ISSN No. 2454 - 1427

CDE
July 2024

Decisiveness, Correctness and Accuracy in Criminal Adjudication

Francesco Parisi
Email: parisi@umn.edu
University of Minnesota
and University of Bologna

Ram Singh
Email: ramsingh@econdse.org
Delhi School of Economics
University of Delhi

Working Paper No. 350

Centre for Development Economics
Delhi School of Economics
Delhi- 110007

Francesco Parisi[1] – Ram Singh[2]

Decisiveness, Correctness and Accuracy in Criminal Adjudication

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. Currently, states retain autonomy to determine the composition of their juries and to determine the Blackstonian ratios that their respective jurisdictions are to follow. In this paper, we expose the critical interdependence between these elements to assess how these variations in jury structure affect the accuracy and decisiveness of the jury process. We further show how the tradeoff between different combinations of jury size and burdens of proof is affected by the prosecutorial selectivity and the frequency with which hung-jury mistrials are brought up for a retrial.

Keywords: *jury size, standard of proof, Blackstonian ratios, mistrial, prosecutorial selectivity*

JEL Codes: *K0, K4*

---

1 Oppenheimer Wolff & Donnelly Professor of Law, University of Minnesota and Distinguished Professor of Public Economics, University of Bologna. Visiting Professor of Law, University of Miami. E-mail: parisi@umn.edu. 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.

2 Professor, Delhi School of Economics, University of Delhi. Email: ramsingh@econdse.org. The author would like to thank the Science and Engineering Research Board, Government of India, for research funding support (File Number: CRG/2021/007007). Institutional support by the Centre for Development Economic at Delhi School of Economics is appreciated.

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.[3] 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.[4] Additionally, many U.S. jurisdictions have reduced jury sizes over the years.[5] Great variations are also observed in the manner in which the "beyond a reasonable doubt" standard is implemented across U.S. jurisdictions.

In this paper, we consider the tradeoffs between jury size and standard of proof and evaluate their impact on the accuracy and decisiveness of the jury process. Our analysis shows how ideal jury size can vary depending on changes in the standard of proof. Furthermore, we show how this equilibrium is influenced by prosecutorial selectivity within the criminal justice system and by the frequency of retrials (i.e., the refiling of cases that were dismissed after a hung-jury mistrial). These insights offer a valuable framework for evaluating the impacts of institutional decisions on the criminal justice system across U.S. jurisdictions.

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. Section 3 presents a theoretical setup of jury design. We develop a formal model to consider the interdependent effects of jury size and standards of proof on the (i) correctness and (ii) decisiveness of the jury process. Section 4 extends the model to evaluate the tradeoffs under different conditions of the criminal justice system, focusing on the relevance of retrials. Section 5 concludes with a summary of the results observing how states' choices regarding jury structure may at times betray those states' proclaimed policy objectives. Finally, we offer suggestions for future research on this important topic.

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

[4] 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).

[5] 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 of Florida, 339 So.3d 506 (Fla. Dist. Ct. App. 2022) (rejecting the argument that the appellant was entitled to a 12-person jury when tried for capital sexual battery), *cert. denied*, 143 S. Ct. 1027 (2023).

2. U.S. Juries in a Flux

Historically, several changes have affected the way an American jury functions. In particular, these changes relate to three main characteristics of a trial by jury: (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.[6] Several subsequent Supreme Court decisions affirmed the constitutionality of juries with fewer than twelve members, imposing some limits.[7] Currently, approximately half of U.S. jurisdictions 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 2023 case *Pretell v. Florida*.[8] The U.S. Supreme Court did not grant certiorari to this case.[9]

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

---

[6] 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.

[7] 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 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.

[8] Pretell v. State, 339 So.3d 506 (Fla. Dist. Ct. App. 2022) (rejecting the argument that the appellant was entitled to a twelve-person jury when tried for capital sexual battery), *cert. denied*, 143 S. Ct. 1027 (2023).

[9] Given the turmoil associated with declaring non-unanimous juries unconstitutional and the resulting quest for retroactive application of unanimity, the denial of certiorari in *Pretell* is not surprising. This inconvenience aside, the Supreme Court will likely have to tackle a similar issue as that raised by *Pretell* in a future docket.

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).[10] 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.[11] During the last century, the use of non-unanimous juries has faced numerous legal and political challenges at both state and federal levels.[12] Ultimately, in 2020, the U.S. Supreme Court took on this important question and declared criminal convictions reached through non-unanimous verdicts to be unconstitutional. In *Ramos v. Louisiana*, the Supreme Court overruled its decision in *Apodaca*, establishing that non-unanimous verdicts for criminal felony convictions violate the Sixth Amendment's guarantee of a trial by jury.[13] 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.[14] Whether the Florida statutory amendment will face a challenge (either in state court or the U.S. Supreme Court) remains to be seen.[15]

[10] See Rule 31 of the Federal Rules of Criminal Procedure. See also the interesting discussion of non-unanimous convictions by Glasser (1996).

[11] 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). According to legal historians, the Louisiana reform stemmed from the increase in ethnic diversity in the composition of juries, and the fear that criminal outcomes could be driven by ethnic splits among jurors—minority jurors could veto a proposed verdict and single-handedly bring about a mistrial. See, for example, Hannaford-Agor et al.'s (2002) multi-phased NCSC research on mistrials, motivated by the concern that mistrials were occurring at unacceptably high frequencies in some jurisdictions. Likewise, upon admission to the Union in 1907, Oklahoma relaxed the unanimity requirement to allow non-unanimous convictions in misdemeanour cases. In 1934, Oregon became the third state to allow juries to convict with non-unanimous verdicts, having passed a constitutional amendment to that effect.

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

[13] 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 (2023). 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).

[14] 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).

[15] *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.

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 many criticisms, the "beyond a reasonable doubt" (hereinafter, BARD) standard continues to stand as a cornerstone of the American legal system.[16] 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.[17] 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 on States in implementing the BARD standard leaves room for heterogeneous interpretations of this standard, for courts and juries alike.[18] As recently shown in the empirical study carried out by Pi et al. (2019), behind the nominal uniformity of the BARD standard, this evidentiary pillar that protects defendants' rights is at risk of being implemented inconsistently across U.S. jurisdictions.[19]

3. A Theoretical Model

3.1 Related Literature

Prior contributions have focused on how large a jury should be (e.g., Paroush, 1997; Ben-Yashar and Paroush, 2000; Dharmapala and McAdams, 2003; Helland and Raviv, 2008), how juries should vote to convict (e.g., Klevorick and Rothschild, 1979; Klevorick et al., 1984; Ladha, 1995), and how to optimally set the standard of proof (e.g., Demougin and Fluet, 2005; Ognedal, 2005; Lando, 2009). Other scholars have explored the optimal combination of voting requirement and jury size (e.g., Urken and Traflet, 1983; Neilson and Winter, 2005; Luppi and Parisi, 2013; Guerra et al., 2020).[20] Studies

[16] 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 20th centuries.

[17] 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.

[18] In providing jury instructions, courts often refer to the Blackstonian ratio that is adopted in their jurisdiction. The choice between different Blackstonian ratios necessarily implies a jurisdiction's commitment to different standards of proof (Pi et al., 2019). 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 determining the choice of different BARD standards.

[19] Pi et al. (2019) researched the official proclamations of Blackstonian ratios across U.S. jurisdictions, unveiling substantial differences among states' BARD standards. 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. For an extensive literature review on Blackstonian ratio problems, see Rizzolli and Saraceno (2013).

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

covering similar terrain 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 the 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 these factors' critical interdependence in maximizing the correctness of jury verdicts and the accuracy of trial outcomes. Along with jury size and voting requirements, in this paper 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 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.[21] That said, if a quantitative measure of the standard of proof is deemed desirable for guiding jury decisions, determining which measure would be optimal 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 tradeoff 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 and (ii) decisiveness of the jury process to evaluate the tradeoffs 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, and how this balance is in turn affected by the level of prosecutorial selectivity and retrial rates.

[21] Case law 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) ("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.") *See also* Polinsky and Shavell (1989); Rizzolli and Stanca (2012) and Guerra et al. (2018).

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 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 Type-G defendants) or not committed the crime (innocent, or Type-I defendants). Denote the number of guilty defendants as G and the number of innocent defendants as I. Let $\pi(G)$ be the proportion of guilty individuals in the population of defendants, which will be taken as equivalent to the probability that a defendant is Type-G.

