

September 1980

## And Then There were Twelve: Statistical Reasoning, the Supreme Court, and the size of the Jury

David Kaye

Follow this and additional works at: <http://scholarship.law.berkeley.edu/californialawreview>

---

### Recommended Citation

David Kaye, *And Then There were Twelve: Statistical Reasoning, the Supreme Court, and the size of the Jury*, 68 CAL. L. REV. 1004 (1980). Available at: <http://scholarship.law.berkeley.edu/californialawreview/vol68/iss5/2>

### Link to publisher version (DOI)

<http://dx.doi.org/https://doi.org/10.15779/Z386T8H>

This Article is brought to you for free and open access by the California Law Review at Berkeley Law Scholarship Repository. It has been accepted for inclusion in California Law Review by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact [jcera@law.berkeley.edu](mailto:jcera@law.berkeley.edu).

# And Then There Were Twelve: Statistical Reasoning, the Supreme Court, and the Size of the Jury

DAVID KAYE†

Before 1970, it was commonly thought that the sixth amendment right to trial by jury<sup>1</sup> precluded the use of criminal juries consisting of less than twelve persons selected from the community at large.<sup>2</sup> In *Williams v. Florida*,<sup>3</sup> however, a divided United States Supreme Court held that a decrease in the size of the criminal jury from twelve to six members did not offend the sixth amendment. Within months, the majority opinion was subjected to harsh criticism, especially for its reliance on four “experiments” that, according to the majority, indicated “no discernible difference” in the results reached by six- and twelve-member juries.<sup>4</sup> It was readily shown that these studies were inconclusive and had little bearing on the question of how size affects jury verdicts.<sup>5</sup> In addition, the critics pointed out that elementary statistical

---

† Professor of Law, Arizona State University. S.B. 1968, Massachusetts Institute of Technology; A.M. 1969, Harvard University; J.D. 1972, Yale Law School. This paper was supported in part by a grant from the Arizona State University Faculty Grant-in-Aid Program. I am also grateful to Dennis Karjala for many helpful observations about the Poisson model discussed in Part III, to Richard Lempert and Dennis Young for comments on an earlier version of this Article, and to Nancy Kaye for her assistance in the preparation of the Appendix.

1. “In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed, which district shall have been previously ascertained by law. . . .” U.S. CONST. amend. VI. The sixth amendment right to a jury is applicable to state trials through the fourteenth amendment. *Duncan v. Louisiana*, 391 U.S. 145 (1968). The Court is divided as to the extent to which the states have latitude not available to the federal government to modify the details of the jury system. *See, e.g., Ballew v. Georgia*, 435 U.S. 223, 246 (1978) (Powell, J., concurring).

2. *See, e.g., Rochin v. California*, 342 U.S. 165, 169-70 (1952); *Patton v. United States*, 281 U.S. 276 (1930); *Thompson v. Utah*, 170 U.S. 343 (1898); Wells, *Unanimity Verdicts*, 63 ALB. L.J. 360, 362 (1901). *But cf. Maxwell v. Dow*, 176 U.S. 581 (1900) (state statute providing for jury of eight in non-capital criminal cases held not a denial of due process). Those who drafted and ratified the Constitution arguably conceived of juries as bodies of 12. *See* authorities cited note 24 *infra*.

3. 399 U.S. 78 (1970).

4. *Id.* at 101.

