the role of our three institutional variables from the possibility of informational cascades (e.g., Luppi and Parisi 2013), and strategic voting (e.g., Ladha 1992; Feddersen and Pesendorfer 1998; Kaniovski and Zaigraev 2011). It should further be noted that our analysis focuses on binary unanimous decisions (guilty/not guilty), under which truthful voting becomes a dominant strategy, so strategic manipulation is not a primary concern. In a different setting with majority decision rules and continuous strategies—such as majority voting on the severity of a sentence—strategic considerations would become significant.

[28] See, however, how this result may change when free-riding and informational cascades in decision-making are introduced (Luppi and Parisi 2013).

[29] For an extensive analysis on appeals and retrials, see Mukhopadhaya (2003). and Neilson and Winter (2005).

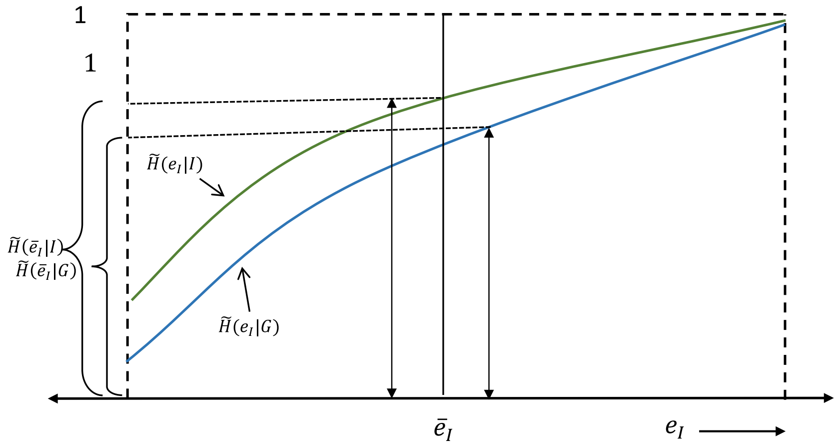


**Fig. 3** Probability that signal meets the BARD, conditional on defendant's Type

That is, a juror will correctly find an I-defendant not-guilty with probability $\tilde{H}(e_I|I)$. With probability $1-\tilde{H}(e_I|I)$, the juror will incorrectly find an I-defendant guilty.

Next consider the case where the defendant committed the crime (i.e., at $t=1$, the defendant is a G-defendant). In this case, the evidence will be generated following distribution function $F(e|G)$. This means that a juror will find the defendant not guilty with probability $\int H(e_I|e)\,dF(e|G) = \int H(e_I|e)\,f(e|G)\,de$, and guilty with probability $1-\int H(e_I|e)\,dF(e|G)$. Let,

$$\tilde{H}(e_I|G) = \int H(e_I|e)\,dF(e|G)\,.$$

That is, a juror will incorrectly find a G-defendant not guilty with probability $\tilde{H}(e_I|G)$; whereas with probability $1-\tilde{H}(e_I|G)$, the juror will correctly find a G-defendant guilty.

It can be seen that $\tilde{H}(e_I|G)$ and $\tilde{H}(e_I|I)$ are increasing in the BARD standard, $e_I$. Moreover, the assumption that $F(e|G)$ first order stochastically dominates $F(e|I)$ and $H(.|e)$ is decreasing in $e$ implies that:

$\tilde{H}(e_I|G) < \tilde{H}(e_I|I)$ and $1-\tilde{H}(e_I|G) > 1-\tilde{H}(e_I|I)$.

In other words, the signals received by the jurors are informative. For any given BARD, $e_I$, an I-defendant is more likely to be found not guilty by a juror than a G-defendant. On the other hand, a G-defendant is more likely to be found guilty, compared to an I-defendant. See Fig. 3.

Before proceeding further, we should note the following properties of the $\tilde{H}(e_I|I)$ and $\tilde{H}(e_I|G)$ plots in Fig. 3: Both $\tilde{H}(e_I|I)$ and $\tilde{H}(e_I|G)$ are increasing in $e_I$, such that $\lim\limits_{e_I\to-\infty}\tilde{H}(e_I|I)=0$, $\lim\limits_{e_I\to-\infty}\tilde{H}(e_I|G)=0$. Further, $\lim\limits_{e_I\to\infty}\tilde{H}(e_I|I)=1$, $\lim\limits_{e_I\to\infty}\tilde{H}(e_I|G)=1$. That is, the two plots are arbitrarily close to each other at both a very low and an extremely high BARD standard.[30]

[30] When the BARD standard is set at zero, all defendants will be found guilty so the I-defendants and G-defendants face similar prospects. Likewise, at extremely high BARD standards, defendants of neither type will be found guilty.

Moving ahead, the likelihoods of a jury reaching a unanimous verdict or a mistrial behave similarly for both I-defendants and G-defendants. At time $t = 1$, the probability of an acquittal (a not-guilty verdict) for an I-defendant is: $P(A|I) = \left[\tilde{H}(e_I|I)\right]^N$. The probability of a conviction (a guilty verdict) for I-defendants (wrongful conviction or Type-1 error) is: $P(C|I) = \left[1 - \tilde{H}(e_I|I)\right]^N$. So, for any given BARD standard, the probability of acquitting an I-defendant, as well as the probability of wrongfully convicting an I-defendant (or Type-1 error), decreases with jury size. This means that for I-defendants, the probability of a unanimous verdict is,

$P[U|I] = \left[\tilde{H}(e_I|I)\right]^N + \left[1 - \tilde{H}(e_I|I)\right]^N$ which decreases with jury size. The complementary probability of a mistrial is,

$1 - P[U|I] = 1 - \left[\tilde{H}(e_I|I)\right]^N - \left[1 - \tilde{H}(e_I|I)\right]^N$, which is increasing in jury size.

Similarly, at $t = 1$, the probability of a not-guilty verdict for a G-defendant (wrongful acquittal or Type-2 error) is: $P(A|G) = \left[\tilde{H}(e_I|G)\right]^N$. The probability of a correct conviction (a guilty verdict) for a G-defendant is: $P(C|G) = \left[1 - \tilde{H}(e_I|G)\right]^N$. That is, for any given BARD standard, the probability of convicting a G-defendant, as well as the number of correct convictions, decreases with the jury size. Consequently, for G-defendants, the probability of a unanimous verdict,

$P[U|G] = \left[\tilde{H}(e_I|G)\right]^N + \left[1 - \tilde{H}(e_I|G)\right]^N$, decreases with jury size. Therefore, also in this case, the complementary probability of a mistrial,

$1 - P[U|G] = 1 - \left[\tilde{H}(e_I|G)\right]^N - \left[1 - \tilde{H}(e_I|G)\right]^N$, increases with the jury size. Jury size and the BARD standard, $e_I$, conjunctly determine the probability of a unanimous verdict and a mistrial.

Since $\tilde{H}(e_I|I)$ and $\left[1 - \tilde{H}(e_I|I)\right]$ are respectively increasing and decreasing in $e_I$, if the BARD standard is set at a sufficiently high level, $\tilde{H}(e_I|I) > \left[1 - \tilde{H}(e_I|I)\right]$ and hence $\left[\tilde{H}(e_I|I)\right]^N > \left[1 - \tilde{H}(e_I|I)\right]^N$ will hold. That is, I-defendants are more likely to be acquitted than convicted by a unanimous jury. This is equivalent to stating that $\tilde{H}(e_I|I) > \left[1 - \tilde{H}(e_I|I)\right]$, i.e., $\tilde{H}(e_I|I) > \frac{1}{2}$. This means that, with a sufficiently high standard of proof, a juror is more likely to find an I-defendant not guilty than guilty. Formally, there exists a "lower bound" of the BARD standard, $e_I^{Min}$, such that for all $e_I > e_I^{Min}$ $\tilde{H}(e_I|I) > \left[1 - \tilde{H}(e_I|I)\right]$ holds.

In a well-functioning trial system, we also expect that a G-defendant should be more likely to be convicted than acquitted. That is, $\left[1 - \tilde{H}(e_I|G)\right]^N > \left[\tilde{H}(e_I|G)\right]^N$. This would entail $\left[1 - \tilde{H}(e_I|G)\right] > \tilde{H}(e_I|G)$, i.e., $\tilde{H}(e_I|G) < \frac{1}{2}$. In this case, a juror should be more likely to find a G-defendant guilty than not-guilty. For this result to be observed, the standard of proof $e_I$ should not be excessively high so as to satisfy $\tilde{H}(e_I|G) < \frac{1}{2}$. This allows us to identify the "upper bound" of the BARD standard, $e_I^{Max}$, such that for all $e_I < e_I^{Max}$, $\left[1 - \tilde{H}(e_I|G)\right] > \tilde{H}(e_I|G)$.

Also, we know that for all $0 < e_I < \infty$, $\left[\tilde{H}(e_I|G)\right] < \left[\tilde{H}(e_I|I)\right]$ holds. Therefore, $e_I^{Min} < e_I^{Max}$. Summing up, under plausible conditions, we expect to have a feasible range of BARD standards $e_I$ contained between the lower and upper bounds identified above:

$$e_I^{Min} < e_I < e_I^{Max}.$$

We assume that the BARD standard $e_I$ is such that the following holds:

$$e_I^{Min} \left\langle e_I \left\langle e_I^{Max} : \left[\tilde{H}\left(e_I|I\right)\right]\right\rangle \left[1-\tilde{H}\left(e_I|I\right)\right] \,\&\, \left[1-\tilde{H}\left(e_I|G\right)\right]\right\rangle \left[\tilde{H}\left(e_I|G\right)\right].$$

At this point we should note that, for any given $e_I$ and $N$, the exact value of the various probabilities defined above will depend on the characteristics and complexities of the case, as captured by the functions $H\left(e_I|e\right)$ and $\tilde{H}\left(e_I|I\right)$. The qualitative results in Sects. 3 and 4 hold across cases of varying complexity, whether low or high. In Sect. 5, when comparing error ratios across states, we initially assume similar case complexity across jurisdictions. In our concluding remarks, we revisit this assumption and outline key directions for future research.