In a well-functioning criminal justice system, it is expected that the processes of screening by police investigators and case selection by prosecutors will serve as initial, impartial filters. These processes aim to ensure that only the strongest cases proceed to trial. Police investigators are responsible for conducting thorough and unbiased screenings of potential cases, while prosecutors carefully evaluate the gathered evidence and circumstances to determine which cases have sufficient merit to pursue. Consequently, cases with robust evidence and a higher likelihood of conviction are advanced, whereas weaker cases, lacking in substantial evidence or legal viability, are dismissed early on. This systematic approach helps maintain fairness and efficiency in the judicial process. Hereinafter, we shall refer to these combined characteristics of the criminal justice system as "prosecutorial selectivity". In an environment with good prosecutorial selectivity, we would expect a high proportion of Type-G and a smaller proportion of Type-I defendants in the population of defendants, $G \gg I$.[22]

During trial, prosecution and defense present evidence supporting their respective positions. Let, e denote the balance of incriminating evidence presented through the adversarial discovery from

[22] As it will be seen in our analysis, prosecutorial selectivity will play an important role in determining if the effects of changes in decisiveness, correctness, and accuracy affect primarily Type-G or Type-I defendants. This can be illustrated very simply by considering an ideal world of perfect prosecutorial selectivity, where no Type-I defendants are prosecuted. It is obvious that in such ideal world, errors by juries could only lead to false acquittals, and would never lead to wrongful convictions. Similarly, the effect of hung juries and lack of decisiveness would only reduce the accuracy of the criminal trials of Type-G defendants.

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 Type-G defendant can take on low values of e (e.g., the prosecution did not find much evidence to prove a Type-G defendant's guilt). On the other hand, e might take high values for Type-I defendants. Therefore, it is not possible to infer the defendant's guilt with certainty based on any given value of e.

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

Figure 1: *Timeline*

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 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, if the defendant is Type-G, then 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 $\mathrm{F}(e|I)$ be the distribution functions.

The assumption that evidence is more likely to be incriminating when the defendant committed the crime—i.e., a Type-G defendant generates stronger evidence (in expected terms) than a Type-I defendant—means that $F(e|G)$ first order stochastically dominates $\mathrm{F}(e|I)$, which is a standard assumption in criminal models (e.g., Rubinfeld and Sappington, 1987; Miceli, 1990; Miceli, 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.

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.[23] 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

$$s_i = e + \epsilon_i.$$

Each juror receives a different signal. A juror receiving a strong signal is more likely to believe that the defendant is guilty than a juror receiving a weak signal. For simplicity, assume for any given e, ϵ_i are independently and identically distributed such that $\epsilon_i \sim N(0, \sigma^2)$ for all $i \in \{1, \dots, N\}$.[24]

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

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, $H(.|e)$ is a decreasing function of e. See Figure 2 showing the probability of signals generated by evidence levels e_1 and e_2, where $e_2 > e_1$. $H(\bar{s}\,|e_i)$ is the probability that the signal value is less than or equal to a chosen level, say $\bar{s}$, for evidence level e_i.

[23] 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.

[24] Assuming a normal distribution for ϵ_i simplifies the formal analysis. Our results are qualitatively robust to other distributions.

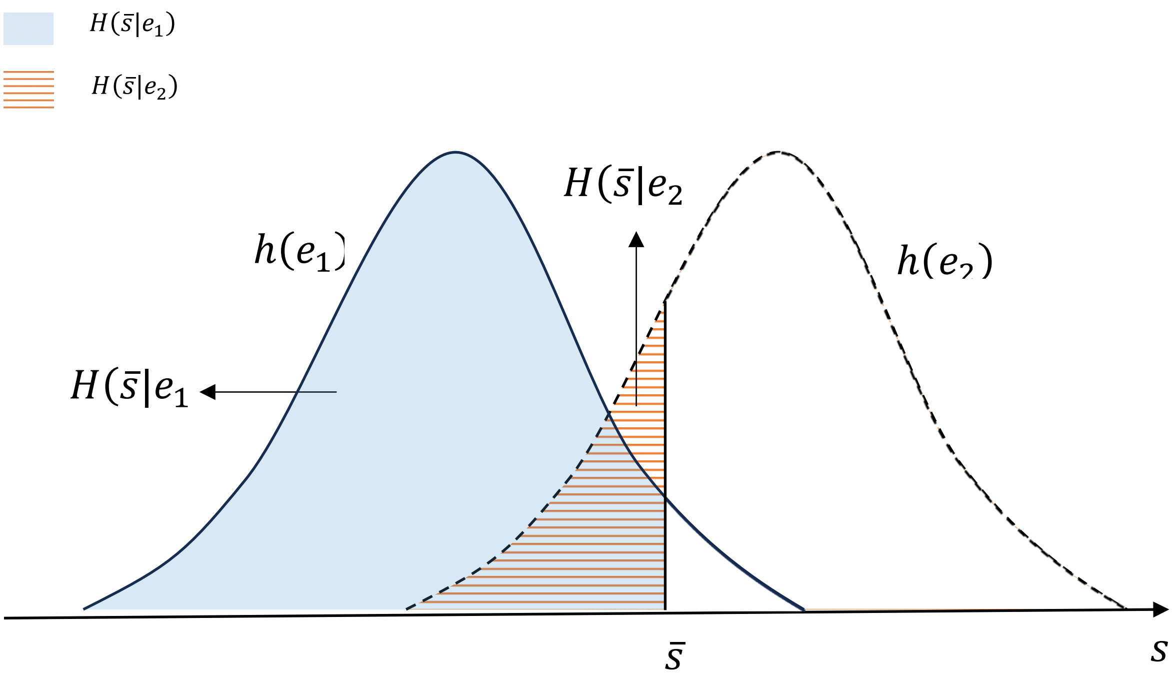

Figure 2: *Probability that the signal meets the BARD, conditional on strength of evidence*

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.[25] 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 $1 - H(e_I|e)$. Clearly, 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 $i \in \{1, \dots, N\}$. If $s_i \leq e_I$ for all $i \in \{1, \dots, N\}$, 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 not-guilty verdict

[25] 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).

decreases with jury size (N); and the probability of guilty verdict also decreases with jury size. So, decisiveness decreases with jury size (i.e., the probability of a mistrial increases with jury size).[26]

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 $t = 1$.[27]

Consider the case where the defendant did not commit the crime (i.e., at $t = 1$, the defendant is Type-I). 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,

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

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

Next consider the case where the defendant committed the crime (i.e., at $t = 1$, the defendant is Type-G). 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,

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

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

It can be seen that $\widetilde{H}(e_I|G)$ and $\widetilde{H}(e_I|I)$ are increasing in 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:

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

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

[26] See, however, how this result may change when free-riding and informational cascades in decision-making are introduced (Luppi and Parisi, 2013).
[27] For an extensive analysis on appeals and retrials, see Neilson and Winter (2005).