5. The “experiments” relied on by the Court are speculative or impressionistic reports based on little or no experience with the six-member jury. *See* Zeisel, . . . *And Then There Were None: The Diminution of the Federal Jury*, 38 U. CHI. L. REV. 710, 713-15 (1971); Note, *The Effect*

reasoning indicated that juries of six could be expected to prove less representative than juries of twelve.<sup>6</sup>

Three years later the Court returned to the issue of jury size.<sup>7</sup> In *Colgrove v. Battin*,<sup>8</sup> the Court indicated that its faith in its unproven "functional-equivalence" hypothesis was unshaken. In upholding a reduction in the size of the civil jury from twelve to six,<sup>9</sup> the Court continued to rely, without apparent embarrassment, on the studies noted in *Williams*. Moreover, it cited four additional studies as providing "convincing empirical evidence" in support of the theory.<sup>10</sup> Again, it was not long before telling criticism of the social science studies found its way into print,<sup>11</sup> further undermining the factual premise of the majority opinions in *Williams* and *Colgrove*.<sup>12</sup>

In the years following *Colgrove*, more articles were written, more studies were commissioned, and more doubts about the effectiveness of smaller juries were raised.<sup>13</sup> Then, eight years after *Williams*, every

---

*of Jury Size on the Probability of Conviction: An Evaluation of Williams v. Florida*, 22 CASE W. RES. L. REV. 529 (1971).

6. See, e.g., Zeisel, *supra* note 5, at 715-24; Note, *supra* note 5, at 547-52.

7. During this time, the Court held that the traditional requirement of jury unanimity was not constitutionally essential. *Johnson v. Louisiana*, 406 U.S. 356 (1972); *Apodaca v. Oregon*, 406 U.S. 404 (1972).

8. 413 U.S. 149 (1973).

9. The assumption that "jury" denoted a twelve-person jury may also have been held with regard to the seventh amendment and civil juries, particularly in the federal system. The seventh amendment reads in part: "in [s]uits at common law, where the value in controversy shall exceed twenty dollars, the right of trial by jury shall be preserved. . . ." Early cases in dicta equated the jury mentioned in the seventh amendment with a jury of 12 people. *Maxwell v. Dow*, 176 U.S. 581, 586 (1900); *Capital Traction Co. v. Hof*, 174 U.S. 1, 13-14 (1899). However, such references almost always addressed federal juries; the states were left free to administer the civil jury as they saw fit. By 1952, many states had implemented juries of less than 12 for certain civil cases. See *Rochin v. California*, 342 U.S. at 170 n.3. It was not until 54 federal district courts had promulgated rules of court implementing six-person juries that the constitutionality of smaller civil juries was decided. *Colgrove v. Battin*, 413 U.S. at 150 n.1. The local rules of court were also attacked on statutory grounds, to no avail. *Id.* at 150-51, 161-64.

10. 413 U.S. at 159 n.15.

11. Zeisel & Diamond, "Convincing Empirical Evidence" on the Six-Member Jury, 41 U. CHI. L. REV. 281 (1974). For additional criticism of some of the post-*Williams* studies, see, e.g., Padawer-Singer, *Justice or Judgments?*, in THE AMERICAN JURY SYSTEM 45 (Roscoe Pound-American Trial Lawyers Foundation 1977); M. SAKS, JURY VERDICTS 47 (1977).

12. Lempert, *Uncovering "Nondiscernible" Differences: Empirical Research and the Jury-Size Cases*, 73 MICH. L. REV. 643, 645 n.9 (1975); Saks, *Ignorance of Science Is No Excuse*, 10 TRIAL 18 (1974).

13. Much of the intervening literature is cited in Justice Blackmun's opinion in *Ballew v. Georgia*, 435 U.S. at 231-39. See also Davis, Bray & Holt, *The Empirical Study of Decision Processes in Juries: A Critical Review*, in LAW, JUSTICE, AND THE INDIVIDUAL IN SOCIETY: PSYCHOLOGICAL AND LEGAL ISSUES 326 (J. Tapp & F. Levine eds. 1977); Lermack, *No Right Number? Social Science Research and the Jury-Size Cases*, 54 N.Y.U.L. REV. 951, 956-61 (1979); Penrod & Hastie, *Models of Jury Decision-Making: A Critical Review*, 86 PSYCH. BULL. 462 (1979); Sperlich, . . . *And Then There Were Six: The Decline of the American Jury*, 63 JUDICATURE 262 (1980); Volrath & Davis, *Jury Size and Decision Rule*, in THE JURY: ITS ROLE IN AMER-

member of the Court concluded in *Ballew v. Georgia*<sup>14</sup> that a criminal conviction returned by a five-member jury stretched the sixth amendment beyond the breaking point.<sup>15</sup>