### 3.3 Decisiveness: jury size vs. BARD standard

In this section we examine the effect of jury size and BARD standards on the decisiveness of the jury process (in Sect. 3.4 we will consider their effects on the correctness of the verdict). For analytical clarity, we will look at the effects on I-defendants and G-defendants separately, and later consider the aggregate effects.

#### 3.3.1 I-Defendants

Let us begin by considering the effect of changes in jury size and BARD standards on the decisiveness of the jury process for I-defendants. We can do so by looking at the probability of a unanimous verdict—a unanimous acquittal or a unanimous conviction. At time $t=1$, or before evidence is presented at trial, the probability of acquittal (acquitting an I-defendant, in this case) is $P\left(A|I\right)=\left(\tilde{H}\left(e_I|I\right)\right)^N$. The expected number of correct acquittals is $I \times \left[\tilde{H}\left(e_I|I\right)\right]^N$. The corresponding probability of a (wrongful) conviction in this case is $P\left(C|I\right)=\left[1-\tilde{H}\left(e_I|I\right)\right]^N$, and the expected number of wrongful convictions is $I \times \left[1-\tilde{H}\left(e_I|I\right)\right]^N$.

Let $P\left[\left(U|I\right)\right]=\left[\tilde{H}\left(e_I|I\right)\right]^N+\left[1-\tilde{H}\left(e_I|I\right)\right]^N$ denote the probability of a unanimous verdict for an I-defendant. $P\left(A|I\right)$, $P\left(C|I\right)$ and $P\left[\left(U|I\right)\right]$ are all decreasing in jury size. Also, the total number of unanimous verdicts becomes:

$$I \times P\left[\left(U|I\right)\right]=I \times \left[\left(\tilde{H}\left(e_I|I\right)\right)^N+\left(1-\tilde{H}\left(e_I|I\right)\right)^N\right].$$

The total number of unanimous verdicts also decreases with the jury size.

In contrast, the effect of BARD standard on the decisiveness of a jury is, at first impression, indeterminate. However, given that $\tilde{H}\left(e_I|I\right) \geq \left[1-\tilde{H}\left(e_I|I\right)\right]$, i.e., I-defendants are at least as likely to be acquitted by a juror as to be convicted, the probability of a unanimous verdict turns out to be increasing in $e_I$. Here is why,

$$\frac{\mathrm{d\ P}\left[\mathrm{U|I}\right]}{\mathrm{d}\ e_I}=N\left[1-\tilde{H}\left(e_I|I\right)\right]^{N-1}\left(-\frac{\partial\tilde{H}\left(e_I|I\right)}{\partial e_I}\right)+N\left[\tilde{H}\left(e_I|I\right)\right]^{N-1}\left(\frac{\partial\tilde{H}\left(e_I|I\right)}{\partial e_I}\right)$$

$$= N\left(\frac{\partial \tilde{H}\left(e_I|I\right)}{\partial e_I}\right)\left(\left[\tilde{H}\left(e_I|I\right)\right]^{N-1} - \left[1 - \tilde{H}\left(e_I|I\right)\right]^{N-1}\right) \tag{1}$$

$N > 1$ and $\tilde{H}\left(e_I|I\right) \geq \left[1 - \tilde{H}\left(e_I|I\right)\right]$, so $\left[\tilde{H}\left(e_I|I\right)\right]^{N-1} - \left[1 - \tilde{H}\left(e_I|I\right)\right]^{N-1} > 0$ holds. Also, $\frac{\partial \tilde{H}(e_I|I)}{\partial e_I} > 0$, thus $\frac{\mathrm{d}\ \mathrm{P[U]}}{\mathrm{d}\ e_I} > 0$. In other words, as far as I-defendants are concerned, decisiveness in terms of the expected number of unanimous verdicts increases with $e_I$. We thus observe the following:

*(I) Effects of Jury Size and BARD on Decisiveness*. An increase in jury size has a negative effect on the decisiveness of the jury: the total number of unanimous verdicts (correct acquittals and wrongful convictions) decreases with the jury size. In contrast, the net effect of an increase in the BARD standard is a desirable increase in the decisiveness of the jury (a larger increase in correct acquittals and a smaller decrease in wrongful convictions).

### 3.3.2 G-Defendants

We now proceed to consider the effect of changes in jury size and BARD standards on the decisiveness of the jury process for G-defendants. As before, we consider the probability of a unanimously correct verdict (convicting a G-defendant, in this case), at time $t = 1$ —i.e., $P\left(C|G\right) = \left[1 - \tilde{H}\left(e_I|G\right)\right]^N$—and the probability of a unanimously wrongful acquittal of a G-defendant—i.e., $P\left(A|G\right) = \left[\tilde{H}\left(e_I|G\right)\right]^N$. Let,$P\left[U|G\right] = \left[1 - \tilde{H}\left(e_I|G\right)\right]^N + \left[\tilde{H}\left(e_I|G\right)\right]^N$ denote the probability of a unanimous verdict in case of G-defendant.

In view of the above, the total number of unanimous verdicts is:

$$G \times \left[\left[1 - \tilde{H}\left(e_I|G\right)\right]^N + \left[\tilde{H}\left(e_I|G\right)\right]^N\right].$$

The probability of a unanimous verdict as well as the total number of unanimous verdicts decrease with the jury size.

Unlike with I-defendants, the effect of the BARD standard on the decisiveness of the jury seems indeterminate. Note that:

$$\frac{dP\left[U\right]}{de_I} = N\left[1 - \tilde{H}\left(e_I|G\right)\right]^{N-1}\left(-\frac{\partial \tilde{H}\left(e_I|G\right)}{\partial e_I}\right) + N\left[\tilde{H}\left(e_I|G\right)\right]^{N-1}\left(\frac{\partial \tilde{H}\left(e_I|G\right)}{\partial e_I}\right)$$

Thus:

$$\frac{dP\left[U\right]}{de_I} = N\left(\frac{\partial \tilde{H}\left(e_I|I\right)}{\partial e_I}\right)\left(\left[\tilde{H}\left(e_I|G\right)\right]^{N-1} - \left[1 - \tilde{H}\left(e_I|G\right)\right]^{N-1}\right) \gtrless 0 \tag{2}$$

$N > 1$ and $\frac{\partial \tilde{H}(e_I|I)}{\partial e_I} > 0$, but $\left[1 - \tilde{H}\left(e_I|G\right)\right] \gtrless \tilde{H}\left(e_I|G\right)$ can hold. Therefore, increasing the BARD standard has an indeterminate effect on jury decisiveness. However, this indeterminacy is resolved if our assumption $\left[1-\tilde{H}\left(e_I|G\right)\right] > \tilde{H}\left(e_I|G\right)$ holds. Therefore, in our setup $\frac{\mathrm{d}\ \mathrm{P[U]}}{\mathrm{d}\ e_I} < 0$. $\left[1-\tilde{H}\left(e_I|G\right)\right] \left\langle \tilde{H}\left(e_I|G\right) \frac{\mathrm{d}\ \mathrm{P[U]}}{\mathrm{d}\ e_I} \right\rangle 0$.

*(II) Effects of Jury Size and BARD on Decisiveness*. An increase in jury size has a negative effect on the decisiveness of the jury: the expected number of unanimous verdicts (correct acquittals and wrongful convictions) decreases with jury size. An increase in the BARD standard also has a net negative effect on the decisiveness of the jury: the total number of unanimous verdicts decreases with the BARD standard (a larger decrease in correct convictions and a smaller increase in wrongful acquittals).

Summing up the above results, we can formulate the following Propositions.

**Proposition 1:** *Increases in jury size have an unambiguous negative effect on decisiveness for both I-defendants and G-defendants.*

However, the effect of an increase in BARD standard on decisiveness is ambiguous. Specifically, Proposition 2 holds.

**Proposition 2:** *An increase in the BARD standard has an ambiguous effect on decisiveness. Specifically, it increases decisiveness for I-defendants and decreases it for G-defendants.*

As shown above, an increase in BARD standard improves decisiveness for I-defendants. However, $\frac{\mathrm{d}\ \mathrm{P[U]}}{\mathrm{d}\ e_I} < 0$. The result is a decrease in decisiveness for G-defendants leading to Proposition 2.