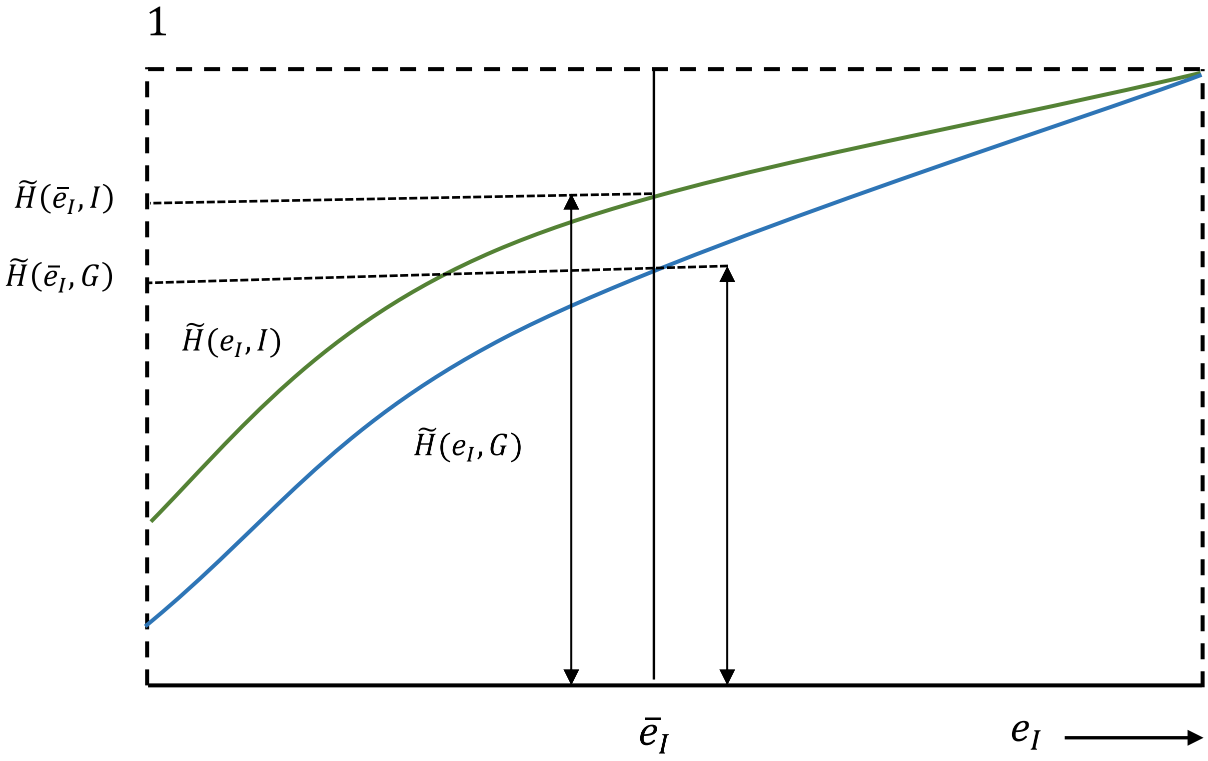

Figure 3: *Probability that signal meets the BARD, conditional on defendant's Type*

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

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

Jury size and the BARD standard, e_I, conjunctly determine the probability of a unanimous verdict. Moreover, we assume that $\widetilde{H}(e_I|G) < \widetilde{H}(e_I|I)$ and $1-\widetilde{H}(e_I|G) > 1-\widetilde{H}(e_I|I)$. Therefore, $[\widetilde{H}(e_I|G)]^N < [\widetilde{H}(e_I|I)]^N$ and $[1-\widetilde{H}(e_I|G)]^N > [1-\widetilde{H}(e_I|I)]^N$. In other words, even before evidence is presented at trial, at time $t=1$, for any given e_I, a Type-I defendant can be expected to be acquitted by the jury with a greater probability than a Type-G defendant. For the same reason, a Type-G defendant faces a higher probability of conviction.

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 Section 3.4 we will consider their effects on the correctness of the verdict). For analytical clarity, we will look at the effects on Type-I and Type-G defendants separately, and later consider the aggregate effects.

3.3.1 *Type-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 Type-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 a Type-I defendant, in this case) is $P(A|I) = (\widetilde{H}(e_I|I))^N$. The expected number of correct acquittals is $I \times [\widetilde{H}(e_I|I)]^N$. The corresponding probability of a (wrongful) conviction in this case is $P(C|I) = [1-\widetilde{H}(e_I|I)]^N$, and the expected number of wrongful convictions is $I \times [1-\widetilde{H}(e_I|I)]^N$.

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

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

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. This indeterminacy can be resolved by assuming that Type-I defendants are at least as likely to be acquitted as to be convicted—by assuming $P(A|I) \geq P(C|I)$ (or $\widetilde{H}(e_I|I) \geq [1-\widetilde{H}(e_I|I)]$, the probability of a unanimous verdict will be increasing in e_I. Now,

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

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

Since $N > 1$, it can be seen that $\left[\widetilde{H}(e_I|I)\right]^{N-1} - \left[1-\widetilde{H}(e_I|I)\right]^{N-1} > 0$. Also, $\frac{\partial\widetilde{H}(e_I|I)}{\partial e_I} > 0$, thus $\frac{\mathrm{d}\,\mathrm{P[U]}}{\mathrm{d}\,e_I} > 0$. In view of this result, it is clear that the expected number of unanimous verdicts also increases with e_I. In other words, as far as Type-I defendants are concerned, decisiveness increase with e_I. We can 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 (larger increase in correct acquittals and smaller decrease in wrongful convictions).

3.3.2 Type-G Defendants

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

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

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

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

As with Type-I defendants, the effect of the BARD standard on the decisiveness of the jury is, at first impression, indeterminate. However, in the context of Type-G defendants, the indeterminacy can be resolved if we assume that, overall, the criminal justice system is sufficiently functional to ensure that the probability of a correct conviction for a Type-G defendant is at least as large as the probability of a wrongful acquittal. Specifically, if we assume that $P(C|G) > P(A|G)$ (i.e., $\left[1-\widetilde{H}(e_I|G)\right]^N \geq \left[\widetilde{H}(e_I|G)\right]^N$), it will follow that:

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

Since $N > 1$ and $\left[1-\widetilde{H}(e_I|I)\right] > \widetilde{H}(e_I|I)$, we know that $\left[\widetilde{H}(e_I|G)\right]^{N-1} < \left[1-\widetilde{H}(e_I|G)\right]^{N-1}$. Thus:

$$\frac{\mathrm{d}\,\mathrm{P[U]}}{\mathrm{d}\,e_I} = N\left(\frac{\partial \widetilde{H}(e_I|I)}{\partial e_I}\right)\left(\left[\widetilde{H}(e_I|I)\right]^{N-1} - \left[1-\widetilde{H}(e_I|I)\right]^{N-1}\right) < 0 \qquad (2)$$

Therefore, an increase in the BARD standard will have a negative net effect on the decisiveness of the jury.

(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 has a negative net effect on the decisiveness of the jury: the total number of unanimous verdicts decreases with the BARD standard (larger decrease in correct convictions and smaller increase in wrongful acquittals).

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

Proposition 1: *For Type-I defendant, (i) an increase in jury size has a negative effect on decisiveness, and (ii) an increase in the BARD standard has a positive effect on decisiveness. For Type-G defendants, (i) increases in jury size, and (ii) increases in the BARD standard, have a negative effect on decisiveness.*

3.3.3 Aggregate Effects

In view of the above, an increase in jury size has a negative effect on the decisiveness of the jury, regardless of the types of defendants and their population shares. The overall effect of changes in jury size and BARD standards will depend on the proportion of Type-I to Type-G defendants in the criminal justice system. To see this, note that the total number of unanimous verdicts is given by:

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

Differentiating this with respect to e_I

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

From above, we can see that $\frac{\mathrm{d}\,\mathrm{P[U|I]}}{\mathrm{d}\,e_I} > 0$ and $\frac{\mathrm{d}\,\mathrm{P[U|G]}}{\mathrm{d}\,e_I} < 0$. That is, an increase in the BARD standard, e_I, increases decisiveness with respect to Type-I defendants but reduces decisiveness for Type-G defendants. The overall effect of e_I on decisiveness will thus depend on the proportion of the two types of defendants. Specifically, when $\pi(G) = \frac{G}{G+I} \approx 1$, i.e., $\frac{I}{G+I} \approx 0$, we get $I \times \frac{\mathrm{d}\,\mathrm{P[U|I]}}{\mathrm{d}\,e_I} + G \times \frac{\mathrm{d}\,\mathrm{P[U|G]}}{\mathrm{d}\,e_I} < 0$. On the other hand, when $\pi(G) \approx 1$, i.e., $\frac{I}{G+I} \approx 1$, we get $I \times \frac{\mathrm{d}\,\mathrm{P[U|I]}}{\mathrm{d}\,e_I} + G \times \frac{\mathrm{d}\,\mathrm{P[U|G]}}{\mathrm{d}\,e_I} > 0$.

In an environment with good prosecutorial selectivity, we would expect a high proportion of Type-G defendants in the population, with only a smaller number of Type-I defendants, $G \gg I$. With high levels of prosecutorial selectivity, jury decisiveness will decrease with the BARD standard. At the opposite end of the spectrum, when prosecutorial selectivity is very low, decisiveness can increase with the BARD standard—it will be easier to reach a unanimous not-guilty verdict for the many innocent defendants facing prosecution.

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

Proposition 2: *For the overall population of defendants, (i) an increase in jury size has a negative effect on decisiveness; and (ii) an increase in the BARD standard has a negative (positive) effect on decisiveness when prosecutorial selectivity is high (low).*

Proposition 2 shows the interdependence between prosecutorial selectivity (the diligent and unbiased activity of police departments and the effective screening of cases by the prosecutor's office) and decisiveness of the BARD standard.

