Rethinking Guild, Juries, and Jeopardy

George C. Thomas III
Rutgers University School of Law-Newark

Barry S. Pollack
Chadbourne & Parke

Follow this and additional works at: https://repository.law.umich.edu/mlr
Part of the Criminal Law Commons, and the Law and Philosophy Commons

Recommended Citation
Available at: https://repository.law.umich.edu/mlr/vol91/iss1/2

This Article is brought to you for free and open access by the Michigan Law Review at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.
INTRODUCTION

Commentators have vigorously debated the fairness, as well as the ability, of jurors and juries.1 Dean Erwin Griswold once wrote: “The jury trial [in civil cases] is the apotheosis of the amateur. Why should anyone think that 12 persons brought in from the street, selected in various ways for their lack of general ability, should have any special capacity for deciding controversies between persons?”2 It is easier to defend a critical role for juries in criminal cases. As the Supreme Court has noted, juries provide “an inestimable safeguard against the corrupt or overzealous prosecutor and against the compliant, biased, or eccentric judge.”3 While this role seems self-evident, a difficult question lurks beneath the surface. What characteristics define a jury? One is, of course, that the jury must be impartial, a requirement derived from the Constitution itself.4 Beyond impartiality, however, one strains to identify the inalienable characteristics of a jury. Does the Constitution require a minimum number of jurors in a criminal case? A guilty vote from a minimum percentage of the total number of jurors to convict a defendant?

The answers to these questions must be affirmative. No one would consider a vote of 1-1 a jury verdict. As jury size increases, however, the questions become more difficult. The crux of the problem appears

---

1. See HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY 3-11 (1966) (providing broad outline of the debate over the use of the jury system).
2. Id. at 5 (quoting 1962-1963 HARVARD LAW SCHOOL DEAN'S REPORT 5-6).
in the Court's treatment of 9-3 and 5-1 verdicts. The Court has held the former, but not the latter, constitutional.\(^5\) Does any principle — other than the trite remark of Justice Rehnquist that "lines must be drawn somewhere"\(^6\)—justify finding a distinction of constitutional magnitude between 9-3 and 5-1? Why must a six-member jury reach a unanimous verdict when a seventy-five percent margin is constitutionally acceptable in a twelve-member jury?\(^7\) And what about verdicts of 8-4? 6-1?

To seek a nonarbitrary answer to these questions, we will follow the advice of Gertrude Stein, who told a young Ernest Hemingway: "Begin over again and concentrate."\(^8\) We "begin over again" by rethinking the concept of guilt. After settling on a plausible mechanism for evaluating jury verdicts of guilty or not guilty, we construct statistical models to compare the accuracy of various jury configurations. These models produce results consistent with the outcomes that the Court has reached by what appears to be a rather haphazard route.\(^9\)

We also argue, again based on our mechanism for evaluating verdicts, that the Supreme Court has failed to extend the logic behind its jury configuration cases to the acquittal side of the balance. If, as our models show and the Court has held, a 9-3 verdict in favor of the state is sufficiently certain to justify a verdict of guilty,\(^10\) a 9-3 verdict in favor of the defendant is also sufficiently certain to require an acquittal. In those jurisdictions that require jury unanimity,\(^11\) a jury vote of 9-3 for acquittal is a hung jury in form but, we believe, an acquittal in substance.\(^12\) If we are correct, a second trial for that offense would


\(^6\) Burch, 441 U.S. at 137.

\(^7\) The difference between 5-1 and 9-3 verdicts is the presence of six additional jurors who vote guilty by a 4-2 split. But why would adding six jurors who can manage only a 67% margin for guilt make the 9-3 verdict constitutionally superior to the 5-1 verdict? For an answer, see infra text accompanying note 107.

\(^8\) CARLOS BAKER, ERNEST HEMINGWAY: A LIFE STORY 87 (1969). Although Baker does not report that the remark offended Hemingway, Hemingway savaged Stein many years later in a posthumously published essay recounting his version of the conversation. See ERNEST HEMINGWAY, Miss Stein Instructs, in A MOVEABLE FEAST 11-21 (1964).

\(^9\) We do not mean to imply that the Court has never scrutinized statistical models. See Ballew v. Georgia, 435 U.S. 223, 231-36 (1978). Rather, we mean to say that the Court has never relied on a consistent principle to arrive at answers to these questions. Thus, we were surprised when we first performed our statistical tests and found that they tended to confirm the Court's jurisprudential outcomes.

\(^10\) The Court has held that 9-3 votes to convict possess sufficient accuracy and integrity to function as criminal verdicts. See Johnson v. Louisiana, 406 U.S. 356, 360 (1972).

\(^11\) Most states still require unanimity for criminal verdicts, as does the federal system. See FED. R. CRIM. P. 23(b), 31 (requiring 12-member juries as well as unanimity).

\(^12\) In essence, we argue for a concept of "de facto" acquittal. See Burks v. United States, 437 U.S. 1 (1978) (holding that an appellate reversal based on insufficient evidence is functionally
violate the Double Jeopardy Clause.

I. EVALUATING VERDICTS

The purpose of a mechanism for deciding guilt is to impose punishment only on those who deserve it. This goal raises two questions. First, who decides? Second, how can an observer evaluate whether the decisionmaker was right or wrong?

Although citizen juries played a role in the English criminal process at least as far back as 1201 A.D., their role initially was limited to screening cases, much like the present-day grand jury. Thus, a jury would decide whether to hold the accused for the guilt-determining phase, which might be battle or ordeal. Battle on an accusation of felony ended in death for the accused if he lost — the offender would perish either in the battle or from immediate hanging. Ordeal consisted principally of trial by fire and water, which required the accused (and sometimes the accuser) to walk through fire, carry red-hot iron, plunge his hand or arm into boiling water, or be thrown into a pool of water.

The premise underlying both battle and ordeal as guilt-determining mechanisms was that "God would always interpose miraculously to vindicate the guiltless." Thus, the innocent defendant who carried hot iron or cast his arm into boiling water would not show evidence of burns. A perverse twist on this theory explains the water ordeal. If the accused "floated therein, without any action of swim-
ming, it was deemed an evidence of his guilt; but if he sank he was acquitted.” 20 Here, apparently, the theory was that a person could only float without swimming through Satan’s intervention, and Satan would spare only guilty defendants.

With the dawning of the Renaissance, however, Western culture gradually ceased seeing the hand of God or Satan in every physical event, and the role of citizen juries was extended from screening defendants for trial to determining guilt. 21 As long as God was making the decision, believers did not accept the possibility of error. Once the ultimate decision rested in the hands of citizens, however, error became possible. The difficult, and usually unappreciated, question is how error can occur.

The most superficially plausible way for error to occur is to consider guilt as an empirical determination — the only judgment that the relevant facts could sustain. If an evaluator external to the jury22 knew the relevant facts, she could then judge whether the jury reached the right result. 23 But if the relevant facts can sustain only one verdict, how can the jury reach a contrary result? One answer is that the rules of evidence and procedure may have denied the jury some of the relevant facts, but this gets us nowhere. We seek to determine whether juries consistently reach the correct result within the parameters of the legal system, not whether the legal system perfectly identifies factors

20. 4 BLACKSTONE, supra note 15, at *343-44. Blackstone was critical of the rationale underlying the water ordeal. He wrote, “[o]ne can not but be astonished at the folly and impiety of pronouncing a man guilty, unless he was cleared by a miracle, and of expecting that all the powers of nature should be suspended by an immediate interposition of Providence to save the innocent, whenever it was presumptuously required.” Id. at *344.


22. Judges often act as the trier of fact in criminal cases, either because a defendant has waived the right to trial by jury or because the offense is “petty” and does not implicate that right. See Baldwin v. New York, 399 U.S. 66 (1970) (plurality) (holding that offenses punishable by six months or less do not require trial by jury); THE PRESIDENT'S COMMISSION ON LAW ENFORCEMENT AND ADMINISTRATION OF JUSTICE, THE CHALLENGE OF CRIME IN A FREE SOCIETY 262, Figure 5 (1967) (noting roughly twice as many bench trials for FBI index crimes as jury trials in 1965). Because our concern lies with jury configurations, however, we will limit our discussion to jury verdicts.

23. See Jonathan J. Koehler & Daniel N. Shaviro, Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods, 75 CORNELL L. REV. 247, 250 (1990) (noting that “lack of a truth criterion does not mean that verdicts are neither accurate nor inaccurate; it merely indicates limits to our knowledge about particular cases”). Appellate courts review convictions to determine whether they are based on sufficient evidence. See Tibbs v. Florida, 457 U.S. 31 (1982); Burks v. United States, 437 U.S. 1 (1978). This function resembles that of the external evaluator described in the text, though procedural rules that effectively create a presumption of correctness for trial verdicts make it simpler. See Jackson v. Virginia, 443 U.S. 307, 319 (1979) (phrasing “relevant question” for federal habeas review of convictions as “whether, after viewing the evidence in the light most favorable to the prosecution, any rational trier of fact could have found the essential elements of the crime beyond a reasonable doubt”).
indicating guilt. Thus, our external evaluator should know only what the jury knew.

Within that limitation, and assuming guilt is an empirical concept, a jury could make a mistake about guilt in the same way that a mathematician makes a mistake proving a theorem. That empirical questions have only one right answer does not mean every attempt to derive the answer is right. The difficulty, of course, is to establish a method capable of evaluating the attempt to derive the right answer.

Guilt as an empirical concept seems intuitively plausible. Charles Nesson has argued that public acceptance of a verdict and its behavioral message requires a belief that a verdict is a statement about what happened.24 For example, viewers of the videotaped beating of Rodney King may believe that they were in a position to conclude that the jury verdict was wrong. But practical and epistemological problems abound. Take a simple case where the only issue is whether defendant $Y$ performed act $X$. Since no way exists to go back in time and observe whether $Y$ did $X$, the most the legal system can do is gather evidence about what happened in the past and provide that evidence to a factfinder.25 In most cases, the external evaluator will not have access to a videotape of the relevant encounter and must rely on human testimony recalling distant events. This practical limitation typically requires judgments about witness credibility before the “fact” about $X$ is known, a task our system assigns to the jury.