Unlike the case of very high BARD standard, at moderate levels the overall effect of changes in BARD standards will depend on the proportion of I-defendants to G-defendants in the criminal justice system. To see this, note that the total number of unanimous verdicts is given by:

$$I \times P\left[(U|I)\right] + G \times P\left[(U|G)\right].$$

Differentiating this with respect to $e_I$

$$I \times \frac{\mathrm{d}\ \mathrm{P}\left[\mathrm{U|I}\right]}{\mathrm{d}\ e_I} + G \times \frac{\mathrm{d}\ \mathrm{P}\left[\mathrm{U|G}\right]}{\mathrm{d}\ e_I}$$

From above, $\frac{\mathrm{d}\ \mathrm{P[U|I]}}{\mathrm{d}\ e_I} > 0$. Also, $e_I < e_I^{Max}$, so $\frac{\mathrm{d}\ \mathrm{P[U|G]}}{\mathrm{d}\ e_I} < 0$. Therefore, the overall effect of $e_I$ on decisiveness will thus depend on the proportion of the two types of defendants.[31]

[31] It is pertinent to note that as a consequence of informational cascades, a dissenting juror is more likely to be persuaded by the majority view than holding on to their own view when the majority group is larger. So, we may expect the probability of mistrial to decrease with jury size and the corresponding decisiveness to increase with jury size. This will not affect our qualitative results. Without further specification, it is not possible to conjecture at a theoretical level about the net effects of informational cascades on accuracy and correctness.

### 3.4 Correctness: jury Size vs. BARD standard

We now proceed to study the effects of jury size and BARD standards on the correctness of jury verdicts. For analytical clarity, we shall do this in two steps, first considering effects on I-defendants and then considering G-defendants.

#### 3.4.1 I-Defendants

The effect of changes in jury size and BARD standards on the correctness of jury verdicts for I-defendants can be considered by looking at the probability of a correct verdict (acquitting an I-defendant, in this case) at time $t=1$, $P(A|I)=(\tilde{H}(e_I|I))^N$ and at the resulting number of expected correct acquittals, $I\times\left[\tilde{H}(e_I|I)\right]^N$.[32] The corresponding probability of a wrongful conviction in this case is $P(C|I)=\left[1-\tilde{H}(e_I|I)\right]^N$, and the expected number of wrongful convictions is $I\times\left[1-\tilde{H}(e_I|I)\right]^N$. $P(A|I)$ is increasing in the BARD standard $e_I$ but decreasing in the jury size $N$, whereas $P(C|I)$ is decreasing in both.

#### 3.4.2 G-Defendants

We now proceed to consider the effect of changes in jury size and BARD standards on the correctness of jury verdicts for G-defendants. We can do so by looking at the probability of a correct verdict (convicting a G-defendant, in this case), at time $t=1$, before the evidence is presented at trial, $P(C|G)=\left[1-\tilde{H}(e_I|G)\right]^N$. The expected number of correct convictions thus becomes $G\times\left[1-\tilde{H}(e_I|G)\right]^N$. The probability of a wrongful acquittal of a G-defendant is $P(A|G)=\left[\tilde{H}(e_I|G)\right]^N$ and the expected number of wrongful acquittals becomes $G\times\left[\tilde{H}(e_I|G)\right]^N$. Wrongful acquittals, $P(A|G)$, are increasing in the BARD standard, $e_I$, but decreasing in the jury size, $N$, whereas correct convictions, $P(C|G)$, are decreasing in both.

Here we can see that an increase in the BARD standard, $e_I$, leads to an increase in the probability of wrongful acquittals and to a decrease in the probability of correct convictions. These are two undesirable effects that reduce the overall correctness of jury verdicts.

#### 3.4.3 Effects of BARD on correctness

In view of the above, for I-defendants, an increase in the BARD standard has two positive effects: (i) an increase in correct acquittals, and (ii) a decrease in wrongful convictions. Both effects are desirable and lead to an overall increase in correctness. For G-defendants, an increase in the BARD standard has two undesirable effects: (i) an increase in wrongful acquittals, and (ii) a decrease in correct convictions.

**Proposition 3:** *For I-defendants, an increase in the BARD standard has a positive effect on correctness. For G-defendants, increases in the BARD standard has a negative effect on correctness.*

[32] Recall that evidence has not yet been presented as of time t=1.

In other words, the BARD standard poses the familiar trade-off between the two types of errors—wrongful conviction vs. wrongful acquittal. The aggregate effect of an increase in the BARD standard correctness will depend on the population shares of the two types.

### 3.4.4 Effects of jury size on correctness

For I-defendants, an increase in jury size, on the contrary, has two countervailing effects on the correctness of jury verdicts: (i) a desirable decrease in wrongful convictions, and (ii) an undesirable decrease in correct acquittals. Therefore, to measure the overall effect of an increase in jury size on correctness, we must examine and compare the two countervailing effects. First, we prove the following result.

**Lemma 1:** *As jury size increases, the reduction in the number of correct acquittals for I-defendants exceeds the reduction in their wrongful convictions.*

Proof: See Appendix 1.

The lemma suggests that increasing jury size reduces overall correctness because the undesirable decline in correct acquittals outweighs the beneficial reduction in wrongful convictions. Consequently, for I-defendants, a larger jury size negatively impacts correctness.

An increase in jury size has two countervailing effects on the correctness of jury verdicts for G-defendants as well: (i) a desirable decrease in wrongful acquittals, and (ii) an undesirable decrease in correct convictions. To show this, first, we prove the following result.

**Lemma 2:** *As the jury size increases, the fall in the number of correct convictions of G-defendants is greater than the fall in wrongful acquittals.*

Proof: See Appendix 1.

The lemma suggests that when a G-defendant is more likely to be found guilty than not, an increase in jury size reduces correctness because the (undesirable) decrease in correct verdicts is greater than the (desirable) decrease in incorrect verdicts.

Lemmas 1 and 2 lead to the following results.

**Proposition 4:** *Increases in jury size have a net negative effect on correctness. Larger juries unambiguously reduce correct convictions more than they reduce wrongful acquittals.*

This proposition shows that in well-functioning legal system (with $[\widetilde{H}(e_I|I)] > \left[1-\widetilde{H}(e_I|I)\right]$ & $\left[1-\widetilde{H}(e_I|G)\right] > \left[\widetilde{H}(e_I|G)\right]$) an increase in jury size unambiguously leads to a decrease in correctness of unanimous jury verdicts. Contrary to Condorcet's prediction, this means that that the correctness and decisiveness of unanimous juries do not increase with greater jury size.

Summing up, in general, jury size and the BARD standard are not substitutes. They work differently vis-à-vis the correctness of unanimous jury verdicts for I-defendants. An increase in jury size reduces the chances of a correct acquittal for I-defendants, whereas an increase in BARD increases the correctness of unanimous jury trials. For G-defendants, increases in BARD standards have a negative effect on correctness, while the jury size has an indeter-

**Table 2** Decisiveness and correctness

| | Increase in Jury Size | | Increase in BARD | |
|---|---|---|---|---|
| | Decisiveness | Correctness | Decisiveness | Correctness |
| I-Defendants | ↓ | ↓ | ↑ | ↑ |
| G-Defendants | ↓ | ↓ | ↓ | ↓ |
| Ratio of G/I Defendants | ↓/↓ | ↓/↓ | ↓/↑ | ↓/↑ |

minate effect on correctness. Increases in jury size have a net negative effect on correctness for both types of defendants.

Table 2 summarizes the previously stated results regarding the effects of jury size and BARD standards on the decisiveness and correctness of the jury deliberation. Table 2 illustrates that an increase in jury size reduces both jury decisiveness for I-defendants and G-defendants alike. These decreases result regardless of whether the number of I-defendants is larger or smaller than the G-defendants. Moreover, an increase in jury size reduces correctness both for I-defendants and G-defendants. Meanwhile, increasing the BARD standard increases in jury decisiveness and correctness with respect to I-defendants. For G-defendants, increase in BARD standard reduces decisiveness. The aggregate effects are variable depending on the number of I-defendants vis-à-vis G-defendants.

Reducing jury size, $N$, enhances decisiveness for *both* G-defendants and I-defendants. Additionally, in our set up (with $[\widetilde{H}(e_I|I)] > \left[1-\widetilde{H}(e_I|I)\right]$ & $\left[1-\widetilde{H}(e_I|G)\right] > \left[\widetilde{H}(e_I|G)\right]$), a reduction in jury size also increases accuracy for both types.

Conversely, a reduction in BARD standards leads to mixed results. Thus, starting from reasonable levels of BARD standard, given the choice between adjusting either or both policy variables—jury size and BARD standards—reducing jury size appears to be the more favorable option for states. However, as discussed in Sect. 4, the effects of jury size and BARD standards can converge when focusing on the accuracy of trial outcomes rather than the correctness of individual jury verdicts.

## 4 Accuracy of the jury process and the relevance of retrials

When a jury fails to reach a unanimous decision on the verdict, a mistrial occurs. This form of mistrial is technically referred to as a "hung-jury mistrial," which distinguishes it from mistrials declared by the judge for other procedural reasons.[33] A hung-jury mistrial does not constitute an acquittal for the defendant, and the case does not necessarily conclude at that time. The prosecution may choose to retry the case, which would lead to a new trial starting

[33] Courts have the discretion to declare a mistrial in other circumstances where the integrity of the legal process and the defendant's right to a fair trial have been compromised, such as (i) prejudicial jury misconduct, (ii) serious error in procedure or law; (iii) illness or incapacitation of a participant (judge, attorney, or a juror); (iv) prosecutorial or defense misconduct; and (v) jury tampering or intimidation. For an interesting discussion of deadlocked juries and their impact on the justice system, see also Flynn (1977) and Hannaford-Agor et al. (1999).

with the selection of a new jury, presentation of evidence and legal arguments, and closing statements. Alternatively, the prosecution may decide to not pursue the case further. In such instances, the case ends without resolution. In this latter scenario, no criminal sanctions are imposed on the criminal defendant.[34] In other cases, after a mistrial the prosecution and defense may enter negotiations for a plea bargain. This could result in the defendant either pleading guilty to a lesser charge or receiving a reduced sentence in exchange for avoiding a new trial.