3.4 Correctness: Jury Size vs. BARD Standard

We can 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 Type-I defendants and then considering Type-G defendants. We will later consider the aggregate effects.

3.4.1 Type-I Defendants

The effect of changes in jury size and BARD standards on the correctness of jury verdicts for Type-I defendants can be considered by looking at the probability of a correct verdict (acquitting Type-I defendant, in this case) at time $t = 1$, $P(A|I) = (\widetilde{H}(e_I|I))^N$ and at the resulting number of expected correct acquittals, $I \times \left[\widetilde{H}(e_I|I)\right]^N$.[28] The corresponding probability of a wrongful conviction in this case is $P(C|I) = \left[1 - \widetilde{H}(e_I|I)\right]^N$, and the expected number of wrongful convictions is $I \times \left[1 - \widetilde{H}(e_I|I)\right]^N$. $P(A|I)$ is increasing in the BARD standard e_I but decreasing in the jury size N, whereas $P(C|I)$ is decreasing in both.

(I) *Effects of Jury Size and BARD on Correctness*. 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. An increase in jury size,

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

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 is greater than the reduction in the number of wrongful convictions.*
Proof: See the Appendix.

The lemma suggests that the net effect of an increase in jury size is a decrease in correctness because the undesirable reduction in correct acquittals is more significant than the desirable reduction in wrongful convictions. From the above, it follows that, for Type-I defendants, (i) an increase in jury size has a negative effect on correctness; and (ii) an increase in the BARD standard has a positive effect on correctness.

3.4.2 Type-G Defendants

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

(I) *Effects of Jury Size and BARD on Correctness*. An increase in the BARD standard has two undesirable effects: (i) an increase in wrongful acquittals, and (ii) a decrease in correct convictions. This implies that an increase in the BARD standard leads to an overall decrease in correctness. In contrast, an increase in jury size has two countervailing effects on the correctness of jury verdicts: (i) a desirable decrease in wrongful acquittals, and (ii) an undesirable decrease in correct convictions.

3.4.3 Aggregate Effects

Therefore, to estimate the overall effect of an increase in jury size on correctness, we must compare the magnitude of the two countervailing effects. First, we prove the following result.

Lemma 2: *As the jury size increases, the fall in the number of correct convictions is greater than the fall in the wrongful acquittals, leading to a decrease in correctness.*
Proof: See the Appendix.

This lemma suggests that an increase in jury size leads to a decrease in correctness because correct convictions decrease more than wrongful acquittals, leading to a net decrease in correctness.

The following proposition sums up the results so far.

Proposition 3: *For Type-I defendants, (i) an increase in jury size has a negative effect on correctness; and (ii) an increase in the BARD standard has a positive effect on correctness. For Type-G defendants, increases in (i) jury size and/or (ii) BARD standards have a negative effect on correctness.*

Summing up, an increase in jury size reduces correctness because the (undesirable) decrease in correct verdicts is greater than the (desirable) decrease in incorrect verdicts. This is true regardless of the prosecutorial selectivity (i.e., the ratio of Type-G/Type-I defendants). Conversely, the overall effect of an increase in the BARD standard depends on the prosecutorial selectivity of the system. In view of Propositions 4 and 5, when prosecutorial selectivity is poor, an increase in BARD mitigates the problem, increasing the correctness of the jury verdict. On the other hand, when the prosecutorial selectivity is high, an increase in the BARD standard leads to a decrease in the correctness of the jury verdict.

Jury size and BARD standard are substitutes for Type-G trials. As far as Type-I defendants are concerned, jury size and BARD standards are not substitutes. They work differently vis-à-vis the correctness of jury verdicts. An increase in jury size reduces the chances of a correct acquittal for Type-I defendants, whereas an increase in BARD increases the probability of a unanimous verdict.

In light of Proposition 3, we can summarise the results for the overall population of defendants as:

Proposition 4*:* *For the overall population of defendants, an increase in jury size always has a negative effect on the correctness of jury verdicts. An increase in the BARD standard has a negative (or positive) effect on correctness when prosecutorial selectivity is high (low).*

Let us now step back and summarize the previously stated results on the effects of jury size and BARD standards on the decisiveness and correctness of the jury deliberation in Table 1 below. Table 1 illustrates that an increase in jury size reduces both jury decisiveness and correctness for Type-I and Type-G defendants alike. These decreases result regardless of whether prosecutorial selectivity is high or low. Meanwhile, increasing the BARD standard increases jury decisiveness and correctness with respect to Type-I defendants, but decreases these outcomes for Type-G defendants. The effects are variable when prosecutorial selectivity is high or low—decisiveness and correctness decrease in high prosecutorial selectivity contexts and increase in low prosecutorial selectivity contexts.

	Increase in Jury Size		**Increase in BARD**	
	Decisiveness of Deliberation	Correctness of Jury Verdicts	Decisiveness of Deliberation	Correctness of Jury Verdicts
Type-I Defendants	↓	↓	↑	↑
Type-G Defendants	↓	↓	↓	↓
Total (High/Low Selectivity)	↓/↓	↓/↓	↓/↑	↓/↑

Table 1: *Decisiveness and Correctness*

4. Hung Juries 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.[29] 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 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

[29] Courts have the discretion to declare a mistrial in other circumstances where the integrity of the legal process and 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).

criminal sanctions are imposed on the criminal defendant.[30] In yet 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.[31] 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.[32] The likelihood that a jury will hang twice in the same case is only 2.4% (Hannaford-Agor, 2002). This may suggest that most hung juries are caused by individual errors, rather than some inherent closeness in the evidence recognized across different groups.[33]

The incidence of hung juries and the determination of what happens after a mistrial is of great relevance for the understanding of the impact of jury size and BARD standards on the correctness and decisiveness of the jury process. In this section we will proceed to study how retrial decisions affect the results we derived in Section 3.

---

[30] 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 only in 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 to leave the case undecided through a hung jury, which has no corresponding possibility in a bench trial.

[31] 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).

[32] 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).

[33] 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.

4.1 Accuracy of Trial Outcomes Without Retrials

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.[34]

In this section, we shall consider the combined effects of correctness and decisiveness of juries on the accuracy of trial outcomes when hung-jury cases are never brought up for a retrial. In this case, hung juries grant de facto impunity to criminal defendants, and mistrials creates a wedge between the correctness of jury verdicts (when unanimity is reached) and the overall accuracy of trial outcomes. This point can be made clearer by noting that, as per Condorcet's jury theorem (de Caritat, 1785) an increase in jury size increases the correctness of a verdict.[35] Our results in Section 3 show that when a jury decides under a unanimity rule, the *correctness of jury verdicts* as well as the *decisiveness of the jury process* decrease with jury size. In this section we will focus on the combined effects of correctness and decisiveness on the *accuracy* of trial outcomes, showing that when a hung-jury mistrial is not followed by a retrial, Condorcet's jury theorem no longer holds: the *accuracy of trial outcomes* does not necessarily increase with jury size.

The same distinction between correctness of jury verdicts and accuracy of trial outcomes applies to the BARD standard. An increase in BARD standard always increases the correctness of (unanimous) verdicts, but its effect on the accuracy of trial outcomes is indeterminate and hinges upon the level of prosecutorial selectivity. We will proceed in our analysis, to consider the effect of mistrials in the limiting case where prosecutors do not file for a retrial. For analytical clarity, we will look at the effects on Type-I and Type-G defendants separately, and later consider the aggregate effects.

[34] 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, 2023).

[35] 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.

4.1.1 Type-I Defendants

When a mistrial effectively produces the results of an acquittal—i.e., no criminal sanctions imposed on the criminal defendant—the probability that a Type-I defendant avoids punishment becomes:

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

The complementary probability of a wrongful conviction of a Type-I defendant becomes $\left[1 - \widetilde{H}(e_I|I)\right]^N$. The probability of wrongful conviction, $P(C|I)$, is decreasing in jury size, N, as well as in the BARD standard, e_I. Meanwhile the probability of a correct non-conviction, $P(NC|I)$, 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, $P(C|I)$, and an increase in the probability of correct acquittal, $P(NC|I)$, of innocent defendants. Similarly, for any given jury size, an increase in the BARD standard, e_I, improves the accuracy of trial outcomes because $P(C|I)$ is decreasing in e_I and $P(NC|I)$ is increasing in e_I.