More fundamentally, even if we permit the evaluator to judge credibility in determining the “fact” about $X$, no particular reason exists to prefer the evaluator's conclusion to that of the jury. Whoever seeks to evaluate the judgment of the jury stands in no higher relation to the “truth” than does the jury.26 Viewers of the Rodney King videotape could justifiably assert that they were surprised by the verdict in that case or that they would have voted differently, but to assert that the


25. In many cases, the issue is not so clear-cut as whether $Y$ did $X$. It might be, instead, $Y$'s state of mind when $Y$ did $X$ or the inferences that might be drawn from $Y$'s having done $X$ or whether $Y$ was justified in doing $X$. Or it may be that no witness saw $Y$ do $X$, and jurors are asked to make inferences from other facts to determine whether $Y$ did $X$. In cases where the issue is state of mind or the availability of defenses, even a time-traveling observer would be hard pressed to pronounce the jury's decision “right” or “wrong.” Of course, we could grant omniscience to the time-traveling observer, but that would return us to the pre-Renaissance days of deferring to the judgment of a deity.

26. When the state permits appellate review of the sufficiency of the evidence, see supra note 23, the state places appellate judges procedurally closer to the “truth.” But procedural relation to the truth is not the same as epistemological truth.
jury was wrong implies a privileged access to truth.\textsuperscript{27}

Thus, no math-like method for verifying a jury verdict exists. Any "truth" about guilt that exists in the universe is undiscoverable. The legal system can attempt to uncover its errors only by performing a post hoc review of its fallible human actors.\textsuperscript{28} We can avoid epistemological difficulty, however, by adopting a positivist view of guilt: guilt is what the system (the sovereign) says it is.\textsuperscript{29} Defendants are guilty if found guilty.\textsuperscript{30}

The narrow sense of this positivist conception holds that guilt or lack of guilt is simply the outcome of a complex process that limits the evidence that a jury can hear, requires proof beyond a reasonable doubt and a certain quality of representation of defendants, disqualifies some citizens from jury service, provides for the appeal of convictions (but not acquittals), and has a host of other rules that impede or facilitate the task of prosecuting individual cases. The final result of this elaborate process is, in a positivist sense, the reality about guilt.\textsuperscript{31}

While the narrow positivist conception of guilt is both practical and epistemologically sound, its narrowness renders it unsatisfactory as a framework for evaluating jury verdicts. Indeed, the compass is so

\textsuperscript{27} Some readers of an earlier draft insisted that viewers of the tape could justifiably conclude that the jury was wrong because it was biased. But we believe that to conclude the jury was biased presupposes that it was wrong. If it was right, why would a viewer infer bias? Moreover, basing a judgment about the jury decision on the videotape also presents a practical difficulty: the tape was not the only evidence presented at trial. See Morton I. Greenberg (Circuit Judge, Third Circuit Court of Appeals), Letter to the Editor, N.Y. TIMES, May 15, 1992, at A28 (noting that "those who have declared that the verdict was wrong have not judged the same case as did the jury"). The external evaluator discussed in the text must have access to all of the evidence that the jury considered. Even so, the epistemological difficulty remains.

\textsuperscript{28} See Barry S. Pollack, Note, Death Penalty Query: Is "Desert" Served?, 44 RUTGERS L. REV. (forthcoming 1992) ("Certainly our criminal justice system is filled with errors. Jurors can err in their findings of fact. Judges can err in their legal determinations and in the exercise of discretion. Witnesses can err in their recall. Lawyers can err in their strategy."). We think it too late in the day to return to the idea that an omniscient evaluator could make the ultimate judgment, as in the twelfth century battle and ordeal. See supra notes 14-21 and accompanying text.

\textsuperscript{29} See RICHARD O. LEMPERT & STEPHEN A. SALTZBURG, A MODERN APPROACH TO EVIDENCE 148 (2d ed. 1982):

The law is too practical to concern itself with the nice epistemological question of whether the truth can ever be known. It assumes that it can, and it assumes that the way to find truth is to present [to] the jury ... evidence which bears on the issue to be decided.

\textit{Id.} at 148.

\textsuperscript{30} At one level, this statement merely expresses a truism. A defendant is never guilty as a legal proposition in the absence of a verdict. See Thomas, supra note 12, at 835. Indeed, regardless of the quantity of evidence, the jury is free to "nullify" the evidence and vote not guilty. See Alan W. Schefflin, Jury Nullification: The Right to Say "No," 45 S. CAL. L. REV. 168 (1972). But we mean the statement in the text to offer a more fundamental, positivist proposition — that the system outcome is the truth about guilt in every sense of the term.

\textsuperscript{31} See Nesson, supra note 24, at 1367-68 (calling this a "more complex and sophisticated understanding of the meaning of a verdict" that entails "no judgment about the underlying factual event").
narrow that it enables no evaluation of the outcome, as opposed to the process. Once the highest court in the jurisdiction affirms a conviction, the defendant is guilty, and talking about whether the verdict is correct is incoherent.

Of course, we can use mathematical models to evaluate the process even under a narrow positivist conception of guilt. For example, suppose juries deliberated to pick a number between one and three. Then, if the defendant guessed the number, the verdict would be not guilty; otherwise the verdict would be guilty. The conviction rate would be 67%, similar to what our criminal justice system produces now. But this process would not produce the "truth" about guilt as the system currently contemplates that concept, in part because the current process is designed to produce a judgment about the strength of the admissible evidence of guilt.

We can adjust this "guess-game" method of reaching a verdict for the strength of the state’s case and thus illustrate another problem. Assume a trial is conducted under the current rules, nine of twelve jurors vote guilty, and the other three vote not guilty. In response to the 25% not guilty vote, the foreperson of the jury could select a number between one and four instead of one and three. The result is an acquittal if the defendant can guess the number. The defendant in this example has a 25% chance of being acquitted, a probability that reflects the fraction of jurors voting in favor of acquittal and thus the strength of the evidence presented. The difficulty with the evidence-adjusted "guess game," even under a positivist conception of guilt, is that the outcome ultimately depends on random probability. Guilt, whether viewed as a positivist or empirical notion, requires an individual judgment about an individual case.

Once an acceptable process is in place, however, the narrow positivist conception posits that the outcome of the process answers the question whether the defendant is guilty. If this is the best way to conceptualize guilt, then asking whether the jury verdict was substan-

32. See Kalven & Zeisel, supra note 1, at 58.
33. See Stuart S. Nagel & Marian Neef, Deductive Modeling to Determine an Optimum Jury Size and Fraction Required to Convict, 1975 WASH. U. L.Q. 933, 936 (If "in effect the fate of defendants in criminal cases would be decided by the flip of a coin, [the result] would be unconstitutional."). Other process aspects of the jury system are important in reaching a positivist judgment about guilt. Consider the role of jury deliberation. If deliberation were not important, the following method of reaching a verdict would presumably be acceptable. Members of the jury are secretly numbered before trial and a computer randomly generates a number in the range of the size of the jury. While all the jurors would hear the case, only the juror who corresponds to the generated number would decide the case. Thus, as in the example in the text, if nine out of twelve jurors thought the defendant was guilty, the defendant would have a 25% chance of being acquitted.
tively correct is incoherent. But this conception seems intuitively wrong— we do want to be able to critique the substance of the verdict in the Rodney King case rather than merely content ourselves with observing that the legal system works. Indeed, the concept of justice seems to entail an evaluation of the result, not just the process.34

Can guilt be conceived in a way that avoids the practical and epistemological problems of empirical guilt without accepting the ultimate verdict as inevitably correct? One answer lies in a broader version of positivism. Guilt might be what the sovereign, broadly construed, says it is. As both H.L.A. Hart and John Austin noted, the sovereign in a democracy is ultimately the electorate.35 A broader version of positivism, then, suggests that guilt is what the electorate would decide in any given case. Viewed in this light, a jury is simply a sample of the larger universe of the electorate, and a jury verdict can be compared to the result the electorate would have reached had it judged the case.

In a similar vein, Ronald J. Allen has observed, “The jury system is based upon a relativistic or communitarian theory of knowledge. Community consensus determines what is true for purposes of after the fact determinations.”36 This broad positivist conception of guilt comports with our societal preoccupation with sensational jury trials—for example, the trials of William Kennedy Smith, Mike Tyson, and the officers accused of beating Rodney King.37 One reading of this public attention is that society reserves for itself the ultimate decision on the defendant’s guilt,38 with the jury merely functioning as a convenient way of attempting to express the societal judgment. Thus, reporters asked many people concerning these three trials, “Did you agree with the verdict?” Implicit in this question is the notion that the

34. See, e.g., RONALD DWORKIN, LAW’S EMPIRE 177 (1986).
38. See, e.g., Murphy v. Florida, 421 U.S. 794, 799 (1975) (describing trial of defendant whose filmed confession had been telecast as a “hollow formality,” noting that “the real trial had occurred when tens of thousands of people, in a community of 150,000, had seen and heard the defendant admit his guilt before the cameras”) (referring to Rideau v. Louisiana, 373 U.S. 723 (1963)); Irvin v. Dowd, 366 U.S. 717, 727 (1961) (noting that almost 90% of 430 prospective jurors “entertained some opinion as to guilt — ranging in intensity from mere suspicion to absolute certainty”).
defendants' "true" guilt depends on society's view rather than on the actual jury decision. 39

Other reasons support the positivist argument that society 40 is the appropriate benchmark for evaluating whether a trial jury reached the "right" verdict. First, society bears the systemic consequences of jury verdicts — incorrect acquittals free guilty defendants while incorrect convictions reduce confidence in the process. The entity that bears the systemic consequences of trial jury decisions ought to provide the ultimate standard by which these decisions are evaluated. 41 Second, the society-as-jury concept parallels the Greek understanding of juries. Some Greek juries were as large as 1001 citizens, 42 a discernible percentage of the total citizenry. Today, of course, a jury of twelve is not a discernible percentage of even the smallest state's adult citizenry. Moreover, the Greek juries voted without deliberation or discussion, 43 suggesting that guilt was a societal determination and that the jurors were exercising society's right directly rather than as representatives of a much larger group.