The prosecutor's office enjoys complete discretion in deciding which cases, if any, the office will pursue for retrial. Jurisdictions do not provide prosecutors with any guidelines for determining how often and under which circumstances cases should be retried.[35] Research carried out by the National Institute of Justice suggests that only 32% of cases that ended in hung-jury mistrial were filed for a retrial, and less than half of the cases that were refiled advanced to a full trial. The likelihood that a jury will hang twice in the same case is only 2.4% (Hannaford-Agor 2002).[36] From an exculpatory point of view, a criminal defendant would naturally prefer a unanimous "not guilty" verdict rather than a mistrial without a verdict. However, when a hung-jury mistrial occurs and the prosecution decides not to pursue the case further, the mistrial produces effects that are de facto equivalent to an acquittal—the charges against the defendant are effectively dropped and no criminal sanctions are imposed on the criminal defendant.[37]

The incidence of hung juries and the determination of what happens after a mistrial is of great relevance for understanding the impact of jury size and BARD standards on the correctness and decisiveness of the jury process. Our analysis will first examine the impact of mistrials on jury accuracy in the more common scenario where hung-jury cases are never brought up for a retrial. Focusing on this case is justified in view of the National Institute of Justice's findings that, on average, only a small fraction (5–13%) of trials end up with hung

[34] Kalven and Zeisel (1966) interestingly observe that the possibility of deadlocks is an important, and often overlooked, differentiating factor between jury trials and bench trials. In their extensive research on the American Jury, based on 3,576 trial questionnaires filled out by trial judges throughout the United States, the authors asked judges how they would have decided the same case in the absence of a jury. The survey revealed that judges agreed with the decision of the jury in only 75% of the cases, disagreeing with the jury deliberation in 20% of the cases. The remaining 5% difference between jury verdicts and bench decisions was driven by the possibility of leaving the case undecided through a hung jury, which has no corresponding possibility in a bench trial.

[35] Quite surprisingly, jurisdictions do not keep records of mistrial rates or retrial rates. The logistical reason for this is that a hung jury is not a final case disposition—if there is a subsequent action like a plea bargain or a retrial, that information replaces the hung-jury data. So, determining the aggregate frequency of hung juries is virtually impossible, as is determining the frequency of what follows hung juries. We are grateful to Scott Dewey, Faculty Librarian at the University of Minnesota Law Library, for his painstaking effort to research retrial rates in U.S. state jurisdictions and to Paula Hannaford-Agor, Director of the Center for Jury Studies at the National Center for State Courts, for sharing and explaining her challenging experience in collecting data for their 2002 hung jury study, in which the state general jurisdiction courts of the 75 most populous counties in the country were contacted in an effort to collect information on felony case dispositions. See Hannaford-Agor et al. (2002) and Hans et al. (2003).

[36] Of those cases that resulted in a second jury trial, 69% resulted in a conviction, 19% in acquittal, 8% in a second hung jury, and 4% resulted in a mistrial for reasons other than a hung jury. (Hannaford-Agor 2002). Multiple retrials are very rarely observed.

[37] In the early 1980s some American legal scholars suggested that retrials after hung juries create a double jeopardy issue (Findlater 1981). However, the U.S. Supreme Court in *Richardson v. United States* 468 U.S. 317, 104 S. Ct. 3081 (1984) decided that a mistrial did not preclude a retrial. Some subsequent suggestions on preventing retrials after hung juries have therefore focused on procedural rules rather than constitutional challenges (Gelman 2024).

jury,[38] and just 32% of cases ending in mistrial result in another, subsequent jury trial. When hung-jury cases are not retried, a mistrial effectively grants de facto impunity to the defendant and drives a wedge between the correctness of unanimous jury verdicts and the overall accuracy of trial outcomes.[39] This point becomes especially salient when contrasted with Condorcet's jury theorem (de Caritat 1785), which holds that increasing jury size always improves the accuracy of collective decisions.[40] As discussed below, the apparent tension between our results and Condorcet's theorem arises from a key difference in institutional design: Condorcet assumes decisions made under a majority rule, whereas we consider unanimous verdicts.[41]

Building on these results, this section examines the combined effects of correctness and decisiveness on accuracy. Recall that in our framework, "correctness" measures the conditional probability that, if a verdict is rendered it is factually accurate, whereas "accuracy" is the unconditional probability that the trial process both produces a verdict and that the verdict is correct (i.e., is the ratio of the probability of acquitting over the probability of convicting defined for either the innocent or the guilty). We will show that, in the absence of retrials, the overall accuracy of the jury process declines with larger jury sizes.[42] We will carry out a similar analysis to study the effect of changes in BARD standards on the accuracy of jury outcomes. As it will be shown, although an increase in the BARD standard always increases the correctness of verdicts, its effect on the accuracy of trial outcomes is indeterminate and hinges upon the number of I-defendants vis-à-vis G-defendants.

We proceed with our analysis by focusing on the effects of mistrials in the limiting case where prosecutors opt not to pursue retrials or plea bargains. For analytical clarity, we will look at the effects on I-defendants and G-defendants separately and later consider the aggregate effects.

*I-Defendants*

[38] Hannaford-Agor, P.L. (2002, September 30). *Are Hung Juries a Problem?* National Center for State Courts. https://www.ncsc-jurystudies.org/__data/assets/pdf_file/0018/6138/hung-jury-final-report.pdf.

[39] In the United States, retrials can occur if a trial ends in a hung jury and the prosecutor subsequently decides to try the case a second time (Gelman 2023). According to studies sponsored by the National Institute of Justice which reviewed cases taking place between 1996 and 1998, 32% of cases ending in mistrial as the result of a hung jury resulted in another, subsequent jury trial, 21.6% of these cases are dismissed, and 31.8% result in a guilty plea (Hannaford-Agor 2002).

[40] Marie-Jean-Antoine-Nicolas de Caritat (1743–94), generally known as the Marquis de Condorcet, considered the process of jury deliberation. He proved that, if individual jurors are more likely than not of being correct in their convictions, an increase in the number of jurors will increase the chance that the collective majority decision will be correct (de Caritat 1785). This theorem, which can be seen as a consequence of the law of large numbers, has played an important role in jury design, providing a strong theoretical basis for arguments in support of larger juries.

[41] In Condorcet's framework, juries decide by majority rule rather than unanimity. The dynamics of decision-making change significantly under a unanimity rule, as larger juries are less likely to reach unanimous consensus. This raises important questions about the optimal jury size and the trade-offs between unanimous and majority decision rules. While unanimity can enhance deliberation, it may also increase the likelihood of hung juries, thereby reducing decisiveness of trials (Mueller 2003, p. 128). The gap between verdict correctness and outcome accuracy diminishes with the introduction of retrials, ultimately disappearing entirely when cases are retried indefinitely, until a unanimous verdict is reached.

[42] Recall that in our framework, "correctness" measures the conditional probability that, if a verdict is rendered it is factually accurate, whereas "accuracy" is the unconditional probability that the trial process both produces a verdict and that the verdict is correct.

When a mistrial effectively produces the results of an acquittal—i.e., no retrials or plea bargains follow a mistrial and no criminal sanctions are imposed on the criminal defendant—the probability that an I-defendant avoids punishment becomes:

$$P\left(NC|I\right) = \tilde{H}\left(e_I|I\right) + \left[1 - \tilde{H}\left(e_I|I\right) - \left[1 - \tilde{H}\left(e_I|I\right)\right]\right] = 1 - \left[1 - \tilde{H}\left(e_I|I\right)\right]^N.$$

The complementary probability of a wrongful conviction of an I-defendant remains $\left[1 - \tilde{H}\left(e_I|I\right)\right]^N$.[43] The probability of wrongful conviction, $P\left(C|I\right)$, is decreasing in jury size, $N$, as well as in the BARD standard, $e_I$. Meanwhile, the probability of a correct non-conviction, $P\left(NC|I\right)$, is increasing in both.

For any given threshold, $e_I$, an increase in jury size has two desirable effects: a decrease in the probability of wrongful conviction, $P\left(C|I\right)$, and an increase in the probability of correct non-conviction, $P\left(NC|I\right)$, of innocent defendants. Similarly, for any given jury size, an increase in the BARD standard, $e_I$, improves the accuracy of trial outcomes because $P\left(C|I\right)$ is decreasing in $e_I$ and $P\left(NC|I\right)$ is increasing in $e_I$.

The accuracy of trial outcomes can be taken as the ratio of the number of effectively correct acquittals over the number of wrongful convictions, i.e.[44];

$$\frac{1 - \left[1 - \tilde{H}\left(e_I|I\right)\right]^N}{\left[1 - \tilde{H}\left(e_I|I\right)\right]^N}.$$

The above accuracy ratio is increasing in the jury size, $N$, as well as the BARD standard, $e_I$.

*Effects of Jury Size and BARD on the accuracy of Trial Outcomes*. An increase in jury size or BARD standard has two desirable effects: (i) an increase in correct acquittals, and (ii) a decrease in wrongful convictions. So, as far as I-defendants are concerned, both jury size and BARD standard increase the accuracy of trial outcomes.