However, most of the Justices had grown noticeably gun-shy of the social science data. Only Justice Blackmun, in his separate opinion announcing the judgment of the Court, saw fit to discuss the burgeoning mass of jury-size studies. His opinion, the most elaborate to issue from the highly fractured Court,<sup>16</sup> conceded that the studies mentioned in *Williams* and *Colgrove* were less than convincing, but insisted that the holdings of those cases remained intact. Indeed, the Blackmun opinion undertook a comprehensive review of the research literature and cited yet another two studies—these deductive rather than empirical—as specifically supporting the conclusion that six jurors are enough but five are too few.<sup>17</sup>

Three of the Justices sought to justify the distinction between six- and five-member juries in another fashion. Justice Powell, joined by Chief Justice Burger and Justice Rehnquist, preferred to rely entirely on the familiar refrain that a line must be drawn somewhere, even if it

ICAN SOCIETY 73 (R. Simon ed. 1980). Despite the chorus of voices raised against smaller juries, by April 1978, 38 states had authorized juries of fewer than 12 in some civil actions, 34 had sanctioned smaller juries for some criminal trials, and 80 of the 95 federal district courts had adopted rules reducing the size of juries in civil actions. N.Y. Times, Apr. 26, 1978, at A1, col. 5; NATIONAL CENTER FOR STATE COURTS, FACETS OF THE JURY SYSTEM: A SURVEY 5 (1976).

14. 435 U.S. 223 (1978).

15. *Ballew* was hardly the first instance in which the use of five-person juries was questioned. See Note, *The Sacred Twelve*, 5 GREEN BAG 550, 551 (1893).

16. There was no majority opinion. Justice Blackmun announced the judgment of the Court and wrote a separate opinion joined only by Justice Stevens. Justice White concurred in the judgment, but his one-paragraph statement reveals little regarding his views toward Justice Blackmun's reasoning. Justice Powell concurred, joined by Chief Justice Burger and Justice Rehnquist. Justice Brennan penned a one-paragraph opinion joining Justice Blackmun's opinion "insofar as it holds that the Sixth and Fourteenth Amendments require juries in criminal trials to contain more than five persons." 435 U.S. at 246. This "joining opinion" was itself joined by Justices Stewart and Marshall.

17. The two studies are Friedman, *Trial by Jury: Criteria for Convictions, Jury Size and Type I and Type II Errors*, 26-2 AM. STAT. 21 (Apr. 1972) and Nagel & Neef, *Deductive Modeling to Determine an Optimum Jury Size and Fraction Required to Convict*, 1975 WASH. U.L.Q. 933.

Discussions of Justice Blackmun's use of the social science literature can be found in Lermack, *supra* note 13; Sperlich, *supra* note 13; Sperlich, *Trial by Jury: It May Have a Future*, 1978 SUP. CT. REV. 191; Sperlich, *The Changing American Jury: Observing the Eighth Anniversary of the Beginning of Its End?* (paper prepared for delivery at 1978 Annual Meeting of the American Political Science Association); Tanke & Tanke, *Getting Off a Slippery Slope: Social Science in the Judicial Process*, 34 AM. PSYCH. 1130 (1979). With the exception of Lermack, these papers commend Justice Blackmun on his understanding and study of this literature. They fault the Justice only for not citing more such studies and for ignoring the arguments in many of them that the six-person jury is functionally dissimilar to the twelve-member body. In contrast, this Article argues that many of the statistical models favored by Justice Blackmun and many social scientists are of limited usefulness. See text accompanying notes 85-107 *infra*.