Our working hypothesis is that "true" guilt means nothing more, or less, than the judgment that society as a whole would reach in a given case. We may, therefore, evaluate verdicts by considering the trial jury a sample of society. Obviously, for practical reasons, society's judgment cannot function in individual cases as a finding of guilt, but it can serve as a frame of reference for evaluating jury findings.

We recognize that using society as a mechanism for ascertaining "true" guilt is somewhat unsatisfying. Society is simply a larger (in-

39. The question "did you agree?" is also consistent with a view of guilt as an empirical fact ascertainable by an external evaluator; thus, the person to whom the question is addressed assumes the role of external evaluator. While this view is plausible, it entails the epistemological problems discussed in the text, supra notes 26-28 and accompanying text. Why should someone external to the jury — even a legal expert — be better able to perceive the empirical truth about guilt than the jurors present at the trial?

40. We intend society to mean the adult members of society and thus be coextensive with electorate.

41. Society can already demonstrate its interest in systemic consequences of the criminal process through politics. If a societal consensus arose that too many innocent persons were being convicted, legislation increasing the burden of demonstrating guilt would likely follow. Cf. Thomas, supra note 4 (discussing role of societal consensus in shaping judicial interpretations of law); In re Winship, 397 U.S. 358, 363-64 (1970) (noting value that society places on avoiding erroneous convictions). Similarly, if a consensus arose that too many guilty defendants were being acquitted, one could expect legislation permitting nonunanimous verdicts or otherwise easing the state's burden for conviction. If the consensus were powerful enough, a constitutional amendment limiting defendants' rights might result.

42. The number of jurors ranged from 201 to 1001, depending on the nature of the case. See MELVYN B. ZERMAN, BEYOND A REASONABLE DOUBT — INSIDE THE AMERICAN JURY SYSTEM 16 (1981). Zerman concludes that "[s]omething was lost when the size of juries dwindled from 1001 to 12." Id. at 18.

43. Id. at 17.
deed, the largest possible) jury panel. Why should reference to a larger panel produce the "truth" about guilt? Our only answer is that empirical guilt is unverifiable. Without recourse to the broad positivist conception of guilt as society's judgment, no standard exists for determining which jury verdicts are erroneous. Ultimately, we believe that deciding which jury configurations are likely to produce verdicts consistent with society's judgment is a more satisfying basis for jurisprudence than the anemic doctrine that "lines must be drawn somewhere." 44

We will briefly explore the Court's rationales for evaluating jury configurations and will then, by means of the laws of statistics, compare certain jury configuration verdicts with the verdict that the entire society would reach. Our goal is to provide statistical proof of which jury configurations promise to constitute reasonably reliable samples of society as a whole.

II. JURY CONFIGURATION RATIONALES

In a jury trial, the fate of an accused usually hinges on the factual, moral, and intuitive decisions of six to twelve people. 45 As noted earlier, the U.S. Supreme Court has struggled with rules governing jury size and with the requirement of unanimity or, alternatively, the minimal fractions necessary to convict. 46 Without recounting the struggle, the import of the Supreme Court cases on the issues of size and required majority is that jury votes of 9-3 and 6-0 satisfy the Constitution while jury votes of 5-0 and 5-1 do not. 47 One Justice has stated in dicta that a 7-5 conviction would be constitutionally deficient. 48 The

45. Jurors are not limited to factfinding. See Kalven & Zeisel, supra note 1, at 495: [T]he jury imports its values into the law not so much by open revolt in the teeth of the law and the facts, although in a minority of cases it does do this, as by what we termed the liberation hypothesis. The jury, in the guise of resolving doubts about the issues of fact, gives reign to its sense of values. It will not often be doing this consciously; as the equities of the case press, the jury may, as one judge put it, "hunt for doubts." Its war with the law is thus both modest and subtle. The upshot is that when the jury reaches a different conclusion from the judge on the same evidence, it does so not because it is a sloppy or inaccurate finder of facts, but because it gives recognition to values which fall outside the official rules. Id. at 495; see also United States v. Powell, 469 U.S. 57 (1984) (holding that inconsistent jury verdicts on multiple counts are beyond constitutional review, thereby giving juries the power to compromise when they are unsure whether the defendant is guilty); United States v. Dougherty, 473 F.2d 1113 (D.C. Cir. 1972) (discussing power of jury to disregard both uncontradicted evidence and the instructions of a judge in order to acquit).
46. See supra notes 5-7 and accompanying text.
48. See Johnson, 406 U.S. at 366 (Blackmun, J., concurring). Our models clearly vindicate
rationale for drawing these lines begins with the premise that the purpose of a jury trial is to act as a “safeguard against the corrupt or overzealous prosecutor and against the compliant, biased, or eccentric judge.” The Court found, quite reasonably, that a six-member jury could adequately serve this purpose.

But the Court was forced to seek a complementary rationale when, eight years later, it faced the issue of five-member juries in *Ballew v. Georgia*.

Why would five-member juries not also safeguard against prosecutorial and judicial failings? One reason might be the greater risk of reaching an erroneous verdict, but the Court principally relied on empirical data regarding jury dynamics. These studies suggest that decreasing the size of a jury diminishes the effectiveness of its deliberation — greater inconsistencies develop, the number of hung juries decreases, and minority viewpoints disappear from a majority of juries.

Despite *Ballew*'s focus on jury dynamics, we believe that risk of error lay at the heart of the Court's rationale. The Court's concern with deliberation dynamics, fewer hung juries, and minority viewpoints adds up to the principle that a jury should express views that are "truly representative of the community." Thus, the concern implicit in the Court's disapproval of small juries is that they will reach a verdict inconsistent with the verdict the community would reach. This understanding of *Ballew* meshes with our conclusion in Part I that jury configurations must provide reliable samples of society as a whole because society constitutes the only relevant standard for determining "true" guilt.

Varying the size and vote requirements of a jury obviously affects its ability to serve as a representative sample of society. The mathematical law of large numbers indicates that increasing sample sizes

---

Justice Blackmun's judgment on this point. *See infra* Table 3 (disclosing 29.1% risk of error for 7-5 vote as lower of two estimates).

49. *Williams*, 399 U.S. at 100 (quoting *Duncan v. Louisiana*, 391 U.S. 145, 156 (1968)).
50. *See* 399 U.S. at 103.
52. *See* 435 U.S. at 232-33. The Court did mention that, as jury size decreases, the risk of convicting an innocent defendant increases. 435 U.S. at 234. For sources of empirical data utilized by the Court, see 435 U.S. at 231-32 n.10.
53. For example, smaller juries produce more extreme compromises. 435 U.S. at 235.
54. 435 U.S. at 236-37.
55. *See* *Smith v. Texas*, 311 U.S. 128, 130 (1940); *see also* *Batson v. Kentucky*, 476 U.S. 79 (1986) (holding that prosecutors may not systematically discriminate by race during jury selection); *Taylor v. Louisiana*, 419 U.S. 522 (1975) (holding that jury venires must be drawn from a fair cross-section of the community).
simultaneously increases the accuracy of representation.\textsuperscript{56} For example, suppose one attempted to determine the average height of adults in a community where the true average is five feet, eight inches. For the scientific observer, the likelihood of making precise estimates increases as the sample size grows. The first individual randomly sampled might be six feet tall, the second person six feet, four inches. As more people are measured, however, the average will approach five feet, eight inches. Obviously, once the entire community is sampled, no error can exist, and the average of the sample will be five feet, eight inches.\textsuperscript{57}

When guilt is the issue, each subset of society sampled with respect to a particular case may differ, but a poll of the entire society will achieve the same result every time, thereby providing consistency.\textsuperscript{58} For example, if three jury panels hear the same evidence, one may vote guilty, one not guilty, and one may deadlock. But the vote of the entire society can produce only one result.

We believe that, to this point, we have made a tenable case that "true" guilt is only what society would determine and that the risk of error in making this determination explains the Court's concern with the size and vote requirements of the jury. Next, we will discuss the concept of "what society would determine." It is not as easy a concept as it first appears.

\section*{III. Controlling Percentages of Society}

We have argued that guilt is what society says it is, a judgment that the entire society would reach on the merits of an individual case. The next step is to determine what voting percentage of society should constitute its judgment. Because that percentage controls the meaning of guilt under our theory, we will refer to it as the \textit{controlling percentage}.

An infinite number of possible controlling percentages exists. We begin by rejecting unanimity. While unanimity might be appropriate, or at least practicable, for trial juries, it would effectively abolish guilt if required of the entire society. At the other end of the spectrum, any

\textsuperscript{56} This mathematical principle first appeared in Jakob Bernoulli's 1713 treatise, \textit{Ars Conjectandi}. See \textsc{Carl B. Boyer}, \textit{A History of Mathematics} 459 (1968). It holds true, as a statistical principle, only when the sample is drawn randomly.

\textsuperscript{57} Height sampling is offered here to enable the reader to envision the impact of the law of large numbers. Obviously, the height example does not pose the probabilistic/metaphysical problem that arises from guilt determinations.

\textsuperscript{58} Of course, as time passes the composition of society changes. Likewise, opinions of what constitutes "true" guilt in any given case could change. Nevertheless, society's judgment and the definition of "true" guilt are always the same at any given time.
percentage less than 50% seems insufficient to constitute a societal judgment of guilt. But perhaps one vote over 50% would be sufficient. We will refer to this theory as "majority control." Our democratic structure supports majority control as a guilt-defining theory. Historically, in democratic societies, at least a simple majority of jurors had to vote guilty in order to convict a defendant. Socrates was convicted by a vote of something like 280-221 — 56% to 44% — pursuant to the Greeks' simple majority definition of guilt.