---

[43] Based on the available evidence, let's assume that there is a single retrial and that no hung jury occurs in this retrial. Additionally, let's assume that the retrial follows the same judicial procedures as the initial trial. Under these conditions, and given the occurrence of a retrial, the probability of a unanimous yet wrongful conviction of an I-defendant is $\left[1 - \tilde{H}\left(e_I|I\right)\right]^N$. So, the overall probability that an I-defendant is acquitted at trial or retrial is $\left[\tilde{H}\left(e_I|I\right)\right]^N + \left[\tilde{H}\left(e_I|I\right)\right]^N \left[1 - \tilde{H}\left(e_I|I\right)^N - \left[1 - \tilde{H}\left(e_I|I\right)\right]^N\right]$. The overall probability of convicting an I-defendant is $\left[1 - \tilde{H}\left(e_I|I\right)\right]^N + \left[1 - \tilde{H}\left(e_I|I\right)\right]^N \left[1 - \tilde{H}\left(e_I|I\right)^N - \left[1 - \tilde{H}\left(e_I|I\right)\right]^N\right]$. Our assumption that a mistrial without a retrial or a plea bargain is de facto equivalent to an acquittal, means that the term $\left[1 - \tilde{H}\left(e_I|I\right)\right]^N \left[1 - \tilde{H}\left(e_I|I\right)^N - \left[1 - \tilde{H}\left(e_I|I\right)\right]^N\right] \approx 0$. Symmetrically, for G-defendants, this assumption means that $\left[1 - \tilde{H}\left(e_I|G\right)\right]^N \left(1 + \left[1 - \tilde{H}\left(e_I|G\right)^N - \left[1 - \tilde{H}\left(e_I|G\right)\right]^N\right]\right) \approx 0$.

[44] This claim holds more broadly even in cases where hung-jury mistrials are retried indefinitely until a unanimous verdict is reached, i.e., without assuming $\left[1 - \widetilde{H}\left(e_I|I\right)\right]^N \left[1 - \widetilde{H}\left(e_I|I\right)^N - \left[1 - \widetilde{H}\left(e_I|I\right)\right]^N\right] \approx 0$ (refer to footnote 42 above). Indeed, it can be verified that when mistrials are regularly retried, the accuracy ratio for I-defendants is $\frac{\left[\tilde{H}\left(e_I|I\right)\right]^N}{\left[1-\tilde{H}\left(e_I|I\right)\right]^N}$, which, in view of $\tilde{H}\left(e_I|I\right) > 1 - \tilde{H}\left(e_I|I\right)$, is increasing in $N$, as well as, $e_I$. A similar argument applies to the accuracy ratio for G-defendants.

*G-Defendants*

When mistrials are not followed by a retrial or plea bargains, the probability of conviction for a G-defendant is $\left[1-\widetilde{H}\left(e_I|G\right)\right]^N$. The complementary probability of wrongful non-conviction for a G-defendant becomes:

$$P\left(NC|G\right)=\left[\tilde{H}\left(e_I|G\right)\right]^N+\left[1-\left[\tilde{H}\left(e_I|G\right)\right]^N-\left[1-\tilde{H}\left(e_I|G\right)\right]^N\right]=\left[1-\left[1-\tilde{H}\left(e_I|G\right)\right]^N\right].$$

In this case, for any given BARD standard, an increase in jury size leads to an undesirable decrease in the probability of conviction for G-defendants. This is accompanied by an undesirable increase in the probability of wrongful non-conviction for G-defendants. Similarly, for any given jury size, an increase in the BARD standard reduces the expected accuracy of trial outcomes for G-defendants.

We can see these effects in a more compact form by looking at the accuracy ratio introduced above:

$$\frac{\left[1-\tilde{H}\left(e_I|G\right)\right]^N}{1-\left[1-\tilde{H}\left(e_I|G\right)\right]^N}.$$

It can be verified that the above accuracy ratio is decreasing in the jury size, $N$, as well as in the BARD standard, $e_I$.

Summing up the above results, we formulate the following Proposition.

**Proposition 6:** *When mistrials are not followed by a retrial, for I-defendants: (i) increases in jury size and (ii) increases in the BARD standard have positive effects on the accuracy of trial outcomes. For G-defendants: (i) increases in jury size and (ii) increases in the BARD standard have negative effects on the accuracy of trial outcomes.*

## 5 Aggregate effects

Proposition 6 captures the trade-off between the desire to protect innocent defendants and the need to hold guilty defendants accountable. When a mistrial is not followed by a retrial, changes in jury size and BARD standards trigger these counterbalancing effects.

However, depending on the number of I-defendants vis-à-vis G-defendants, an increase in jury size or the BARD standard can have desirable or undesirable consequences. To see this, note that the number of correct convictions and correct non-convictions is:

$$I\times\left[1-\left[1-\tilde{H}\left(e_I|I\right)\right]^N\right]+G\times\left[1-\tilde{H}\left(e_I|G\right)\right]^N$$

The first term is increasing in N whereas the second term is decreasing in N. Therefore, the overall effect of an increase is jury size on the accuracy of trial outcomes will depend on the number of I-defendants vis-à-vis G-defendants.

A similar effect can be observed with respect to the BARD standard. The overall effect of an increase in the BARD standard on accuracy hinges upon the number of I-defendants vis-à-vis G-defendants. An increase in the BARD standard is desirable to mitigate false conviction problems. Otherwise, an increase in the BARD standard has an indeterminate effect on the number of accurate outcomes.

However, under reasonable conditions, the proportion of G-defendants is expected to dominate. When most of the accused are G-defendants, the second term will dominate the first. In this case, contrary to the intuition suggested by Condorcet's jury theorem, an increase in jury size will reduce the probability of a correct verdict or accurate jury process. Similarly, raising the BARD standard results in fewer correct/accurate outcomes.[45]

It is worth emphasizing that the jury size and BARD are substitutes regarding accuracy of the outcomes, regardless of the share of the two types of defendants among the accused. Therefore, in order to preserve a chosen accuracy level, any decrease in $N$ should be accompanied by a corresponding increase in $e_I$, and vice versa.

As shown in Table 1 above (in Sect. 2), $N$ ranges from six to twelve. Now, consider the states with $N = 6$, such as Minnesota (F); Montana (F); Nebraska (F), and compare them to states with $N = 12$ such as Pennsylvania (F); Tennessee (M and F); Texas (F); Virginia (F). Other things remaining the same, to achieve the same level of accuracy, the former set of states should have a higher $e_I$, compared to the latter set. However, in either category of states, there are no clear instructions given to the jurors about $e_I$. Therefore, wide variations is the jury size and ambiguity regarding the BARD can cause accuracy levels to jump wildly even among the states with same stated level of accuracy objectives, an issue that has not received adequate attention in the literature.

## 6 Exposing contradictions: the consequences of jury structure reforms on blackstonian ratios

Echoing Sowell's (2011) well-regarded insight, legal policies should be evaluated by their actual outcomes, not merely by their aspirations.[46] Across the last several decades, numerous reforms to jury structure have shared a common goal: enhancing the criminal justice system. The previous analysis contributed to the analytical jury literature by illustrating how various combinations of jury size and BARD standards influence the decisiveness and correctness of the jury process, and how these, in turn, jointly impact the expected accuracy of trial outcomes.

In this section, we conclude our analysis examining how the observed variations in jury structure align with the Blackstonian ratios adopted by U.S. states, as reported in Table 1 above. Our analysis brings to light the inherent trade-offs that states face in their choices of jury size, BARD standards, and Blackstonian ratios, particularly due to the differing impacts that jury size and BARD standards have on I-defendants and G-defendants. As shown in our pre-

[45] In this section, consecutive retrials eliminate the issue of decisiveness, as every case ultimately results in either a conviction or an affirmative acquittal. Consequently, the correctness of verdicts and the accuracy of the jury process converge.

[46] Sowell (2011, ch. 3): "Economic policies need to be analyzed in terms of the incentives they create, rather than the hopes that inspired them."

vious analysis, any effort to minimize $P(G|I) = \left[1 - \tilde{H}(e_I|I)\right]^N$ by increasing $e_I$ or N also affects the probability of a wrongful acquittal for a G-defendant, $P(A|G) = \left[\tilde{H}(e_I|G)\right]^N$.

Blackstonian ratios measure the degree of protection afforded to defendants by quantifying the acceptable balance between (the probability of) convicting an innocent person $P(G|I)$ and (the likelihood of) acquitting a guilty one $P(A|G)$. For example, a jurisdiction adopting a Blackstonian ratio of $\beta = 1:10$ would require $P(G|I) = 10 \times P(A|G)$. A jurisdiction adopting a ratio of $\beta = 1:20$ would be willing to tolerate a higher probability of false acquittals relative to the probability of false convictions.

To evaluate the effect of jury size and BARD standards on Blackstonian ratios, in the absence of retrials, we can restate the Blackstonian ratio in terms of false conviction vs. false acquittal of jury verdicts as follows:

$$\beta(N, e_I) = \frac{\left[1 - \tilde{H}(e_I|I)\right]^N}{\left[\tilde{H}(e_I|G)\right]^N}.$$

Note that in our set, given the jury size and BARD standard, for any given trial we get a unique value of the Blackstonian ratio, $\beta(N, e_I)$.