Alternatively, the accuracy of trial outcomes can be taken as the ratio of the number of correct acquittals over the number of wrongful convictions, i.e.:

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

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

(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 Type-I are concerned, both jury size and BARD standard increase the accuracy trial outcomes.

4.1.2 Type-G Defendants

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

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

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

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

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

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 can formulate the following Proposition.

Proposition 5: *When mistrials are not followed by a retrial, for Type-I defendants: (i) increases in jury size and (ii) increases in the BARD standard have positive effects on the accuracy of trial outcomes. For Type-G defendants: (i) increases in jury size and (ii) increases in the BARD standard have negative effects on the accuracy of trial outcomes.*

4.1.3 Aggregate Effects

Proposition 5 captures the familiar Blackstonian ratio tradeoff 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 prosecutorial selectivity, 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:

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

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 prosecutorial selectivity. With high prosecutorial selectivity, $\pi(G)\approx 1$, i.e., $\frac{I}{G+I}\approx 0$, the share of Type-G defendants is large and the probability of an accurate trial outcome decreases with jury size. On the other hand, when $\frac{I}{G+I}\approx 1$, the probability of an accurate trial outcome increases with jury size. An increase in jury size is therefore desirable to mitigate situations with low prosecutorial selectivity. In the intermediate range, jury size has an indeterminate effect on the number of accurate trial outcomes.

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 prosecutorial selectivity. An increase in the BARD standard is desirable to mitigate false conviction problems associated with low prosecutorial selectivity. Otherwise, an increase in the BARD standard has an indeterminate effect on the number of accurate outcomes.

Proposition 6: *When mistrials are not followed by a retrial, the probability of conviction for both types of defendants decreases with jury size and BARD standard. An increase in (i) jury size and/or (ii) BARD standard has a negative (positive) effect on the accuracy of trial outcomes when prosecutorial selectivity is high (low).*

4.2 Accuracy of Trial Outcomes With Retrials

The search for truth that drives criminal adjudication should ideally bring all cases to a final resolution, whether a verdict of guilty or not guilty. However, when the jury cannot reach a unanimous agreement with respect to a verdict, the judge declares a mistrial and the case ends with neither verdict. When a hung-jury mistrial is declared, the prosecution may choose to retry the case. If retried, the case may reach a resolution at the end of the second trial, with a final verdict. Under U.S. law, a defendant can be retried more than once, for an indefinite number of times, without violating the principle of double jeopardy—though this occurs infrequently.[36]

In the previous section, we considered the combined effects of jury correctness and decisiveness on the accuracy of trial outcomes for the limiting situation in which hung-jury cases are never brought up for a retrial. In that scenario, hung juries grant de facto impunity to criminal defendants. As a result, the probability of a mistrial creates a wedge between the correctness of jury verdicts (when unanimity is reached) and the overall accuracy of trial outcomes. In this section, we will consider the other limiting case, where a case ending in a hung-jury mistrial can be retried until a unanimous (guilty or not guilty) verdict is delivered. Under this regime, the decisiveness of the jury process is guaranteed. We can thus proceed with our analysis focusing on the effect of the BARD standards on the accuracy of trial outcomes. As in the previous sections, we will look at the effects on Type-I and Type-G defendants separately, and later consider the aggregate effects.

4.2.1 Type-I Defendants

As discussed in Section 3, the probability of a mistrial for a Type-I defendants is $1 - \widetilde{H}(e_I|I)^N - \left(1 - \widetilde{H}(e_I|I)\right)^N$. Suppose that the retrial follows the same judicial procedure as the original trial. Conditional on retrials, the probability of unanimous acquittal of Type-I defendant is $\left[\widetilde{H}(e_I|I)\right]^N$ and the

[36] The double jeopardy rule does not apply in cases of mistrial because, technically, the first trial never led to a final disposition. If a defendant is instead found not guilty in a criminal trial, the charges against him are dismissed and the double jeopardy rule would prohibit trying the acquitted individual again for the same crime.

probability of unanimous conviction is $\left[1-\widetilde{H}(e_I|I)\right]^N$. For instance, if there is only one retrial—i.e., only two trials in total—the probability of acquittal during trial for Type-I will be $\left[\widetilde{H}(e_I|I)\right]^N + \left[\widetilde{H}(e_I|I)\right]^N.\left[1-\widetilde{H}(e_I|I)^N - \left[1-\widetilde{H}(e_I|I)\right]^N\right]$.

With retrials continuing until a jury reaches a unanimous decision, the probability that a Type-I defendant will be acquitted is given by $P^{RT}(A|I)$—i.e., by the sum of the probabilities of acquittal at every trial round—which is

$$[\widetilde{H}(e_I|I)]^N(1+Z+Z^2+Z^3+\cdots,$$

where $Z=\left[1-\widetilde{H}(e_I|I)^N-\left[1-\widetilde{H}(e_I|I)\right]^N\right]<1$. Summing the terms in the series on the right-hand side, the probability that a Type-I defendant will be acquitted during the process is given by:

$$P^{RT}(A|I)=\frac{[\widetilde{H}(e_I|I)]^N}{[\widetilde{H}(e_I|I)]^N+\left[1-\widetilde{H}(e_I|I)\right]^N}.$$

Similarly, the probability that at some point during the process, a Type-I defendant will be convicted is

$$P^{RT}(C|I)=\frac{\left[1-\widetilde{H}(e_I|I)\right]^N}{[\widetilde{H}(e_I|I)]^N+\left[1-\widetilde{H}(e_I|I)\right]^N}.$$

Note that, due to the repeated retrial process, the final outcome is either an acquittal or a conviction, so we have $P^{RT}(A|I)+P^{RT}(C|I)=1$. Moreover, in this case there is no wedge between the correctness of jury verdicts (when unanimity is reached) and the overall accuracy of trial outcomes.

***Lemma 3**: For any given BARD standard, the probability of correct acquittal for a Type-I defendant increases with jury size, whereas the probability of wrongful conviction decreases with jury size. An increase in BARD standard increases the probability of a correct acquittal of a Type-I defendant and decreases the probability of a wrongful conviction.*

Proof: See the Appendix.

Formally, $\frac{\partial P^{RT}(A|I)}{\partial N}>0$ and $\frac{\partial P^{RT}(C|I)}{\partial N}<0$. Moreover, $\frac{\partial}{\partial e_I}(P^{RT}(A|I))>0$ and $\frac{\partial}{\partial e_I}(P^{RT}(C|I))<0$.

Summing up, when hung-jury cases are regularly retried, increases in (i) jury size and/or (ii) BARD standards have a positive effect on the accuracy of trial outcomes for Type-I defendant—the probability of a correct acquittal increases and probability of a wrongful conviction decreases.[37]

[37] It should be noted that, in the absence of indecisiveness, the wedge between the correctness of jury verdicts and the accuracy of trial outcomes identified in Section 4.1 disappears. Hence the results formulated in Proposition 10 with respect to accuracy of trial outcomes could be restated in terms of correctness of jury verdicts.

4.2.2 Type-G Defendants

Turning to Type-G defendants, recall that in every round the probability of acquittal is $\left[\widetilde{H}(e_I|G)\right]^N$. Also, note that a juror is more likely than not to find a Type-G defendant guilty. So, $\widetilde{H}(e_I|G) < \frac{1}{2}$. The probability that a Type-G defendant will be convicted at some round of repeated trials is:

$$P^{RT}(C|G) = \frac{[1 - \widetilde{H}(e_I|G)]^N}{\left[\widetilde{H}(e_I|G)\right]^N + [1 - \widetilde{H}(e_I|G)]^N}$$

The probability that a Type-G defendant will be acquitted at some round of repeated trials is:

$$P^{RT}(A|G) = \frac{[\widetilde{H}(e_I|G)]^N}{\left[\widetilde{H}(e_I|G)\right]^N + [1 - \widetilde{H}(e_I|G)]^N}$$

For a Type-G defendant, we get the following result.

Lemma 4: *For any given BARD standard, the probability that a Type-G defendant is incorrectly acquitted decreases with jury size, whereas the probability of a correct conviction increases with jury size. An increase in the BARD standard increases the probability of a wrong acquittal and decreases the probability of a correct conviction for a Type-G defendant.*
Proof: See the Appendix.

Formally $\frac{\partial P^{RT}(C|G)}{\partial N} > 0$ and $\frac{\partial P^{RT}(A|G)}{\partial N} < 0$. Moreover, it can be seen that $\frac{\partial}{\partial e_I}(P^{RT}(A|G)) > 0$ and $\frac{\partial}{\partial e_I}(P^{RT}(C|G)) < 0$.