A simple majority definition of guilt appears to raise troublesome questions. Assume a society of 200,000,000 adults who vote only on the merits of criminal cases; defendant A is deemed guilty because he receives 100,000,001 guilty votes while defendant B is deemed not guilty because she receives only 100,000,000 guilty votes. Are constitutional or general fairness principles violated when a difference of one vote out of 200,000,000 determines guilt? A due process argument suggests itself: if identically situated defendants are acquitted nearly as often as convicted when indistinguishably prosecuted before randomly selected juries, enforcement of any guilty verdict appears arbitrary.

The premise of this due process argument fails. An observer cannot know whether A and B are similarly situated until a judgment is reached about the concept of guilt. If majority control is the appropriate definition of guilt, then A and B are not similarly situated, and the one-vote societal majority therefore does not offend due process concerns.

The relevant question is whether any principle could justify defining guilt as a concurrence of greater than 50% of society. We are not concerned here with the risk of error that any small sample may not accurately represent the larger universe; this concern could obviously

60. See Zerman, supra note 42, at 17 (noting simple majority and secret ballot attributes of Greek system).
62. Our argument is premised on jurors' deciding a case only on the merits. In real cases, some jurors might be influenced by the defendant's race, ethnicity, appearance, and so forth. We mean to exclude all factors not relevant to the defendant's guilt by assuming that the jurors, whether 200,000,000 or six, consider only the merits of the case.
63. We assume here that a tie vote fails to establish guilt because it is one less than a simple majority.
justify requiring small sample verdicts to be greater than 50%. Instead, we are defining guilt by the judgment of society. Assume, for a moment, a system that premised guilt on a finding by a supermajority standard of 67%. Presumably, the rationale for preferring this to majority control would be to decrease the likelihood of convicting innocent people. But we cannot know that we are convicting innocent people until we have a definition of guilt. Thus, concern about convicting too many innocent people cannot be used to define guilt.

Perhaps the rationale can be restated: a supermajority controlling percentage would reduce the incidence of guilt and indulge a quasi-presumption in favor of innocence. While true, that principle does not seem susceptible to limitation. If 67% is a good way to reduce the incidence of guilt, then 75% is better and 90% is still better and, of course, 100% is the best of all.

Some may believe that a supermajority is necessary to effectuate the constitutional standard of "proof beyond a reasonable doubt." While this argument appears to disconnect the supermajority issue from the question of "too much guilt," it also runs into trouble. Requiring jurors to find defendants guilty beyond a reasonable doubt describes the reasoning process individual jurors should use; the evidentiary burden is theoretically unrelated to the question of what voting majority best defines guilt.

Moreover, the practical question of what voting majority would be equivalent to proof beyond a reasonable doubt is thorny. Why should 67% or even 75% be sufficient to constitute proof beyond a reasonable doubt? Giving effect to the reasonable doubt standard by requiring a supermajority vote of society would tend to make guilty verdicts rare events. Finally, the proof-beyond-a-reasonable-doubt argument ulti-

64. See infra Part IV.
65. See In re Winship, 397 U.S. 358 (1970) (requiring proof beyond a reasonable doubt in criminal cases as part of Fourteenth Amendment Due Process Clause).
66. The Court rejected the evidentiary burden argument in Johnson v. Louisiana, 406 U.S. 356 (1972), by approving a 9-3 guilty verdict despite the contention that three votes for acquittal "demonstrate that guilt was not in fact beyond a reasonable doubt." 406 U.S. at 362. The Court concluded "that verdicts rendered by nine out of 12 jurors are not automatically invalidated by the disagreement of the dissenting three." 406 U.S. at 363.
67. One reader of a draft of this paper argued that the existence of one million not guilty votes out of two hundred million total votes would cause him to conclude that a reasonable doubt existed. But the chance of obtaining a societal vote (on any issue) of 99.5% strikes us as extremely low. A certain percentage of society will, we believe, always act as contrarians; in deciding guilt, a few would credit testimony or defenses that most of us would reject while others would impose a standard of absolute certainty, rather than proof beyond a reasonable doubt, on the state. Moreover, our society-as-jury could not deliberate, further increasing the likelihood that votes approaching unanimity would be extremely rare. While we do not know what percentage required for guilt would make societal votes of guilty acceptably common, it would probably have to be sufficiently low to accommodate millions of not guilty votes.
mately does not succeed in disconnecting the rationale for a supermajority from the question of "too much guilt." The contention is premised on a concern with convicting only those whose guilt is manifest.

We thus have difficulty conceiving of guilt by any standard other than majority control. But one value of our statistical methodology is its applicability to any chosen controlling standard. The reader may disagree with our preference for majority control; if so, she may apply any percentage she wishes and come up with statistical judgments of the various jury configurations that are quite different from those produced by a majority control standard. We will next develop some jury models and use statistical principles to determine the risk of misrepresenting society posed by various permutations of jury dimensions. We will use both the majority control definition of guilt and an alternate controlling fraction of two thirds in these models. 68 Under the two thirds control model, society would have to vote guilty by two thirds or more before a defendant would be deemed guilty. 69

IV. JURY MODELS

Ethical principles forbid observing actual juries while they deliberate; 70 jury models thus play a major role in predicting and evaluating verdicts. 71 This section will construct two basic jury models, the "Close-Call-But-Innocent Defendant" Model and the "Sample Accuracy" Model. The figures and conclusions generated in this section permit comparisons between variations in jury sizes and in fractions required to convict.

68. In our jury models, both the majority control and two-thirds control theories refer to society, not actual jurors.

69. A reason for selecting two thirds as the supermajority standard is the significant role this fraction plays in the Constitution. See U.S. CONST. art. I, § 5, cl. 2 (requiring two-thirds vote to expel a member of each house); id. art. I, § 7, cl. 2 (requiring two-thirds vote to override a presidential veto); id. art. II, § 1, cl. 3 (requiring quorum of two thirds of the states for a presidential runoff); id. art. II, § 2, cl. 2 (requiring two-thirds vote of Senate to concur in treaties made by the President); id. art. V (requiring two-thirds vote in both houses or concurrence of two thirds of state legislatures to propose constitutional amendments); id. amend. XII (requiring two thirds of the states to form a quorum for conducting presidential elections); id. amend. XIV, § 3 (requiring two thirds of both houses to remove bar to political office implemented against former office holders who have been guilty of rebellion or insurrection); id. amend. XXV, § 4 (requiring two-thirds vote to bar the President's return to office after invocation of the disability provision in amendment).

70. While we suspect this comment needs no amplification or justification, consider that an observation must either be known or unknown to the jurors. If known, the fact of observation would potentially contaminate the deliberation. If it were unknown, the privacy of the jurors would be violated. Kalven and Zeisel avoided these problems by questioning jurors after they had reached a verdict. See KALVEN & ZEISEL, supra note 1, at 482.

71. For a discussion on requirements and objectives of jury models, see Nagel & Neef, supra note 33.
A. The Close-Call Defendant Model

In *Ballew v. Georgia*, the Supreme Court admonished that "[w]hen the case is close, and the guilt or innocence of the defendant is not readily apparent, a properly functioning jury system will insure evaluation by the sense of the community and will also tend to insure accurate factfinding." The *Ballew* Court accorded the "greatest value" of the right to a jury trial to these "close-call" prosecutions. A system that too frequently convicts close-call-but-innocent defendants is both inaccurate and, apparently under *Ballew*, unconstitutional.

By adopting a close-call model, we avoid the problem Stuart Nagel and Marian Neef created in the Independent-Mind Perspective model. Their model relied on data, collected by Harry Kalven and Hans Zeisel, from which the probability of an average juror's voting guilty or not guilty can be computed. Utilizing the Independent-Mind Perspective, however, twelve jurors will contemporaneously vote to acquit fewer than one in a trillion times, a prediction that is spectacularly wrong.

As Nagel and Neef recognized, the error lies in the twin assumptions underlying their model — that the strength of cases does not vary and that jurors act independently of each other. These premises contemplate that individual votes to acquit are randomly distributed across all juries, a distribution belied by common sense and the Kalven and Zeisel data. Consider a spectrum of cases from "strongest" to

---

73. 435 U.S. at 238.
74. 435 U.S. at 237-38. The *Ballew* Court cited numerous empirical studies in addition to a statistical model by Stuart Nagel and Marian Neef that we will discuss infra notes 75-81 and accompanying text.
75. Nagel and Neef calculated the probability that an average juror will vote guilty to be 96.4% by working backwards from the national average conviction rate of 64% as reported by Kalven and Zeisel. See Nagel & Neef, supra note 33, at 941; *Kalven & Zeisel*, supra note 1, at 58. If each juror will vote guilty 96.4% of the time and all jurors vote independently of each other, 12 jurors will vote guilty contemporaneously 64% of the time (96.4% raised to the twelfth power). The probability that an average juror will vote to acquit an average defendant would thus be 3.6% (100%-96.4%).
76. This is the chance of any one juror voting not guilty, 3.6%, raised to the twelfth power. See supra note 75.
77. The nationwide probability of jury acquittal, according to the Kalven and Zeisel data, is roughly 33%. See *Kalven & Zeisel*, supra note 1, at 58. This means that a trillion verdicts will contain roughly three hundred billion acquittals, rather than fewer than one, as would follow from the Nagel and Neef Independent-Mind Model.
78. See Nagel & Neef, supra note 33, at 942-43; see also Michael O. Finkelstein & Bruce Levin, Statistics for Lawyers 129-32 (1990) (questioning the accuracy of Nagel and Neef model).
79. See *Kalven & Zeisel*, supra note 1, at 58.
"weakest." "Strongest" cases are those that would require no deliberation to achieve a unanimous guilty verdict; "weakest" cases are those that would require no deliberation to achieve a unanimous acquittal. Given a sufficiently large sample of cases, this spectrum will contain many points between "strongest" and "weakest." As the points grow more distant from the "strongest" end, the cases will require more deliberation; at some point, no amount of deliberation will change enough jurors' minds to avoid a hung jury. As the points continue toward the "weakest" end of the spectrum, deliberation will eventually produce an acquittal, with less and less deliberation again required along the spectrum.