Further, in the limiting case where defendants can avoid punishment as a result of a mistrial, the probability of false acquittals of guilty defendants de facto increases to $1 - \left[1 - \tilde{H}(e_I|G)\right]^N$. The probability of false conviction of an innocent defendant remains at $\left[1 - \tilde{H}(e_I|I)\right]^N$. The Blackstonian ratio therefore becomes:

$$\beta(N, e_I) = \frac{\left[1 - \tilde{H}(e_I|I)\right]^N}{1 - \left[1 - \tilde{H}(e_I|G)\right]^N}.$$

As shown in Table 1, several states adopt a strict Blackstonian ratio of $\beta = 1:1$. These jurisdictions uphold a strict criminal justice policy that gives equal priority to the punishment of wrongdoers and the safeguarding of innocent individuals. In the remaining U.S. states, the range of observed Blackstonian ratios falls below $\beta(N, e_I) = \left[\frac{1 - \tilde{H}(e_I|I)}{\tilde{H}(e_I|G)}\right]^N < 1$. In all these jurisdictions, it is clear that the Blackstonian ratio is decreasing in both $e_I$ and N. In sum, jurisdictions prioritizing the protection of innocent defendants would be expected to utilize larger juries and/or higher BARD standards, increasing the probability of false acquittals relative to false convictions.[47]

We are now well-positioned to evaluate the consistency of states' policy choices with the results of our analysis. It is worth noting that Pi et al. (2020) have exhaustively shown that state courts refrain from offering quantitative definitions of the BARD standard, and their qualitative instructions to jurors are fairly uniform across states. However, as shown

[47] These results show that the objective of achieving accuracy (defined in terms of correctness of trial verdicts and decisiveness of the jury process, as discussed above) can be in conflict with the objective of protecting innocent defendants (in terms of delivering the states' chosen Blackstonian ratios). In jurisdictions with a larger share of type-G defendants, improving accuracy requires reducing the jury size and/or the BARD standard, whereas improving protection of innocent defendants requires increase them.

in Sect. 2, choices of jury sizes and Blackstonian ratios vary widely across states. Do these variations follow an underlying logic, and are they internally consistent?

Some variations in jury sizes and Blackstonian ratios across states can be attributed to substantive policy differences. For instance, Florida utilizes juries of six with a Blackstonian ratio of 1:5, reflecting its policy objective of moderately balancing the risk of wrongful convictions against wrongful acquittals. In contrast, California employs juries of twelve with a higher Blackstonian ratio of 1:10, indicating a stronger policy emphasis on minimizing wrongful convictions. These variations are consistent with the states' criminal justice policies and respective goals of punishing offenders while safeguarding innocent individuals. However, the analysis of Table 1 highlights some glaring inconsistencies in other states' policy choices.

Take Alaska and Wyoming, for instance. Besides being alphabetical bookends, they occupy opposite poles on the Blackstonian ratio spectrum: Alaska holds to a 1:1 ratio, while Wyoming promises a staggering 1:100. Both states employ twelve-member juries for felonies and six-member juries for misdemeanors. The inconsistencies here are two-fold and should be glaringly obvious. First, to arrive at such drastically different Blackstonian ratios with identical jury structures, Alaska and Wyoming would need their juries to apply wildly divergent standards of proof—requiring Wyoming's prosecution to meet a far loftier burden than Alaska's. Going by Pi et al. (2020) this does not seem to be the case, given that courts' qualitative instructions to jurors are fairly uniform across states.

Thus, Alaska and Wyoming likely will not produce these dramatically different Blackstonian ratios unless, perhaps, their supreme courts quietly concede vast disparities in the quality of their respective law enforcement agencies. This suggests that Alaska prosecutes far more innocent defendants than Wyoming. Second, Alaska and Wyoming's policies also reveal internal contradictions. Given their stated Blackstonian ratios, it is implausible for these states to achieve consistent outcomes while employing different jury sizes—twelve members for felony cases and six for misdemeanors. To maintain identical Blackstonian ratios across both crime categories, either lower BARD standards would need to be applied to felonies than misdemeanors, or one would need to assume that police and prosecutors allocate greater resources to prevent wrongful charges in misdemeanor cases. Absent such counter-intuitive and even unplausible adjustments, it is unlikely that defendants receive comparable levels of protection across these categories, despite state supreme courts' commitments to upholding uniform Blackstonian ratios within their jurisdictions.

The contradictions extend well beyond these two examples, pervading nearly every U.S. jurisdiction that claims to adhere to fixed Blackstonian ratios. These contradictions are driven by the interdependence of the policy variables analyzed here. States cannot fully separate these variables; decisions made in one area inevitably limit choices in others, necessitating a careful balance of policy commitments. For a state intent on preserving a specific Blackstonian ratio alongside a fixed jury size, adjustments to the implicit BARD standard are essential. Conversely, if the BARD standard is set as a rigid threshold, the state must permit flexibility in either the Blackstonian ratio or in jury size to ensure coherence in its criminal justice processes. Likewise, a commitment to a particular Blackstonian ratio and BARD standard may demand a reassessment of the optimal jury size.

For instance, states with the same stated Blackstonian ratio—say 1:10—but with jury sizes ranging from twelve (in the case of California, Maine, and Michigan, among others) to a jury size of six (in the case of Florida and Wisconsin) cannot all achieve their stated objectives. Consider two similar crimes tried in two different jurisdictions—say, in California and Wisconsin. Our analysis suggests that a guilty defendant tried in California would have a relatively higher probability of wrongful acquittal than a guilty defendant tried in Wisconsin. Similar inconsistencies would arise in many other cases across U.S. jurisdictions.

While our study does not definitively explain whether these inconsistencies arise from fragmented, piecemeal decisions across jurisdictions or politically motivated choices targeting different constituencies, the pervasive mismatch between state jury structures and professed Blackstonian ratios are a serious source of divergence between the stated intents and (unintended) outcomes across several states.[48] Our analysis has focused on the common scenario in which hung-jury cases are not retried. However, even in cases where hung juries lead to repeated retrials, jury size and BARD standards remain central to determining both the accuracy of jury decisions and protecting of innocent defendants. Consequently, the significant disparity between states' jury structures and their stated adherence to Blackstonian ratios continues to underscore the issues discussed here, highlighting the need to revisit constitutional and political frameworks governing jury design. We hope this paper provides an objective basis to examine these inconsistencies, evaluate the practical impact of Blackstonian ratios, and explore pathways toward more consistent protections for defendants.

Finally, case complexity is also relevant for the trial errors. Our qualitative results in Sects. 3 and 4 hold for all types of cases with little or high complexity. In Sect. 5, while comparing error ratios across states, it is important to keep in mind that case complexity, in principle, can vary across states. We hope the future research will be to address this issue.

## Appendix 1

**Proof of** Lemma 1**:**

The probability of an accurate acquittal of an I-defendant is $(\tilde{H}\left(e_I|I\right))^N$. The number of I-defendants being accurately acquitted is $I \times (\tilde{H}\left(e_I|I\right))^N$. The probability of I-defendants being wrongly convicted is $\left[1-\tilde{H}\left(e_I|I\right)\right]^N$. The number of I-defendants being wrongfully convicted is $I \times \left[1-\tilde{H}\left(e_I|I\right)\right]^N$. The number of both accurate acquittals and wrongful convictions decreases with jury size. However, the absolute impact of an increase in jury size is different for the two numbers. In our setup, $[\tilde{H}\left(e_I|I\right)]^N > \left[1-\tilde{H}\left(e_I|I\right)\right]^N$. Differentiating $[\tilde{H}\left(e_I|I\right)]^N$ with respect to $N$ we obtain $\frac{d[\tilde{H}(e_I|I)]^N}{dN} = [\tilde{H}\left(e_I|I\right)]^N ln\left(\tilde{H}\left(e_I|I\right)\right)$. Similarly, we get $\frac{d\left[1-\tilde{H}(e_I|I)\right]^N}{dN} = \left(1-\tilde{H}\left(e_I|I\right)\right)^N ln\left(1-\tilde{H}\left(e_I|I\right)\right)$. In view of assumptions that $1/2 < \tilde{H}\left(e_I|I\right) < 1$, we have $\left[1-\tilde{H}\left(e_I|I\right)\right] < \tilde{H}\left(e_I|I\right)$. Hence, $0 < \frac{[1-\tilde{H}(e_I|I)]^N}{[\tilde{H}(e_I|I)]^N} < 1$. Also, $\left|\frac{d[\tilde{H}(e_I|I)]^N}{dN}\right| \left\langle \left|\frac{d\left[1-\tilde{H}(e_I|I)\right]^N}{dN}\right| \Leftrightarrow \tilde{H}\left(e_I|I\right)^N \ln\left(\tilde{H}\left(e_I|I\right)\right)\right\rangle$ $\left(1-\tilde{H}\left(e_I|I\right)\right)^N \ln\left(1-\tilde{H}\left(e_I|I\right)\right) \Leftrightarrow \frac{\ln\left(\tilde{H}(e_I|I)\right)}{\ln\left(1-\tilde{H}(e_I|I)\right)} < \frac{\left(1-\tilde{H}(e_I|I)\right)^N}{\tilde{H}(e_I|I)^N}$. Note that since

[48] Several factors could plausibly explain the observed differences in parameters across states, including variations in police enforcement levels, prosecutorial filing rates, retrial rates, selection mechanisms for trial judges (election vs. appointment), and broader social structures. Attempting to identify any specific cause would be purely speculative and beyond the scope of this article.