Summing up the results of this subsection,

Proposition 7: *When hung-jury cases are regularly retried, increases in (i) jury size and/or (ii) BARD standards have a positive effect on the accuracy of trial outcomes for Type-I defendant—the probability of a correct acquittal increases and the probability of a wrongful conviction decreases. For Type-G defendant (i) jury size has a positive effect on the accuracy of trial outcomes; and (ii) BARD standards have a negative effect on the accuracy of trial outcomes.* [38]

4.2.3 Aggregate Effects

The previous Propositions show that the familiar Blackstonian ratio tradeoff between the desire to protect innocent defendants and the need to hold guilty defendants accountable disappears with respect to changes in jury size—*an increase in jury size improves the protection of Type-I and punishment rates*

[38] Also in this cases, as explained for Proposition 10, the results formulated with respect to accuracy of trial outcomes could be restated in terms of correctness of jury verdicts.

for Type-G, simultaneously. The tradeoff remains as far as the BARD standard is concerned. This means that, in the absence of the indecisiveness, jury size and BARD standards are not substitutes.

The number of accurate outcomes (the sum of correctly acquitted Type-I defendants and correctly convicted Type-G defendants) is given by

$$I \times P^{RT}(A|I) + G \times \ P^{RT}(C|G)$$

Since $\frac{\partial P^{RT}(A|I)}{\partial N} > 0$ and $\frac{\partial P^{RT}(C|G)}{\partial N} > 0$, the number of accurate trial outcomes increases with jury size. However, $\frac{\partial}{\partial e_I}(P^{RT}(A|I)) > 0$ but $\frac{\partial}{\partial e_I}(P^{RT}(C|G)) < 0$.

Therefore, jury size and BARD standards have opposing effects only with respect to accurate trial outcomes for Type-G defendants. The overall effect of the BARD standard on the accuracy of trial outcomes depends on the relative shares of the two types of defendants. When the share of Type-I is large, the BARD standard increases the accuracy of trial outcomes. When prosecutorial selectivity is high, an increase in the BARD standard decreases the accuracy of trial outcomes.

In light of the above, we can summarise the results as:

Proposition 8: *When hung-jury cases are regularly retried, (i) jury size has a positive effect on the accuracy of trial outcomes, and (ii) BARD standards have a negative (positive) effect on the accuracy of trial outcomes when prosecutorial selectivity is high (low).*

Let us now step back and summarize the previously stated results on the effects of jury size and BARD standards on the accuracy of trial outcomes in Table 2 below. A key distinction here is that, when retrials are available, an increase in jury size always increases the accuracy of trial outcomes. Additionally, for Type-I defendants, increasing jury size or increasing the BARD standard improves accuracy of trial outcomes regardless of whether retrials are available. Conversely, increasing jury size when mistrials are not followed by retrials or increasing the BARD standard in general will decrease accuracy of trial outcomes for Type-G defendants. Different levels of prosecutorial selectivity can also result in different levels of accuracy. When prosecutorial selectivity is high and retrials are not utilized, increasing jury size, or increasing the BARD standard in general will result in lower trial accuracy. Conversely, increasing jury size or BARD standards will increase trial accuracy in low selectivity jurisdictions.

		Accuracy of Trial Outcomes		
		Type-I Defendants	Type-G Defendants	Total (High/Low Selectivity)
Increase in Jury Size	W/Out Retrials	↑	↓	↓/↑
	With Retrials	↑	↑	↑
Increase in BARD	W/Out Retrials	↑	↓	↓/↑
	With Retrials	↑	↓	↓/↑

Table 2: *Accuracy of Trial Outcomes*

Our analysis suggests that, given a particular level of prosecutorial selectivity, maintaining the same frequency of accurate trial outcomes requires that an increase in the BARD standards be accompanied by a commensurate increase in jury size (to counterbalance the negative effect of BARD on accuracy). Alternatively, assuming constant prosecutorial selectivity across jurisdictions, states with higher BARD standards should also have large juries, and vice-versa. Absent such a correlation between jury sizes and BARD standards, we should expect different effective levels of trial accuracy across states.

5. Conclusions

The previous literature (Guerra et al., 2020) considered the tradeoffs between jury size and voting requirements under exogenously-set standards of proof. The authors found that large, non-unanimous juries or small, unanimous juries offered alternative ways of maximizing the correctness of jury verdicts while preserving the functionality of juries under exogenously given standards of proof. In 2020, the U.S. Supreme Court decision in *Ramos* reestablished the unanimity requirement for felony convictions, continuing to leave states with some freedom in choosing jury size and Blackstonian ratios. This consequently left some margin of discretion for states' implementation of the BARD standard of proof.

In this paper, we contribute an important building block to this analytical literature, showing how different combinations of jury size and BARD standards affect the decisiveness and correctness of the jury process, and how decisiveness and correctness jointly affect the expected accuracy of trial outcomes. Different jurisdictions' choice of jury size affects the optimal specification of the standard of proof. Blackstonian ratios and BARD standards should be gauged according to a jurisdiction's choice of jury size, with resulting quantitative specifications of the BARD standard that differ across jurisdictions adopting different jury designs. The opposite chain of causation may also be true—a jurisdiction's choice of a specific Blackstonian ratio or BARD standard might necessitate a rethinking of that jurisdiction's choice of jury size.

The existing literature conventionally ties the rationale behind using high standards of proof in criminal cases to the asymmetrical costs of Type-I and Type-II errors. In this paper, we shed light on the fact that standards of proof in criminal cases not only serve an evidentiary role for protecting innocent defendants, but also play an important role in conjunction with other institutional characteristics of the criminal process. For example, standards of proof foster correctness of trial outcomes when jurisdictions face challenges in prosecutorial selectivity.

The main normative lesson to be drawn from our exercise is that optimal jury design is greatly affected by the selectivity of the prosecutorial system. Virtually all the normative recommendations that we can formulate with respect to the desirability (or lack thereof) of larger juries or higher BARD standards hinge upon the overall reliability of police and the prosecutorial system. Naturally, changes in jury size and BARD standards are likely to yield several other effects that go beyond the scope of the present analysis. Future extensions may consider incorporating further factors into the analysis, such as effects on deterrence (Porat and Stein, 2001; Demougin and Fluet, 2005, 2008), litigation expenditures (Guerra et al., 2018), and discovery and evidence production (Miceli, 1990)). Further analysis should also consider the impact of changing jury size and BARD standards on prosecutorial strategies and on defendants' decisions to defend or make plea bargains (Demougin and Fluet, 2006).

Future theoretical research in this field should also extend our analysis to investigate how optimal jury design might change when considering correlated votes (Rubinfeld and Sappington, 1987), endogenous social values of adjudication errors (Miceli, 1990), heterogeneous juror competencies, signaling and behavioral cascades (Luppi and Parisi, 2013), and free-riding and strategic voting by jurors (Mukhopadhaya, K., 2003). For each of these extensions, our setup could serve as a useful building block for understanding more complex jury decision-making scenarios. Additionally, as shown in Pi et al. (2019), the choice of different Blackstonian ratios by U.S. jurisdictions indirectly implies a jurisdiction's commitment to different standards of proof. Empirical and comparative research should explore how a jurisdiction's choices regarding jury size has influenced its choices regarding Blackstonian ratios.

Along with these extensions, the results of this paper, and of much of the existing jury literature, could be tested experimentally using mock juries. Experimental findings could provide valuable insights on the effects of alternative jury structures and BARD standards in real-life conditions. Specifically, further empirical and experimental evidence should be gathered to shed light on the relevance of jury size and BARD standards on jury decisions.

Testing the ability of jurors to respond to changes to the required BARD standards could provide an important validation of much of the theoretical literature before any of these results are used for policy analysis or criminal justice reforms.