Although we do not know how many cases correspond to any point along the spectrum, not guilty votes obviously begin to occur at some distance from the "strongest" end, become unchangeable by deliberation at least by the hung jury juncture, and appear in larger numbers as the points approach the "weakest" end. Because the not guilty votes are clustered at the "weakest" end of the spectrum, acquittals will occur much more frequently than they would if the not guilty votes were distributed randomly across the entire spectrum. The spectrum metaphor also suggests that jury size and required voting majority critically affect only those cases that stand some distance from the "strongest" and "weakest" ends.

1. **Majority Control Theory**

A close-call-but-innocent defendant is one who would receive just enough votes of not guilty from society to avoid conviction. Under the majority control theory, society is equally divided over the guilt of close-call-but-innocent defendants. Since a 50% guilty vote falls just shy of the requisite controlling percentage for conviction, these defendants, by definition, should be acquitted.

We wish in this section to determine the risk that a jury vote in favor of conviction will result when society is evenly divided over guilt. In statistical terms, we will measure the probability of picking any particular jury vote in favor of guilt (for example, 9-3) from a universe that is split 50-50. Admittedly, the risk of error for these defendants will be higher than for the total universe of defendants,

---

80. If a jurisdiction permits a nonunanimous verdict, some not-guilty votes could become unchangeable prior to the hung jury point.

81. Nagel and Neef attempted to compensate for the cluster phenomenon by combining the Independent-Mind model with a Collective-Mind model. This modification did not, however, permit the construction of a formula for optimal jury size in jurisdictions allowing nonunanimous verdicts. See Nagel & Neef, supra note 33, at 968-69.
because the total universe contains many defendants who would receive more than a 50% societal vote for acquittal. Obviously, the risk of picking a 9-3 sample in favor of guilt from a universe that is only 50% in favor of acquittal is much higher than if the universe is 99% in favor of acquittal. We will address the question of risk of error for the total universe of defendants in the next section.

A statistical model capable of determining the risk of convicting close-call-but-innocent defendants resembles a simple coin flip. The probability that a balanced coin will land heads when flipped is 50%. Likewise, the probability of choosing a single guilty vote from an evenly divided society is 50%. If heads represents guilty and tails represents not guilty, each coin flip can equate to a randomly selected vote from society.

Given an evenly balanced coin, the probability in a sample of \( n \) flips of receiving \( g \) heads is displayed in Table 1. This probability equates to the likelihood that a close-call-but-innocent defendant, as defined by majority control theory, will receive \( g \) votes of guilty from a jury with \( n \) members. For example, the likelihood that a close-call-but-innocent defendant will receive nine guilty votes from a twelve-person jury appears in the column labeled 9 and the row labeled 12, a probability of 5%.

82. At least one other “coin-flipping” model exists in the literature. Saks and Ostrom weighted coins to represent various certainties of guilt held by jury members and then calculated the likelihood of reaching a verdict for several permutations of jury size and fraction required to render a verdict. Michael J. Saks & Thomas M. Ostrom, Jury Size and Consensus Requirements: The Laws of Probability v. the Laws of the Land, 1 J. CONTEMP. L. 163 (1975). For example, if all jurors are 80% certain of the defendant's guilt, then their decisionmaking can be represented by a coin that will land heads up 80% of the time. Although Saks and Ostrom's model is mathematically correct, it fails to discuss how often each of the various percentages of certainty actually occurs. This failure makes extrapolating the model's results into data for practical studies extremely difficult. See also David F. Walbert, Note, The Effect of Jury Size on the Probability of Conviction: An Evaluation of Williams v. Florida, 22 CASE W. RES. L. REV. 529 (1971) (finding that less certainty of guilt or innocence is needed for smaller juries to convict or acquit, respectively); cf. infra Table 3 (noting higher percentages of error as jury compositions grow smaller and move farther from unanimity).

83. The formula for calculating the probability of receiving \( g \) heads out of \( n \) flips of an equally weighted coin is \( \frac{\binom{n}{g} \cdot 0.5^n}{0.5^n} \). Virtually any introductory text on statistics discusses coin flipping probability. See, e.g., Y. LEON MAKSOUDIAN, PROBABILITY AND STATISTICS, WITH APPLICATIONS 23-30 (1969).

84. All of our models refer to postdeliberation jurors and, therefore, place no value on deliberation. While deliberation among the members of the jury panel might produce ideas that more accurately represent society, any attempt to quantify the effects of deliberation would be fruitless. By considering only postdeliberation jurors, we also ignore the effect of challenges for cause. These challenges eliminate a limited number of jurors who are predisposed to one side or the other. In the ideal system, all of society takes part in defining guilt; a defendant's sister would have a vote on the ideal jury. There, her overall effect would be minimal within the societal vote, but her presence on an actual jury would unfairly affect the outcome. Because challenges for cause should eliminate these unfair allotments of power, they should produce a jury more representative of society, thus justifying our decision to ignore any effect they might have on the risk of error.
TABLE 1

CHANCE OF CONVICTING CLOSE-CALL-BUT-INNOCENT DEFENDANT

MAJORITY CONTROL

(All figures rounded to the nearest percent)

<table>
<thead>
<tr>
<th># jurors (n)</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>16%</td>
<td>5%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>22%</td>
<td>11%</td>
<td>3%</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>16%</td>
<td>7%</td>
<td>2%</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>21%</td>
<td>12%</td>
<td>4%</td>
<td>1%</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>23%</td>
<td>16%</td>
<td>8%</td>
<td>3%</td>
<td>1%</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>19%</td>
<td>12%</td>
<td>5%</td>
<td>2%</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** = less than 1/2 of 1%

To compute the total risk of error in a system that permits nonunanimous verdicts requires summing of the risks of error for each possible verdict. As noted, a close-call-but-innocent defendant will receive a 9-3 vote of guilty 5% of the time. The chance that the same defendant will receive a 10-2 vote is 2%, while the probability of either an 11-1 or 12-0 vote is less than 1/2 of 1%. Adding these percentages, the probability that a close-call-but-innocent defendant will be convicted in a system that permits 9-3, 10-2, 11-1, and 12-0 guilty verdicts is 7% (5% + 2% + 0% + 0%).85 A system that permits 5-1 and 6-0 verdicts will convict 11% (9% + 2%) of close-call-but-innocent defendants.

The lowest risks of error, unsurprisingly, are associated with unanimous verdicts. Unanimous juries have a risk of error of less than 1/2 of 1% as long as they consist of more than seven members, and even a six-member unanimous jury has only a 2% chance of error. A jury consisting of twelve members that requires a vote of 11-1 or 12-0 also has a negligible chance of error.86

2. Two-Thirds Control Approach

Recall that the close-call-but-innocent defendant would receive just enough votes of acquittal to avoid conviction. While under the majority control theory the close-call defendant contemplates an evenly divided society, under the two-thirds control theory, one vote more than 33-1/3% of society in the defendant’s favor is a close-call-

---

85. We counted less than 1/2 of 1% as 0% when summing the probabilities.
86. By negligible, we mean less than 1/2 of 1%.
but-innocent defendant. The pertinent test for error again resembles a coin flip. In a two-thirds control system, however, the coin must be weighted in a manner that causes it to land heads up 66.6% of the time. Allowing heads to represent guilty and tails to represent not guilty, each coin flip equates to a randomly selected vote from a society 66.6% of whose members would vote for conviction and 33.4% for acquittal.

Given such a weighted coin, the probability in a sample of \( n \) flips of receiving \( g \) heads appears in Table 2. This is the probability that a close-call-but-innocent defendant, as defined by two-thirds control theory, will receive \( g \) votes of guilty from a jury with \( n \) members.

<table>
<thead>
<tr>
<th># jurors (n)</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>26%</td>
<td>8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>30%</td>
<td>20%</td>
<td>5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>27%</td>
<td>15%</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>23%</td>
<td>11%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>26%</td>
<td>19%</td>
<td>8%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>23%</td>
<td>15%</td>
<td>6%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>24%</td>
<td>21%</td>
<td>12%</td>
<td>4%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The risk of error with regard to the close-call-but-innocent defendant is significantly higher in the two-thirds control model than in the majority control model. As Table 2 demonstrates, a system permitting 9-3, 10-2, 11-1, and 12-0 verdicts has a total risk of error of 38% (21% + 12% + 4% + 1%). A system permitting 6-0 verdicts has an 8% chance of erroneously convicting a close-call defendant.

---

87. The coin should actually land heads up an infinitesimal amount less than 66-2/3% of the time. We have used 66.6% for convenience in calculation. Even if we could calculate this infinitesimal, it would not change our risk-of-error calculations.

88. The probability of receiving \( g \) heads out of \( n \) flips of a coin weighted to land heads up 66% of the time is \((.66^n)(.34^{n-g}) (n!/(g!)(n-g)!))\). See supra note 83.

89. This is because the 66-2/3% supermajority standard is a more difficult threshold to cross. Consider all convicted defendants who have 33-1/3% to 49.9% of society in their favor. Those guilty verdicts are correct under a majority standard but incorrect under the 66-2/3% supermajority standard. Because more verdicts are incorrect under supermajority, the risk of error appears larger in Table 2 than in Table 1.

90. Even a 12-member jury that requires unanimity will reach an erroneous verdict 1% of the time under the two-thirds control model.
Thus, the *Johnson* and *Williams* Courts implicitly adopted a policy consistent with majority control theory when they approved verdicts of 9-3 and 6-0. The risk of error for these jury verdicts under majority control theory is much less controversial (7% and 2%, respectively) than under two-thirds control (38% and 8%).

These statistics do not reflect what percentage of all defendants are close-call-but-innocent defendants. If this percentage is small, our coin-flip model has limited value, and the necessities of judicial economy could take priority over protecting the few close-call-but-innocent defendants. The Sample Accuracy Model, which follows, overcomes the problem of unknown distribution.

**B. The Sample Accuracy Model (SAM)**

The mathematically ideal method of testing jury configurations would be to poll society after every verdict. As a thought experiment, assume that society sits as a jury in a large number of criminal cases; the votes in each case will, in theory, constitute a distribution along a scale from 0% guilty to 100% guilty. Because only a very few defendants who are clearly innocent will be prosecuted, however, we will assume a distribution from 20% to 100%. We also assume in this section that a sufficiently large number of prosecutions will produce the same number of votes at each percentage from 20% to 100% — a uniform distribution of data points.