$0 < \left[1 - \tilde{H}\left(e_I|I\right)\right]^N < [\tilde{H}\left(e_I|I\right)]^N < 1$, so $0 < \frac{ln\left(\tilde{H}(e_I|I)\right)}{ln\left(1-\tilde{H}(e_I|I)\right)} < 1$. Summing up, $\left|\frac{d[\tilde{H}(e_I|I)]^N}{dN}\right| < \left|\frac{d\left[1-\tilde{H}(e_I|I)\right]^N}{dN}\right|$ holds if and only if $\frac{ln\left(\tilde{H}(e_I|I)\right)}{ln\left(1-\tilde{H}(e_I|I)\right)} < \frac{\left(1-\tilde{H}(e_I|I)\right)^N}{\tilde{H}(e_I|I)^N}$. Similarly, it can be seen that $\left|\frac{d[\tilde{H}(e_I|I)]^N}{dN}\right| > \left|\frac{d\left[1-\tilde{H}(e_I|I)\right]^N}{dN}\right|$ holds if and only if $\frac{\left(1-\tilde{H}(e_I|I)\right)^N}{\tilde{H}(e_I|I)^N} < \frac{ln\left(\tilde{H}(e_I|I)\right)}{ln\left(1-\tilde{H}(e_I|I)\right)}$. Note that $\lim_{N\to\infty} \frac{\left(1-\tilde{H}(e_I|I)\right)^N}{\tilde{H}(e_I|I)^N} = 0$ and $\lim_{N\to 0} \frac{\left(1-\tilde{H}(e_I|I)\right)^N}{\tilde{H}(e_I|I)^N} = 1$. Specifically, it can be seen that for $N \geq 2$, $\frac{ln\left(\tilde{H}(e_I|I)\right)}{ln\left(1-\tilde{H}(e_I|I)\right)} > \frac{\left(1-\tilde{H}(e_I|I)\right)^N}{(\tilde{H}(e_I|I))^N}$ holds for all $\frac{1}{2} < \left(\tilde{H}\left(e_I|I\right)\right) < 1$.[49] That is, for the plausible range of $\tilde{H}\left(e_I|I\right)$ and $N$, we obtain $I \times \left|\frac{d[\tilde{H}(e_I|I)]^N}{dN}\right| > I \times \left|\frac{d\left[1-\tilde{H}(e_I|I)\right]^N}{dN}\right|$. Therefore, the effect of an increase in jury size is a decrease in accuracy, because the decrease in the number of correct acquittals is larger than the decrease in the number of wrongful convictions.

**Proof of** Lemma 2**:**

The probability that a G-defendant is wrongly acquitted is $(\tilde{H}\left(e_I|G\right))^N$, where $0 < [(\tilde{H}\left(e_I|G\right))]^N < \frac{1}{2}$. The number of G-defendants being wrongfully acquitted is $G \times [\tilde{H}\left(e_I|G\right)]^N$. The probability of G-defendants being accurately convicted is $\left[1 - \tilde{H}\left(e_I|G\right)\right]^N$. The number of G-defendants being accurately convicted is $G \times [1 - \tilde{H}\left(e_I|G\right)]^N$. It can be seen that for a G-defendant, $\left|\frac{d[\tilde{H}(e_I|G)]^N}{dN}\right| \left\langle \left|\frac{d\left[1-\tilde{H}(e_I|G)\right]^N}{dN}\right| \Leftrightarrow \tilde{H}\left(e_I|G\right)^N ln\left(\tilde{H}\left(e_I|G\right)\right)\right\rangle$ $\left(1 - \tilde{H}\left(e_I|G\right)\right)^N ln\left(1 - \tilde{H}\left(e_I|G\right)\right) \Leftrightarrow \frac{ln\left(1-\tilde{H}(e_I|G)\right)}{ln\left(\tilde{H}(e_I|G)\right)} > \frac{(\tilde{H}(e_I|G))^N}{\left(1-\tilde{H}(e_I|G)\right)^N}$. And, $\left|\frac{d[\tilde{H}(e_I|G)]^N}{dN}\right| > \left|\frac{d\left[1-\tilde{H}(e_I|G)\right]^N}{dN}\right| \Leftrightarrow \frac{ln\left(1-\tilde{H}(e_I|G)\right)}{ln\left(\tilde{H}(e_I|G)\right)} < \frac{(\tilde{H}(e_I|G))^N}{\left(1-\tilde{H}(e_I|G)\right)^N}$. We find that for any $N \geq 2$ and $0 < \left(\tilde{H}\left(e_I|G\right)\right) < \frac{1}{2}$, $\left|\frac{d[\tilde{H}(e_I|G)]^N}{dN}\right| < \left|\frac{d\left[1-\tilde{H}(e_I|G)\right]^N}{dN}\right|$ holds. Simply put, the effect of an increase in jury size is a reduction in accuracy because the decrease in the number of correct convictions is larger than the decrease in the number of wrongful acquittals.

[49] $N = 1$, $\frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)} < \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N}$ holds. So, $\left|\frac{d[\widetilde{H}(e_I|I)]^N}{dN}\right| < \left|\frac{d\left[1-\widetilde{H}(e_I|I)\right]^N}{dN}\right|$ holds.

## Appendix 2

The following statutes allow the use of smaller juries in both felony and misdemeanor trials. Arizona: Ariz. Const. art. I, § 23. Arizona permits the use of eight-person juries except in capital cases or cases with a possible sentence of 30 years or more. Connecticut: Conn. Const. art. I, § 19. Connecticut allows the use of six-person juries; in capital cases, a six-person jury is permitted only if the defendant consents. Florida: Fla. Stat. Ann. § 913.10 (West 1996). Florida mandates six-person juries for felony cases, except in capital cases, which require a twelve-person jury. Iowa: Iowa R.Cr.P. 2.67(6) the jury is limited to six jurors. Iowa requires twelve-person juries for "serious" or "aggravated" misdemeanors. Indiana: Ind. Code Ann. § 35-37-1-1 (West 1996). Indiana permits six-person juries for felony cases in municipal and county courts, but not at the Circuit or Superior court levels. Kansas: Kan. Stat. Ann. § 22-34-3403 (West). Kansas allows a jury of fewer than twelve persons in felony cases if both parties agree in writing before the verdict and the court approves. Louisiana: La. Const. art. I, § 17. Louisiana allows six-person juries but requires twelve-person juries when sentencing includes confinement and hard labor. Massachusetts: Mass. Ann. Laws ch. 234, §§ 25–26 (Law. Co-op. 1986). Massachusetts permits six-person juries at the Municipal and District court levels. New Jersey: N.J. Stat. Ann. § 2B:23–1(a) (West 1996). In New Jersey, both prosecutor and defendant may consent to a jury of fewer than twelve jurors, except in capital cases. North Dakota: Rule 23(b)(2)(A) of the North Dakota Rules of Criminal Procedure. Specifies that in misdemeanor cases, a jury consists of six qualified jurors unless the defendant demands a jury of twelve. Utah: Utah Const. art. I, § 10. Utah requires eight-person juries in all felony cases, including capital cases. Washington: Wash. Const. art. I, § 21, as applied by Wash. R. Ct. 48 (2012). Washington allows a jury of fewer than twelve members if the defendant opts for a six-member jury, except in capital cases. Wisconsin: Wis. Stat. Ann. § 972.02(2) (West 2012). In Wisconsin, both prosecutor and defendant may agree to a jury of fewer than twelve jurors.

**Acknowledgements** We are grateful to Paula Hannaford-Agor, Director of the Center for Jury Studies at the National Center for State Courts, and to Scott Dewey, Faculty Librarian at the University of Minnesota Law Library for their information and research on retrial rates and juries in U.S. jurisdictions. We are also indebted to Brianna Stringfield for her valuable comments and suggestions. Kelso Horne and Samuel Makikalli provided very valuable research assistance.

**Funding** Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. The authors received no financial support for the research, authorship, or publication of this article.

## References

Alpern, Steve, and Bo. Chen. 2017. The importance of voting order for jury decisions by sequential majority voting. *European Journal of Operational Research* 258 (3): 1072–1081. https://doi.org/10.1016/j.ejor.2016.09.053.

Arce, Ramón, Francisca Fariña, and Jorge Sobral. 1996. From Juror to Jury Decision Making: A Non-Model Approach. In *Psychology, Law, and Criminal Justice: International Developments in Research and Practice*, edited by Graham Davies, Sally Lloyd-Bostock, Mary McMurran, and Clare Wilson, 337–343. Berlin & New York: De Gruyter. https://doi.org/10.1515/9783110879483.337.

Ben-Yashar, Ruth, and Jacob Paroush. 2000. A nonasymptotic Condorcet jury theorem. *Social Choice and Welfare* 17 (2): 189–199. https://doi.org/10.1007/s003550050014.

Blackstone, William. 1769. *Commentaries on the Laws of England.* Oxford: Clarendon Press. (Reprint https://doi.org/10.7208/chicago/9780226161037.001.0001; original text: https://oll.libertyfund.org/titles/sharswood-commentaries-on-the-laws-of-england-in-four-books-vol-1).

Bornstein, Brian H., and Edie Greene. 2017. *The Jury under Fire: Myth, Controversy, and Reform*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190201340.001.0001.

Bureau of Justice Statistics. 2004. *State Court Organization 2004.* Washington, DC: U.S. Department of Justice. https://bjs.ojp.gov/content/pub/pdf/sco04.pdf.

De Caritat, M.-J.-A.-N. (Marquis de Condorcet). 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Paris: Imprimerie Royale. https://gallica.bnf.fr/ark:/12148/bpt6k417181.

Clermont, Kevin M., and Emily Sherwin. 2002. A comparative view of standards of proof. *American Journal of Comparative Law* 50 (2): 243–275. https://doi.org/10.2307/840821.

Coughlan, Peter J. 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *American Political Science Review* 94 (2): 375–393. https://doi.org/10.2307/2586018.