Appendix

Proof of Lemma 1: The probability of Type-I (innocent) defendants being accurately acquitted is $(\widetilde{H}(e_I|I))^N$. The number of Type-I defendants being accurately acquitted is $I \times (\widetilde{H}(e_I|I))^N$. The probability of Type-I defendants being wrongly convicted is $\left[1-\widetilde{H}(e_I|I)\right]^N$. The number of Type-I defendants being wrongfully convicted is $I \times \left[1-\widetilde{H}(e_I|I)\right]^N$. Clearly, 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 set up, $[\widetilde{H}(e_I|I)]^N > \left[1-\widetilde{H}(e_I|I)\right]^N$. Differentiating $[\widetilde{H}(e_I|I)]^N$ with respect to N we obtain $\frac{d[\widetilde{H}(e_I|I)]^N}{dN} = [\widetilde{H}(e_I|I)]^N ln\left(\widetilde{H}(e_I|I)\right)$. Similarly, we get $\frac{d\left[1-\widetilde{H}(e_I|I)\right]^N}{dN} = \left(1-\widetilde{H}(e_I|I)\right)^N ln\left(1-\widetilde{H}(e_I|I)\right)$. In view of assumptions that ½ $< \widetilde{H}(e_I|I) < 1$, we have $\left[1-\widetilde{H}(e_I|I)\right] < \widetilde{H}(e_I|I)$. Hence, $0 < \frac{[1-\widetilde{H}(e_I|I)]^N}{[\widetilde{H}(e_I|I)]^N} < 1$. Also, $\left|\frac{d[\widetilde{H}(e_I|I)]^N}{dN}\right| < \left|\frac{d\left[1-\widetilde{H}(e_I|I)\right]^N}{dN}\right| \Leftrightarrow \widetilde{H}(e_I|I)^N ln\left(\widetilde{H}(e_I|I)\right) > \left(1-\widetilde{H}(e_I|I)\right)^N ln\left(1-\widetilde{H}(e_I|I)\right) \Leftrightarrow \frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)} < \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N}$.

Note that since $0 < \left[1-\widetilde{H}(e_I|I)\right]^N < [\widetilde{H}(e_I|I)]^N < 1$, so $0 < \frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)} < 1$. Summing up, $\left|\frac{d[\widetilde{H}(e_I|I)]^N}{dN}\right| < \left|\frac{d\left[1-\widetilde{H}(e_I|I)\right]^N}{dN}\right|$ holds if and only if $\frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)} < \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N}$. Similarly, it can be seen that $\left|\frac{d[\widetilde{H}(e_I|I)]^N}{dN}\right| > \left|\frac{d\left[1-\widetilde{H}(e_I|I)\right]^N}{dN}\right|$ holds if and only if $\frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N} < \frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)}$.

Note that $\lim_{N\to\infty} \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N} = 0$ and $\lim_{N\to 0} \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{\widetilde{H}(e_I|I)^N} = 1$. Specifically, it can be seen that for $N \geq 2$, $\frac{ln\left(\widetilde{H}(e_I|I)\right)}{ln\left(1-\widetilde{H}(e_I|I)\right)} > \frac{\left(1-\widetilde{H}(e_I|I)\right)^N}{(\widetilde{H}(e_I|I))^N}$ holds for all $\frac{1}{2} < \left(\widetilde{H}(e_I|I)\right) < 1$.[39] That is, for the plausible

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

range of $\widetilde{H}(e_I|I)$ and N, we obtain $I \times \left|\frac{d[\widetilde{H}(e_I|I)]^N}{dN}\right| > I \times \left|\frac{d[1-\widetilde{H}(e_I|I)]^N}{dN}\right|$. 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 Type-G (guilty) defendant is wrongly acquitted is $(\widetilde{H}(e_I|G))^N$, where $0 < [(\widetilde{H}(e_I|G))]^N < \frac{1}{2}$. The number of Type-G defendants being wrongfully acquitted is $G \times [\widetilde{H}(e_I|G)]^N$. The probability of Type-G defendants being accurately convicted is $\left[1 - \widetilde{H}(e_I|G)\right]^N$. The number of Type-G defendants being accurately convicted is $G \times [1 - \widetilde{H}(e_I|G)]^N$. It can be seen that for a Type-G defendant, $\left|\frac{d[\widetilde{H}(e_I|G)]^N}{dN}\right| < \left|\frac{d[1-\widetilde{H}(e_I|G)]^N}{dN}\right| \Leftrightarrow \widetilde{H}(e_I|G)^N ln\left(\widetilde{H}(e_I|G)\right) > \left(1 - \widetilde{H}(e_I|G)\right)^N ln\left(1 - \widetilde{H}(e_I|G)\right) \Leftrightarrow \frac{ln\left(1-\widetilde{H}(e_I|G)\right)}{ln\left(\widetilde{H}(e_I|G)\right)} > \frac{(\widetilde{H}(e_I|G))^N}{\left(1-\widetilde{H}(e_I|G)\right)^N}$. And, $\left|\frac{d[\widetilde{H}(e_I|G)]^N}{dN}\right| > \left|\frac{d[1-\widetilde{H}(e_I|G)]^N}{dN}\right| \Leftrightarrow \frac{ln\left(1-\widetilde{H}(e_I|G)\right)}{ln\left(\widetilde{H}(e_I|G)\right)} < \frac{(\widetilde{H}(e_I|G))^N}{\left(1-\widetilde{H}(e_I|G)\right)^N}$. We find that for any $N \geq 2$ and $0 < \left(\widetilde{H}(e_I|G)\right) < \frac{1}{2}$, $\left|\frac{d[\widetilde{H}(e_I|G)]^N}{dN}\right| < \left|\frac{d[1-\widetilde{H}(e_I|G)]^N}{dN}\right|$ 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.

Proof of Lemmas 3: When retrials happen 'till the jury reaches a unanimous verdict, the probability that a Type-I defendant is correctly acquitted at some point is: $P^{RT}(A|I) = \frac{[\widetilde{H}(e_I|I)]^N}{[\widetilde{H}(e_I|I)]^N+[1-\widetilde{H}(e_I|I)]^N}$. The probability that a Type-I defendant is wrongly convicted at some point during the retrial is $P^{RT}(C|I) = \frac{[1-\widetilde{H}(e_I|I)]^N}{[\widetilde{H}(e_I|I)]^N+[1-\widetilde{H}(e_I|I)]^N}$. Differentiating $P^{RT}(A|I)$ with respect to N, we obtain $\frac{\partial}{\partial N}(P^{RT}(A|I)) = \frac{[\widetilde{H}(e_I|I)]^N[1-\widetilde{H}(e_I|I)]^N\left[\ln\left(\widetilde{H}(e_I|I)\right)-\ln\left(1-\widetilde{H}(e_I|I)\right)\right]}{\left([\widetilde{H}(e_I|I)]^N+[1-\widetilde{H}(e_I|I)]^N\right)^2}$. Since $\widetilde{H}(e_I|I) > \frac{1}{2}$, we have $\frac{\partial P^{RT}(A|I)}{\partial N} > 0$. Differentiating $P^{RT}(C|I)$ with respect to N, we obtain $\frac{\partial P^{RT}(C|I)}{\partial N} = \frac{[\widetilde{H}(e_I|I)]^N[1-\widetilde{H}(e_I|I)]^N\left[-\ln\left(\widetilde{H}(e_I|I)\right)+\ln\left(1-\widetilde{H}(e_I|I)\right)\right]}{\left([\widetilde{H}(e_I|I)]^N+[1-\widetilde{H}(e_I|I)]^N\right)^2}$. It can be seen that $\frac{\partial P^{RT}(C|I)}{\partial N} < 0$.

Next, differentiating $P^{RT}(A|I)$ with respect to e_I, we obtain $\frac{\partial}{\partial e_I}(P^{RT}(A|I)) = \frac{N[\tilde{H}(e_I|I)]^{N-1}\left[\left(1-\tilde{H}(e_I|I)\right)^{N-1}\right]\frac{d\tilde{H}(e_I|I)}{de_I}}{\left([\tilde{H}(e_I|I)]^N+[1-\tilde{H}(e_I|I)]^N\right)^2}$. Since $\frac{d\tilde{H}(e_I|I)}{de_I} > 0$, we get $\frac{\partial}{\partial e_I}(P^{RT}(A|I)) > 0$. Similarly, it can be seen that $\frac{\partial}{\partial e_I}(P^{RT}(C|I)) = \frac{-N[\tilde{H}(e_I|I)]^{N-1}\left[\left(1-\tilde{H}(e_I|I)\right)^{N-1}\right]\frac{d\tilde{H}(e_I|I)}{de_I}}{\left([\tilde{H}(e_I|I)]^N+[1-\tilde{H}(e_I|I)]^N\right)^2} < 0$. That is, $\frac{\partial}{\partial e_I}(P^{RT}(C|I)) < 0$.