---

91. Data collected by Kalven and Zeisel indicate that the number of close-call defendants is fairly large; they found that judges considered 43% of criminal cases “close.” *Kalven & Zeisel*, supra note 1, at 134.

92. We drew the 20% figure from the procedure in ancient Greece that sanctioned prosecutors if less than 20% of the very large Greek juries voted guilty. *Zerman*, supra note 42, at 17. Data collected by Kalven and Zeisel confirm that only a small number of clearly innocent defendants are prosecuted. *Kalven & Zeisel*, supra note 1, at 134-35. While the 20% benchmark for “clearly innocent” is admittedly arbitrary, small changes in this figure — for example, using 15% or 25% — would not generate significant changes in the results our model produces. Similarly, no data points may exist at 100%, for the reasons offered in note 67, supra, but the use of 95% or 99% as the top of the range would not significantly change our results.

93. No data exist to confirm or reject this empirical assumption. We believe that prosecutors will probably charge more cases in the upper reaches of the range. We also believe that fewer such cases exist as the percentages of guilty votes near 100%. Thus, a normal distribution (bell curve) with a mean between 70%-80% might best describe the universe of cases prosecutors file. But we must consider the effect of plea bargaining. Again, lacking data, we surmise that the incentives to plea bargain are strongest in cases near the mean; both sides have something to gain by bargaining and something to lose by going to trial. The weaker the evidence, the less incentive defendants have to plea bargain; when the evidence is strong enough to convince percentages of society that approach 100%, prosecutors have little incentive to bargain. Plea bargaining, therefore, may flatten the curve of cases that go to trial to something approaching a uniform distribution.

Professor David Edelman, Department of Mathematics and Statistics, University of Sydney, assisted us in performing tests on distributions other than a uniform distribution. These tests demonstrated that other distributions had lower risks of error than SAM. For example, Professor Edelman generated results based on a normal distribution centered at approximately 60%
Because we are concerned for the moment only with the risk of erroneous convictions, consider that any jury panel conviction would be erroneous under the majority control theory if 50% or less of society would have voted guilty in a postverdict vote. We therefore want to know the likelihood that 50% or less of society would vote guilty when a particular jury configuration in favor of guilt occurs. Given our assumptions, we know two facts about our universe when we test a jury configuration: how often 50% or less of society votes guilty; and the jury configuration we are testing. From these facts, our Sample Accuracy Model (SAM) computes the probability that 50% or less of society would have voted guilty given a particular jury configuration.

We are not testing a particular vote in a particular case to determine whether it coincides with society's view. Rather, we want to estimate the likelihood that a particular configuration in favor of guilt, say 9-3, will be drawn from a universe that would have acquitted the defendant. This probability is the risk of error for systems that per-

and two beta-distributions with peaks at just below and just above 80%. The risks of error of a 9-3 verdict ranged from 1.0% to 4.5% in these models, while SAM indicates a 4.6% risk of error. The risks of error listed in Table 3, therefore, can be viewed as an upper range—the highest likely risk of error. The results of the other calculations are available from the authors.

We thank Professor David Edelman, Department of Mathematics and Statistics, University of Sydney, for his critique of an earlier model of SAM and for his verification of the model that appears in the text. We originally created a model that used the student t-test to estimate the probability of drawing a particular jury sample from a universe where 50% or less would
mit the configuration. For example, Table 3 displays the probability that 9-3 votes in favor of guilt are derived from a society with fifty percent or less of its members in favor of conviction — not-guilty defendants under the majority control theory. Likewise, Table 3 displays the probability that 9-3 verdicts in favor of guilt derive from a society less than two thirds of whose members favor conviction — not-guilty defendants under the two-thirds control theory.

SAM thus displays the probability that actual juries favor guilt by particular fractions although society as a whole would acquit (by each of the two standards we developed earlier).

have voted to convict. We did this by extrapolating the known mean and variance of a given sample (a given jury configuration) into the probable mean and variance of the sample's source universe (society). See generally MAXSOUDIAN, supra note 83, at 187-89. Since guilty and not guilty votes are not directly compatible with the numerically oriented student t-test, we converted votes into numbers by allowing a 1 to represent a vote of guilty and a 0 to represent a vote of not guilty. For example, conviction by a vote of 9-3 can be converted into a sample of 9 ones and 3 zeros. The mean of this sample is .75, and its variance is approximately .20. The student t-test produces the probability that such a sample would arise from a universe with a particular mean, e.g., .50 or less (indicating acquittal under the majority control theory) or .66 or less (indicating acquittal under the two-thirds control theory).

Although use of the student t-test when dealing with binomial distributions (for example, guilty-not guilty) suffers from some mathematical difficulties, see, e.g., David Edelman, Bounds for a Nonparametric T-Table, BIOMETRIKA, Apr. 1986, at 242, courts have accepted its use in some circumstances. See, e.g., Moultrie v. Martin, 690 F.2d 1078, 1082 (4th Cir. 1982). The results of this earlier SAM closely tracked those of the SAM we ultimately developed (results are available from the authors).
### TABLE 3

**Sample Accuracy Model (SAM)**

Risk That Particular Configurations for Guilt Result When Society Would Acquit

*(All figures rounded to nearest 1/10 percent)*

<table>
<thead>
<tr>
<th>Jury Size</th>
<th>Verdict</th>
<th>Majority</th>
<th>Two-thirds</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5-0</td>
<td>1.6%</td>
<td>8.7%</td>
</tr>
<tr>
<td>5</td>
<td>4-1</td>
<td>10.8%</td>
<td>35.0%</td>
</tr>
<tr>
<td>6</td>
<td>6-0</td>
<td>.8%</td>
<td>5.8%</td>
</tr>
<tr>
<td>6</td>
<td>5-1</td>
<td>6.2%</td>
<td>26.3%</td>
</tr>
<tr>
<td>7</td>
<td>7-0</td>
<td>.4%</td>
<td>3.9%</td>
</tr>
<tr>
<td>7</td>
<td>6-1</td>
<td>3.5%</td>
<td>19.5%</td>
</tr>
<tr>
<td>7</td>
<td>5-2</td>
<td>14.4%</td>
<td>46.7%</td>
</tr>
<tr>
<td>8</td>
<td>8-0</td>
<td>.2%</td>
<td>2.6%</td>
</tr>
<tr>
<td>8</td>
<td>7-1</td>
<td>2.0%</td>
<td>14.3%</td>
</tr>
<tr>
<td>8</td>
<td>6-2</td>
<td>9.0%</td>
<td>37.7%</td>
</tr>
<tr>
<td>9</td>
<td>9-0</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>9</td>
<td>8-1</td>
<td>1.1%</td>
<td>10.4%</td>
</tr>
<tr>
<td>9</td>
<td>7-2</td>
<td>5.5%</td>
<td>29.9%</td>
</tr>
<tr>
<td>9</td>
<td>6-3</td>
<td>17.2%</td>
<td>55.8%</td>
</tr>
<tr>
<td>10</td>
<td>10-0</td>
<td>**</td>
<td>1.1%</td>
</tr>
<tr>
<td>10</td>
<td>9-1</td>
<td>.6%</td>
<td>7.5%</td>
</tr>
<tr>
<td>10</td>
<td>8-2</td>
<td>3.3%</td>
<td>23.4%</td>
</tr>
<tr>
<td>10</td>
<td>7-3</td>
<td>11.1%</td>
<td>47.2%</td>
</tr>
<tr>
<td>11</td>
<td>11-0</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>11</td>
<td>10-1</td>
<td>.3%</td>
<td>5.4%</td>
</tr>
<tr>
<td>11</td>
<td>9-2</td>
<td>1.9%</td>
<td>18.1%</td>
</tr>
<tr>
<td>11</td>
<td>8-3</td>
<td>7.3%</td>
<td>39.3%</td>
</tr>
<tr>
<td>11</td>
<td>7-4</td>
<td>19.4%</td>
<td>63.1%</td>
</tr>
<tr>
<td>12</td>
<td>12-0</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>12</td>
<td>11-1</td>
<td>.2%</td>
<td>3.8%</td>
</tr>
<tr>
<td>12</td>
<td>10-2</td>
<td>1.1%</td>
<td>13.8%</td>
</tr>
<tr>
<td>12</td>
<td>9-3</td>
<td>4.6%</td>
<td>32.2%</td>
</tr>
<tr>
<td>12</td>
<td>8-4</td>
<td>13.4%</td>
<td>55.1%</td>
</tr>
<tr>
<td>12</td>
<td>7-5</td>
<td>29.1%</td>
<td>75.6%</td>
</tr>
</tbody>
</table>

** = less than 0.1%.

We believe that SAM reflects, with reasonable accuracy, the upper bounds of the risk of error in various verdicts. One reason SAM reflects an upper bound is that our assumption about a uniform distribution of data points between 20% and 100% may be too conservative. If the actual distribution is weighted toward the high end (under the assumption that prosecutors will be more likely to bring stronger cases to trial), the risk of error will generally be lower.\(^98\)

---

\(^{98}\) See *supra* note 93.
Moreover, the SAM risks of error for the nonunanimous verdicts apply to those verdicts alone and do not reflect the lesser risk of error for other permissible verdicts. While we summed the risks of error in Tables 1 and 2 to obtain the total risk of error for systems that permit nonunanimous verdicts, precisely the reverse mathematical relationship exists in the SAM risks of error. SAM takes all 9-3 verdicts and calculates what percentage of those verdicts would receive less than a majority vote from society. Whatever the chance of error in a 9-3 verdict of guilty, for example, the chance of error in a 10-2 guilty verdict must be lower. Because a system that permits 9-3 verdicts will also obtain convictions on 10-2, 11-1, and 12-0 verdicts, the total risk of error for all samples drawn from the universe will be a weighted average of all four risks of error and thus lower than the risk for a 9-3 verdict.