Demougin, Dominique, and Claude Fluet. 2005. Deterrence versus judicial error: A comparative view of standards of proof. *Journal of Institutional and Theoretical Economics* 161 (2): 193–206. https://doi.org/10.1628/0932456054193658.

Demougin, D., and C. Fluet. 2006. Preponderance of evidence. *European Economic Review* 50 (4): 963–976. https://doi.org/10.1016/j.euroecorev.2004.11.002.

Demougin, D., and C. Fluet. 2008. Rules of proof, courts, and incentives. *RAND Journal of Economics* 39 (1): 20–40. https://doi.org/10.1111/j.1756-2171.2008.00002.x.

Dharmapala, Dhammika, and Richard H. McAdams. 2003. The condorcet jury theorem and the expressive function of law: A theory of informative law. *American Law and Economics Review* 5 (1): 1–31. https://doi.org/10.1093/aler/5.1.1.

Feddersen, Timothy, and Wolfgang Pesendorfer. 1998. Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review* 92 (1): 23–35. https://doi.org/10.2307/2585926.

Feess, Eberhard, and Ansgar Wohlschlegel. 2009. "Why Higher Punishment May Reduce Deterrence." *Economics Letters*104 (2): 69–71. http://www.sciencedirect.com/science/article/pii/S0165%E2%80%911765(09)00121%E2%80%919

Findlater, Janet E. 1981. Retrial after a hung jury: The double jeopardy problem. *University of Pennsylvania Law Review* 129:701–737. https://doi.org/10.2307/3311743.

Flynn, Leo J.. 1977. "Does justice fail when the jury is deadlocked?" *Judicature* 61:129–134.

Franklin, Benjamin. 1785. *The Works of Benjamin Franklin*, vol. 9, 293. Letter to Benjamin Vaughan, March 14, 1785.

Gelman, Elijah N.. 2024. "Hung Out to Try: A Rule 29 Revision to Stop Hung Jury Retrials." *Northwestern University Law Review* 118:1139.

Glasser, Michael H.. 1996. "Letting the supermajority rule: nonunanimous jury verdicts in criminal trials." *Florida State University Law Review* 24:659–671.

Guarnaschelli, Serena, Richard D. McKelvey, and Thomas R. Palfrey. 2000. An experimental study of jury decision rules. *American Political Science Review* 94 (2): 407–423. https://doi.org/10.2307/2586020.

Guerra, Alice, Barbara Luppi, and Francesco Parisi. 2020. Accuracy of verdicts under different jury sizes and voting rules. *Supreme Court Economic Review* 28 (1): 221–236. https://doi.org/10.1086/709734.

Hannaford-Agor, Paula L.., Valerie P.. Hans, and G. Thomas. Munsterman. 1999. How much justice hangs in the balance? A new look at hung jury rates. *Judicature* 83 (2): 59–67.

Hannaford-Agor, Paula L., Valerie P. Hans, Nicole L. Mott, and G. Thomas Munsterman. 2002. *Are Hung Juries a Problem?* Williamsburg, VA: National Center for State Courts. https://www.ojp.gov/pdffiles1/nij/grants/199372.pdf.

Hans, Valerie P.., Paula L.. Hannaford-Agor, Nicole L.. Mott, and G. Thomas. Munsterman. 2003. The Hung Jury: The American Jury's Insights and Contemporary Understanding. *Criminal Law Bulletin* 39 (1): 33–50.

Helland, Eric, and Yaron Raviv. 2008. The optimal jury size when jury deliberation follows a random walk. *Public Choice* 134 (3–4): 255–262. https://doi.org/10.1007/s11127-007-9222-5.

Kagehiro, David K. 1990. Defining the standard of proof in jury instructions. *Psychological Science* 1 (3): 194–196. https://doi.org/10.1111/j.1467-9280.1990.tb00197.x.

Kagehiro, David K., and W. Clark Stanton. 1985. Legal vs. Quantified Definitions of Standards of Proof. *Law and Human Behavior* 9:159–174. https://doi.org/10.1007/BF01067049.

Kalven, Harry, and Hans Zeisel. 1966. *The American jury*. Boston: Little, Brown.

Kaniovski, Serguei, and Alexander Zaigraev. 2011. Optimal jury design for homogeneous juries with correlated votes. *Theory and Decision* 71 (4): 439–459. https://doi.org/10.1007/s11238-009-9170-2.

Kaplow, Louis. 2011. On the optimal burden of proof. *Journal of Political Economy* 119 (6): 1104–1140. https://doi.org/10.1086/664614.

Kaplow, Louis. 2012. "Burden of Proof." *Yale Law Journal* 121:738–859.

Kerr, Norbert L.., and Robert J.. MacCoun. 1979. Jury size: impact on verdict's correctness. *American Politics Quarterly* 7 (4): 438–452.

King, Kerry A., and Todd M. Nesbit. 2009. The empirical estimation of the cost-minimizing jury size and voting rule in civil trials. *Journal of Economic Behavior & Organization* 71 (2): 463–472. https://doi.org/10.1016/j.jebo.2009.02.015.

Klevorick, Alvin K., and Michael Rothschild. 1979. A model of the jury decision process. *Journal of Legal Studies* 8 (1): 141–164. https://doi.org/10.1086/467604.

Klevorick, Alvin K., Michael Rothschild, and Christopher Winship. 1984. Information processing and jury decisionmaking. *Journal of Public Economics* 23 (3): 245–278. https://doi.org/10.1016/0047-2727(84)90031-8.

Ladha, Krishna K. 1992. The condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science* 36 (3): 617–634. https://doi.org/10.2307/2111584.

Ladha, K. K. 1995. Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *Journal of Economic Behavior & Organization* 26 (3): 353–372. https://doi.org/10.1016/0167-2681(94)00068-P.

Lando, Henrik. 2009. Prevention of crime and the optimal standard of proof in criminal law. *Review of Law & Economics* 5 (1): 33–52. https://doi.org/10.2202/1555-5879.1237.

Luppi, Barbara, and Francesco Parisi. 2013. Jury size and the hung-jury paradox. *Journal of Legal Studies* 42 (2): 399–422. https://doi.org/10.1086/670692.

Miceli, Thomas J.. 1990. Optimal prosecution of defendants whose guilt is uncertain. *Journal of Law, Economics, and Organization* 6 (1): 189–201.

Mueller, Dennis C. 2003. *Public Choice III*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813771.

Mukhopadhaya, Kaushik. 2003. Jury size and the free rider problem. *Journal of Law, Economics, and Organization* 19 (1): 24–44. https://doi.org/10.1093/jleo/19.1.24.

Neilson, William S., and Harold Winter. 2000. Bias and the economics of jury selection. *International Review of Law and Economics* 20 (2): 223–250. https://doi.org/10.1016/S0144-8188(00)00028-4.

Neilson, W. S., and H. Winter. 2005. The Elimination of hung juries: retrials and nonunanimous verdicts. *International Review of Law and Economics* 25 (1): 1–19. https://doi.org/10.1016/j.irle.2005.03.002.

Ognedal, Tone. 2005. Should the standard of proof be lowered to reduce crime? *International Review of Law and Economics* 25 (1): 45–61. https://doi.org/10.1016/j.irle.2005.05.002.

Paroush, Jacob. 1997. Stay away from fair coins: A condorcet jury theorem. *Social Choice and Welfare* 15 (1): 15–20. https://doi.org/10.1007/s003550050088.

Pi, Daniel, Francesco Parisi, and Barbara Luppi. 2020. "Quantifying Reasonable Doubt." *Rutgers University Law Review*72: 455–508. https://rutgerslawreview.com/wp-content/uploads/2020/11/72_Rutgers_Univ_L_Rev_0455_Pi_Parisi_Luppi.pdf.

Rizzolli, Matteo, and Margherita Saraceno. 2013. Better that ten guilty persons escape: punishment costs explain the standard of evidence. *Public Choice* 155 (3–4): 395–411. https://doi.org/10.1007/s11127-011-9867-y.

Rubinfeld, Daniel L.., and David E. M.. Sappington. 1987. Efficient awards and standards of proof in judicial proceedings. *RAND Journal of Economics* 18 (2): 308–315.

Sowell, Thomas. 2011. Basic Economics: A Citizen's Guide to the Economy. Rev. and expanded ed. New York: Basic Books

Sparling, Zane. 2022. "Hundreds of Oregon Criminal Convictions Overturned Due to Nonunanimous Jury Verdicts, Supreme Court Decides." *The Oregonian/OregonLive,* December 31, 2022. https://www.oregonlive.com/crime/2022/12/hundreds-of-criminal-verdicts-now-invalid-after-oregon-supreme-court-rules-unanimous-jury-requirement-applies-to-old-cases.html.

Spencer, Bruce D. 2007. Estimating the accuracy of jury verdicts. *Journal of Empirical Legal Studies* 4 (2): 305–329. https://doi.org/10.1111/j.1740-1461.2007.00090.x.

Suzuki, Jun. 2012. "The Mathematics of Jury Size." *AIP Inside Science,* March 23, 2012. https://www.aip.org/inside-science/the-mathematics-of-jury-size.

Urken, Arnold B., and Stephen Traflet. 1983. Optimal Jury Design. *Jurimetrics Journal* 24:218–235.

Volokh, Alexander. 1997. N Guilty Men. *University of Pennsylvania Law Review* 146 (1): 173–216. https://doi.org/10.2307/3312649.

Watanabe, Takamitsu. 2020. A numerical study on efficient jury size. *Humanities and Social Sciences Communications* 7 (1): 93. https://doi.org/10.1057/s41599-020-00556-1.