Proof of Lemmas 4:

Given the expressions for $P^{RT}(A|G)$ and $P^{RT}(C|G)$, it can be seen that $\frac{\partial P^{RT}(C|G)}{\partial N} = \frac{[\tilde{H}(e_I|G)]^N[1-\tilde{H}(e_I|G)]^N[-\ln(\tilde{H}(e_I|G))+\ln(1-\tilde{H}(e_I|G))]}{\left[[\tilde{H}(e_I|G)]^N+[1-\tilde{H}(e_I|G)]^N\right]^2} > 0$.

Similarly, $\frac{\partial P^{RT}(A|G)}{\partial N} == \frac{[\tilde{H}(e_I|G)]^N[1-\tilde{H}(e_I|G)]^N[\ln(\tilde{H}(e_I|G))-\ln(1-\tilde{H}(e_I|G))]}{\left[[\tilde{H}(e_I|G)]^N+[1-\tilde{H}(e_I|G)]^N\right]^2} < 0$. So, as far as Type-G defendants are concerned, an increase in jury size improves accuracy on both counts: correct convictions increase and wrongful acquittals decrease. Next, differentiating $P^{RT}(A|G)$ and $P^{RT}(C|G)$ with respect to e_I, it can be seen that the expressions for $\frac{\partial}{\partial e_I}(P^{RT}(A|G))$ and $\frac{\partial}{\partial e_I}(P^{RT}(C|G))$ are similar to the expressions for $\frac{\partial}{\partial e_I}(P^{RT}(A|I))$ and $\frac{\partial}{\partial e_I}(P^{RT}(C|I))$. So, it is straightforward to see that $\frac{\partial}{\partial e_I}(P^{RT}(A|G)) > 0$ and $\frac{\partial}{\partial e_I}(P^{RT}(C|G)) < 0$.

References

Alpern, S. and B. Chen (2017). The importance of voting order for jury decisions by sequential majority voting. *European Journal of Operational Research 258*(3), 1072–1081.

Arce, R., F. Fariña, and J. Sobral (1996). From jurors to jury decision making. a non model approach. *Psychology, law and criminal justice. International developments in research and practice*, 337–343.

Ben-Yashar, R. and J. Paroush (2000). A nonasymptotic Condorcet jury theorem. *Social Choice and Welfare 17*(2), 189–199.

Blackstone, W. (1769). *Commentaries on the Laws of England*. Oxford: Clarendon Press.

Clermont, K. M. and E. Sherwin (2002). A comparative view of standards of proof. *The American Journal of Comparative Law*, 243–275.

Coughlan, P. J. (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *American Political Science Review 94*(02), 375–393.

De Caritat, M.J.. (1785). Essai sur l'application de l'analyse `a la probabilite des decisions rendues a la pluralite des voix [Essay on the Application of Analysis to the Probability of Majority Decisions] Paris: Imprimerie Royale.

Demougin, D. and C. Fluet (2005). Deterrence versus judicial error: A comparative view of standards of proof. *Journal of Institutional and Theoretical Economics 161*(2), 193–206.

Demougin, D. and C. Fluet (2006). Preponderance of evidence. *European Economic Review 50*(4), 963–976.

Demougin, D. and C. Fluet (2008). Rules of proof, courts, and incentives. *The RAND Journal of Economics 39*(1), 20–40.

Dharmapala, D. and R. 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.

Feddersen, T. and W. Pesendorfer (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political science review 92*(1), 23–35.

Feess, E. and A. Wohlschlegel (2009). Why higher punishment may reduce deterrence. *Economics Letters 104*(2), 69–71.

Findlater, J. (1981). Retrial after a Hung Jury: The Double Jeopardy Problem. University of Pennsylvania Law Review 129: 701.

Flynn, L. J. (1977). Does justice fail when the jury is deadlocked. *Judicature 61*, 129.

Gelman, E. (2023). "Hung out to Dry: A Rule 29 Revision to Stop Hung Jury Retrials," 118 Northwestern University Law Review 1139.

Glasser, M. H. (1996). Letting the supermajority rule: Nonunanimous jury verdicts in criminal trials. *Florida State University Law Review 24*, 659.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000). An experimental study of jury decision rules. *American Political science review 94*(2), 407–423.

Guerra, A., B. Luppi, and F. Parisi (2018). Standards of proof and civil litigation: A game-theoretic analysis. *The BE Journal of Theoretical Economics 19*(1), 20170005.

Guerra, A., B. Luppi, and F. Parisi (2020). Accuracy of verdicts under different jury sizes and voting rules. *Supreme Court Economic Review 28*(1), 221–236.

Hannaford-Agor, P. L., V. P. Hans, and G. T. Munsterman (1999). How much justice hangs in the balance-a new look at hung jury rates. *Judicature 83*, 59.

Hannaford-Agor, P. L., V. P. Hans, N. L. Mott, and G. T. Munsterman (2002). Are hung juries a problem. *Williamsburg, VA: National Center for State Courts*.

Hans, V. P., P. L. Hannaford-Agor, N. L. Mott, and G. T. Munsterman (2003). The hung jury: The american jury's insights and contemporary understanding. *Criminal Law Bulletin—Boston 39*(1), 33–50.

Helland, E. and Y. Raviv (2008). The optimal jury size when jury deliberation follows a random walk. *Public Choice 134*(3-4), 255–262.

Kalven, H. and H. Zeisel (1966). *The American Jury*. Boston: Little, Brown.

Kaniovski, S. and A. Zaigraev (2011). Optimal jury design for homogeneous juries with correlated votes. *Theory and decision 71*(4), 439–459.

Kaplow, L. (2011). On the optimal burden of proof. *Journal of Political Economy 119*(6), 1104–1140.

Kaplow, L. (2012). Burden of proof. *Yale Law Journal*, 738–859.

King, K. A. and T. 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.

Klevorick, A. K. and M. Rothschild (1979). A model of the jury decision process. *Journal of Legal Studies 8*(1), 141–164.

Klevorick, A. K., M. Rothschild, and C. Winship (1984). Information processing and jury decisionmaking. *Journal of Public Economics 23*(3), 245–278.

Ladha, K. K. (1992). The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, 617–634.

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.

Lando, H. (2009). Prevention of crime and the optimal standard of proof in criminal law. *Review of Law & Economics 5*(1), 33–52.

Luppi, B. and F. Parisi (2013). Jury size and the hung-jury paradox. *Journal of Legal Studies 42*(2), 399–422.

Miceli, T. J. (1990). Optimal prosecution of defendants whose guilt is uncertain. *Journal of Economic Behavior & Organization 6*, 189.

Miceli, T. J. (2009). Criminal procedure. In *Criminal law and economics— Encyclopedia of Law & Economics*, Volume 3. Cheltenham: Edward Elgar.

Mukhopadhaya, K. (2003). Jury size and the free rider problem. *Journal of Law, Economics, & Organization 19*(1), 24–44.

Neilson, W. S. and H. Winter (2000). Bias and the economics of jury selection. *International Review of Law and Economics 20*(2), 223–250.

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.

Ognedal, T. (2005). Should the standard of proof be lowered to reduce crime? *International Review of Law and Economics 25*(1), 45–61.

Paroush, J. (1997). Stay away from fair coins: A Condorcet jury theorem. *Social Choice and Welfare 15*(1), 15–20.

Pi, D., F. Parisi, and B. Luppi (2019). Quantifying reasonable doubt. *Rutgers University Law Review* 72: 455-508

Polinsky, A. M. and S. Shavell (1989). Legal error, litigation, and the incentive to obey the law. *Journal of Law, Economics, & Organization 5*(1), 99–108.

Porat, A. and A. Stein (2001). *Tort Liability Under Uncertainty*. Oxford University Press.

Rizzolli, M. and M. Saraceno (2013). Better that ten guilty persons escape: punishment costs explain the standard of evidence. *Public Choice 155*(3-4), 395–411.

Rizzolli, M. and L. Stanca (2012). Judicial errors and crime deterrence: theory and experimental evidence. *Journal of Law and Economics 55*(2), 311–338.

Rubinfeld, D. L. and D. E. Sappington (1987). Efficient awards and standards of proof in judicial proceedings. *The RAND Journal of Economics*, 308–315.

Sparling, Z. (2023). Hundreds of Oregon criminal convictions overturned due to nonunanimous jury verdicts, Supreme Court decides. The Oregonian/OregonLive. 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.

Urken, A. B. and S. Traflet (1983). Optimal jury design. *Jurimetrics 24*, 218.