As we do not know the distribution of the various permissible configurations, we cannot calculate the total risk of error. All we can state with certainty is that a system that permits 9-3 verdicts will have a total risk of error, under majority control, of less than 4.6% and more than 0%. Thus, SAM's risk-of-error estimates are substantially lower than those of the coin-flip model. Allowing a 9-3 verdict entails a 7% risk in the coin-flip (majority) model and less than a 4.6% risk in SAM. Similarly, allowing a 5-1 verdict has a 11% risk in the coin-flip (majority) model and less than a 6.2% risk in SAM. SAM produces a lower risk of error because it considers all defendants, not just the close-call-but-innocent ones. Some defendants from the total universe of prosecuted defendants will be obviously not guilty, and the jury panel will be unlikely to make a mistake in those cases.

Regardless of the precision of the final numbers, SAM provides a useful comparative tool. While we do not claim that a system permitting 9-3 verdicts has a precise 4.6% risk of error under majority control theory, we can comfortably claim that the risk of error in a 9-3 system significantly exceeds that in a system that permits 8-2, 7-1, or 6-1 verdicts. This kind of comparison should prove useful. A legislature which already permits 9-3 verdicts might want to know whether

99. See supra notes 85, 89-90 and accompanying text.
100. The average will be weighted by the actual frequency of each configuration.
101. By contrast, the coin-flip model sums a series of binomial probabilities; each coin-flip exercise is independent of all others. Thus, the chance of flipping nine heads in twelve tries does not include the chance of flipping ten or eleven heads in twelve tries. In effect, the risks of error must be summed in the coin-flip model but averaged in SAM.
102. Our spectrum metaphor, see supra notes 80-81 and accompanying text, leads us to assume that a system would have more 9-3 verdicts than 10-2, 11-1, or 12-0 verdicts. If that is right, the total risk of error would be closer to 4.6% than to 0%.
economies could be achieved, without loss of accuracy, by permitting 6-1 verdicts. SAM discloses that while 6-1 produces a smaller risk of error than 9-3, 6-0 is far more accurate than either.

Social scientists typically require a risk of error of 5% or less to confirm a statistical hypothesis. In Table 3, the majority control model produces an upper bound risk of error of less than 5% for all nontraditional jury configurations that the Supreme Court has approved. Six to zero is 0.8%, 9-3 is 4.6%, 10-2 is 1.1%, and 11-1 is 0.2%. We argued earlier that the majority control standard made both practical and epistemological sense, and Table 3 demonstrates that it makes jurisprudential sense as well.

However, none of our statistical work requires acceptance of the Court's demarcations. Even readers who accept majority control as the appropriate standard might believe that the 4.6% upper bound for risk of error associated with 9-3 verdicts is too high. Consider a jurisdiction that permits 9-3 verdicts and conducts 1000 criminal trials in a year. Using Kalven and Zeisel's averages, 640 defendants will be convicted. If the verdicts are evenly distributed among 9-3, 10-2, 11-1, and 12-0, the average risk of error will be 1.5%, and ten innocent defendants will be convicted. As a policy matter, ten erroneous convictions may be too high a price to pay for whatever marginal efficiencies are achieved by permitting 9-3 verdicts. If the jurisdiction drew the line at 10-2 verdicts, only three innocent defendants would be convicted. Eleven to one would yield, on average, only one erroneous conviction per year.

Whatever view one takes on this policy question, Table 3 draws a reasonably clear line between the configurations that the Court has approved, those that it has rejected, and those with which states might experiment. For example, SAM provides an answer to the question posed in the Introduction, whether any principle could justify upholding 9-3 but not 5-1 verdicts. Although it may seem counterintui-

---

103. See Nagel & Neef, supra note 33, at 943 n.17 ("The 0.95 probability level [or 5% risk of error] is customarily used in social science to determine whether an hypothesis has been confirmed; in discussing that level, statisticians sometimes analogize to criminal case decision-making.").

104. The Court rejected 5-0 verdicts in Ballew v. Georgia, 435 U.S. 223 (1978). SAM generates a small risk of error for 5-0 verdicts (1.6%), but Ballew was, in large part, predicated on the dwindling opportunity for minority participation in smaller juries. See 435 U.S. at 241.

105. As noted with respect to the close-call model, the large risks of error associated with the two-thirds control theory suggest that the Supreme Court jurisprudence can be statistically validated only under majority control theory.

106. When we assume a distribution in the text, we assume a postdeliberation verdict, which is the same assumption underlying SAM.

107. See supra notes 5-7 and accompanying text.
tive, a 9-3 verdict has a significantly lower risk of error than 5-1 (4.6% versus 6.2%). SAM also leaves no doubt that the constitutional line should be drawn at 9-3 because an 8-4 verdict produces a clearly unacceptable risk of error of 13.4%.

Our analysis ignores the possibility that error may occur when the jury does not hear relevant, exculpatory evidence. Several factors, some of them benign, can cause exculpatory evidence not to be presented. When this occurs, the error rate will be higher than when computed as a function of how society would have voted had it heard a full evidentiary presentation. Thus, even though SAM discloses a risk of error of less than 0.1% for unanimous verdicts of nine or more jurors, we would expect the actual error rate to be higher when judged by the standard of the fully informed society. As our project is to compare various jury configurations, rather than compute the actual rate of error, we have assumed throughout that the jury and society hear the same evidence.

In sum, our findings provide support for the Supreme Court’s jury configuration jurisprudence. But they also raise a troubling and novel question that the Court has not addressed. If any state can accept a 9-3 vote for conviction as sufficiently accurate, are not all states required to give legal effect to a 9-3 vote in favor of acquittal? Presumably, the two votes would be equally accurate. In states that require unanimous verdicts, therefore, refusing to recognize a 9-3 vote of guilty may be a matter of grace, but votes of 9-3 or higher in favor of acquittal should be treated as an acquittal. The next Part examines this question.

V. De Facto Acquittal

The Supreme Court maintains that substance is more important than form in determining whether a defendant has been acquitted.

108. The factors include neglect, incompetence, and bad faith on the part of the police, the prosecutor, and the judge; defense counsel neglect and incompetence; witness bad faith, loss of memory, or inept performance; and the rigid application of rules of evidence.

109. Our analysis also ignores any state interest in judicial economy that impaneling smaller juries or avoiding hung juries might achieve. We believe, however, that judicial economy should be accorded no weight in the balance of interests unless the jury configuration has an acceptable risk of error. Cf. Burch v. Louisiana, 441 U.S. 130, 139 (1979) (rejecting a judicial economy argument in support of a nonunanimous six-person jury and commenting that “when a State [jury system] ... sufficiently threatens the constitutional principles that led to the establishment of the [jury] size threshold ... any countervailing interest of the State should yield”). Conceivably, any risk of error poses both benefits and detriments for defendants. A 9-3 verdict of guilty may arbitrarily convict a defendant who should be acquitted, while a 9-3 verdict of not guilty may acquit a defendant who should be convicted. While the risks might be balanced, erroneous convictions offend the integrity of our criminal justice system and should be avoided whenever practicable. Moreover, the benefit of the 9-3 verdicts is conferred only upon those persons who are, by definition, guilty.
The issue is not whether the judge or jury uttered the magic word *acquittal* but, instead, whether the outcome constitutes a resolution of the facts in the defendant's favor. 110 In upholding 9-3 verdicts, the Court has expressly held that they constitute a constitutionally adequate and accurate resolution of the facts. 111 Thus, a 9-3 verdict in favor of the defendant is, in substance, an acquittal. The question is whether states may refuse to recognize a 9-3 acquittal by adopting a requirement of unanimity for acquittals as well as convictions. While states are free to insist on a 12-0 standard for conviction, they may be constitutionally obligated to recognize a 9-3 verdict in favor of the defendant as a de facto acquittal.

This type of system would be procedurally asymmetrical. If a state requires a unanimous verdict to convict, procedural symmetry would require the same standard for acquittal. But a more powerful symmetry favors our de facto argument — the symmetry of substance rather than form. Our premise is that a 9-3 verdict has a sufficiently low risk of error that it accurately represents how society as a whole would have voted. If society would almost certainly have acquitted a 9-3 defendant, the symmetry we prefer is between what society would have done and what the legal system is required to do.

Calling a 9-3 verdict for acquittal anything other than an acquittal indulges the assumption that a state may recharacterize an adequate and accurate factual resolution as a nonresolution. If a state may thus use procedure to efface substance, might it not also characterize a 12-0 vote to acquit as a "failure to convict," thus permitting the reprosecution of all defendants who are not convicted on the first attempt? While this scenario seems bizarre, we are at a loss to see how it is distinguishable from a 9-3 favorable verdict, given our statistical confirmation of the Court's view that a 9-3 verdict is both adequate and accurate.

The response to the procedural symmetry argument is that the states that require unanimity may achieve symmetry by amending their rules of procedure to allow 9-3, 10-2, and 11-1 verdicts. Of course, symmetry is not always desirable. Proof beyond a reasonable doubt is an asymmetrical requirement in our system because a wrongful acquittal is preferable to a wrongful conviction. 112 In that light, states might want to require unanimity for conviction but permit 9-3

---

110. *See* United States *v.* Scott, 437 U.S. 82 (1978); *see also* Sanabria *v.* United States, 437 U.S. 54 (1978) (holding insufficient evidence dismissal to be an acquittal even though judge erred in suppressing some of government's evidence).


112. *See In re* Winship, 397 U.S. 358 (1970). This is not to say that the reasonable doubt standard requires certain jury votes. *See supra* note 66 and accompanying text. Nevertheless,
verdicts to establish an acquittal. In any event, this asymmetry would probably cost the system very little. If a 9-3 verdict in favor of acquittal represents the judgment of society, subject to a risk of error that is less than 4.6%, then the chance that a second jury would return a unanimous vote of conviction is quite small.

While recognizing 9-3 verdicts as sufficient to constitute an acquittal is thus a low-cost venture for the state, not recognizing these verdicts exacts a potentially exorbitant cost from individual defendants. Assuming additional deliberation does not persuade the three dissenting jurors to change their votes, the defendant who demonstrates his innocence within a 4.6% risk of error will receive a hung jury mistrial. The Supreme Court has consistently ruled that the Double Jeopardy Clause permits a second trial following a hung jury mistrial, and the 9-3 "de facto acquitted" defendant thus confronts precisely the dilemma the Court condemned in a slightly different context:

[T]he State with all its resources and power should not be allowed to make repeated attempts to convict an individual for an alleged offense, thereby subjecting him to embarrassment, expense and ordeal and compelling him to live in a continuing state of anxiety and insecurity, as well as enhancing the possibility that even though innocent he may be found guilty.

To be sure, the prosecutor may intuitively recognize what our models demonstrate — that the verdict represents what any other jury would likely do — and decline to prosecute. But two reasons counsel against relying on prosecutorial discretion. First, the prosecutor may be willing to take her chances, so even if the outcome is the same, the defendant has suffered the ordeal of the second trial. Second, the prosecutor may be able to improve her case and obtain a guilty verdict against a defendant who received a de facto acquittal from the first jury. Indeed, in Ashe v. Swenson, the State turned a unanimous acquittal into a unanimous guilty verdict. While this outcome (one not guilty verdict, one guilty verdict) raises questions about which jury verdict is the "true" one, the Supreme Court hinted in Ashe that the reasonable doubt standard does support the use of asymmetrical procedures that benefit the defendant.


114. Green v. United States, 355 U.S. 184, 187-88 (1957). Green held that conviction of a lesser-included offense constitutes an implied acquittal of the greater offense so long as the jury could have convicted on the greater. Although Green is thus technically distinguishable, the distinction disappears if a 9-3 verdict for acquittal is recognized as an acquittal in substance.


prosecutor may have manipulated the second prosecution in a way inconsistent with the "true" facts.\textsuperscript{117}

Regardless of which set of facts is "true," the Double Jeopardy Clause exists to prevent a second trial following an acquittal.\textsuperscript{118} Thus, we conclude that a defendant who receives nine not guilty votes must be treated for all purposes as acquitted, since the jury vote reflects a sufficient likelihood that society would consider her not guilty.\textsuperscript{119} There remains but one problem — identifying defendants who receive a 9-3 vote for acquittal. State law could, of course, specify that a 9-3 vote for acquittal is an acquittal. But in states that require unanimity, a jury would likely report a hung jury without disclosing the vote. A simple solution would be to poll the jury once a judge concluded that the jury was hopelessly deadlocked. In effect, recognizing that a 9-3 vote for acquittal is a constitutionally mandated acquittal creates an ancillary right to have a hung jury polled.

Support for the right to poll the jury can be found in \textit{Burks v. United States.}\textsuperscript{120} The Court held in \textit{Burks} that a conviction found by an appellate court to rest on insufficient evidence must be treated as an acquittal. We believe that a 9-3 jury split in favor of the defendant is a finding of insufficient evidence because our models show that this verdict almost certainly represents what a majority of society would believe if faced with the same proof. If a finding of insufficient evidence can thus be lurking in a hung jury, the Due Process Clause should entail the right to poll the jury as a way of preventing potential Double Jeopardy Clause violations.\textsuperscript{121}

\begin{itemize}
\item \textsuperscript{117} The Court noted that the state's witnesses' testimony was, the second time around, "substantially stronger on the issue of [defendant's] identity" — the sole contested issue in both trials. 397 U.S. at 440. Moreover, the Court commented, "The State further refined its case at the second trial by declining to call one of the [witnesses] whose identification testimony at the first trial had been conspicuously negative." 397 U.S. at 440. The Court also quoted from the state's brief that the prosecutor "did what every good attorney would do — he refined his presentation in light of the turn of events at the first trial." 397 U.S. at 447. The Court characterized this as treating "the first trial as no more than a dry run for the second prosecution . . . ." 397 U.S. at 447.
\item \textsuperscript{118} See supra notes 114-15 and accompanying text. Hence, a second trial following a 9-3 acquittal vote is not an independent event tied to its own hypothetical society-wide evaluation of the evidence presented in the second trial. If our de facto acquittal argument is correct, the 9-3 verdict was an acquittal, and the second trial should never have occurred.
\item \textsuperscript{119} Other de facto acquittals would be 10-2 and 11-1. See supra Table 3.
\item \textsuperscript{120} 437 U.S. 1 (1978).
\item \textsuperscript{121} The right to appeal a judge's refusal to poll the jury must also follow. If a trial judge is constitutionally required to take certain action at trial, the violation of that duty is grounds for appeal. See Sandstrom v. Montana, 442 U.S. 510 (1979) (reversing defendant's conviction because trial judge refused properly to instruct jury on state's constitutionally imposed burden of proof); Griffin v. California, 380 U.S. 609 (1965) (reversing conviction because trial judge told jurors they could draw negative inference from defendant's exercise of constitutional right not to testify).
\end{itemize}
A potential doctrinal roadblock exists, however. In *Richardson v. United States* 122 the Court held that a hung jury does not terminate jeopardy and that, therefore, a defendant cannot raise a double jeopardy defense regardless of the cause of the hung jury. The *Richardson* procedural bar means that a defendant cannot even be heard on a claim that insufficient evidence caused the jury to hang. If a jurisdiction labels a 9-3 verdict for acquittal a hung jury, *Richardson* seems to hold that jeopardy has not terminated and, by implication, that a defendant cannot assert that the vote was equivalent to an acquittal.

We do not believe, however, that *Richardson* should be read so rigidly. The Court's rationale, to the extent it had one, 123 follows from the premise that a hung jury is not a verdict. Absent a verdict, the Court concluded, the defendant could point to no "event which terminates jeopardy." 124 As the Court noted in a case doctrinally similar to *Richardson*, "[t]he conceptual difficulty for [the defendant] is that he has not been acquitted; he simply maintains that he ought to have been." 125 Moreover, absent a verdict, no appeal would normally occur, and no other mechanism exists to review the adequacy of the state's proof. *Richardson*, in effect, asked the Court to create an additional level of quasi-appellate review for hung jury mistrial defendants.

Our 9-3 acquittal argument avoids both of the problems that plagued *Richardson*. We do not demand that the record be examined to determine whether the state's case is insufficient as a matter of law and, if so, that this legal judgment supplant the verdictless trial outcome. Instead, we argue that the 9-3 defendant received a de facto acquittal from the jury acting as factfinder. Thus, our 9-3 defendant can point to a verdict that terminates jeopardy in her favor. If the balance of our argument is accepted, a constitutional right to poll the jury must follow.

Once the jury has been polled, we believe the judge must enter an acquittal if the vote is 9-3 or higher in favor of acquittal. Judges al-

---

123. See Thomas, *supra* note 12, at 861-62 (criticizing *Richardson*).
124. 468 U.S. at 325.
125. Justices of Boston Mun. Court v. Lydon, 466 U.S. 294, 307 (1984). *Lydon* is, we believe, indistinguishable from *Richardson*. In *Lydon*, the defendant was convicted in the first tier of a two-tier system, exercised his right to void the conviction by requesting a trial de novo, and then sought to demonstrate that the voided conviction had been based on insufficient evidence. The Court held that Lydon, like Richardson, had no right to a judicial determination of the sufficiency of the evidence. Although Lydon acted to vacate the verdict in his case, while Richardson never received a verdict, the analysis in both cases turns on the lack of a verdict that would support a double jeopardy claim. See generally Thomas, *supra* note 12 (arguing that a valid double jeopardy claim requires an acquittal, an extant conviction, or a prior event that is equivalent to a verdict).
ready possess the power to enter acquittals when juries are hung.126 Our argument simply creates a rationale for entering an acquittal that extends beyond the traditional notion of legally insufficient evidence.

CONCLUSION

By permitting nonunanimous decisions and small juries, the U.S. Supreme Court has risked furthering judicial economy at the expense of constitutional interests. High risks of jury error translate into more wrongful convictions, which in turn condemn innocent defendants to punishment. The burden should rest on the government to show that the systems employed do not too often convict innocent defendants.

Small and nonunanimous juries diminish the probability that a verdict represents the judgment that society as a whole would reach. Because the opinion of society as a whole best defines the concept of guilt, any substantial risk that a jury panel would fail to manifest society’s view creates a risk of convicting an innocent defendant. Both our Close-Call-But-Innocent Defendant Model and our Sample Accuracy Model (SAM) indicate that the risks of error created by the Court’s jury configuration holdings have not been substantial, with the possible exception of the risk created by 9-3 verdicts. For those unresolved configurations, however, a bright line should be drawn to invalidate jury configurations correlating to a 5% or greater overall risk of error as predicted by SAM. These inadequate configurations include, at a minimum, 6-2, 7-2, 7-3, 8-3, and 8-4 verdicts.127

Moreover, if a 9-3 vote of guilty is sufficiently reliable to permit a state to impose a conviction, a vote of 9-3 in favor of acquittal must manifest a lack of guilt. Thus, while a jurisdiction may choose to require unanimity for conviction to ensure as few wrongful convictions as possible, no court should permit a second trial if the first one ended in the defendant’s favor by a vote of 9-3, 10-2, or 11-1. A second trial risks a wrongful conviction or, at the very least, the ordeal of a second trial prohibited by the Double Jeopardy Clause. Implementing this principle simply requires a judge to poll the jury at the defendant’s request and enter a judgment of acquittal if the vote is 9-3 or higher in the defendant’s favor.

We have attempted in this article to “begin over again and concentrate” by taking a fresh look at the interplay between guilt and jury verdicts. Somewhat to our surprise, we discovered that guilt is undefi-
nable without reference to the larger society. We also discovered that our risk-of-error experiments implicated the principle of double jeopardy. When we began this thought experiment, we intended only to test the risk of error in various jury configurations and verdicts. We ended, however, by articulating a more fundamental principle: guilt is nothing more, and nothing less, than the judgment of society. Any verdict that accurately represents how society would have voted is valid, and any acquittal, even if de facto, brings the bar of the Double Jeopardy Clause into play.