cannot be drawn perfectly.<sup>18</sup> These Justices chastised Justice Blackmun for his resort to "numerology."<sup>19</sup>

Finally, just last year, in *Burch v. Louisiana*<sup>20</sup> all the Justices agreed that a six-member criminal jury which was not required to return a unanimous verdict fell short of the line fixed in *Williams* and *Ballew*. Justice Rehnquist, writing for the Court and taking no cognizance of the relevant empirical and mathematical literature, at one point reiterated the central dogma of *Williams*: "[A] jury of 12 is neither more reliable as a factfinder, more advantageous to the defendant, nor more representative of the variety of viewpoints in the community than a jury of six."<sup>21</sup>

This Article investigates whether the jury-size conclusions articulated in *Ballew* and *Burch* can be reconciled with the Court's announced adherence to the reasoning of *Williams*. It suggests that such a reconciliation is possible—but only if *Ballew* and *Burch* are construed as establishing a new, and ultimately unsatisfactory, test for the constitutionality of deviations from the common law jury. Part I enumerates three possible standards by which the constitutionality of such

18. 435 U.S. at 246 (Powell, J., concurring). This theme was sounded again by Justice Rehnquist in his opinion for the Court in *Burch v. Louisiana*, 441 U.S. 130, 137-39 (1979).

19. 435 U.S. at 246 (Powell, J., concurring) ("Also, I have reservations as to the wisdom—as well as the necessity—of Mr. Justice Blackmun's heavy reliance on numerology derived from statistical studies. Moreover, neither the validity nor the methodology employed by the studies cited was subjected to the traditional testing mechanisms of the adversary process. The studies relied on merely represent unexamined findings of persons interested in the jury system.") Justice Blackmun's response was powerful and included a classic *tu quoque*:

Some of these studies have been pressed upon us by the parties. . . .

We have considered them carefully because they provide the only basis, besides judicial hunch, for a decision about whether smaller and smaller juries will be able to fulfill the purpose and functions of the Sixth Amendment. Without an examination about how juries and small groups actually work, we would not understand the basis for the conclusion . . . that "a line has to be drawn somewhere." We also note that The Chief Justice did not shrink from the use of empirical data in *Williams v. Florida*, . . . when the data were used to support the constitutionality of the six-person criminal jury, or in *Colgrove v. Battin*, . . . a decision also joined by Mr. Justice Rehnquist.

*Id.* at 232 n.10 (separate opinion). Nevertheless, only 4 of the 20 studies cited by Justice Blackmun were mentioned by the parties, and none of these were discussed in any detail. See Brief for Petitioner at 7-9.

20. 441 U.S. 130 (1979).

21. *Id.* at 135 n.7. Justice Rehnquist asserted that these views were "not questioned" in *Burch*. *Id.* at 137. This pretense seems odd in light of the amicus brief submitted by the American Civil Liberties Union. That brief disputed these claims and urged the Court to overrule *Williams*. See Brief for American Civil Liberties Union as Amicus Curiae.

*Burch* was held to have retroactive application (at least as to cases still pending on direct review at the time it was decided) in *Brown v. Louisiana*, 100 S.Ct. 2214 (1980). Justices Brennan, Stewart, Marshall, and Blackmun, relying on Justice Blackmun's analysis of the social science literature in *Ballew*, found that nonunanimous six-member juries threatened the integrity of the fact-finding process. Whether these first three Justices concur in the view espoused by Justices Blackmun and Stevens in *Ballew* that the jury size studies permit a constitutional distinction between five- and six-person juries is not yet clear. See note 16 *supra*.

departures might be judged, and it describes criteria that can be used to measure the extent to which a modified jury differs from the traditional jury of twelve. Applying rudimentary statistical theory, Part II evaluates the familiar line-drawing argument as a basis for the Court's holdings. It shows that under the conventional interpretation of *Williams*, which invalidates any change in jury size that would discernibly alter jury functioning, Justice Powell's glittering generality in *Ballew* that a line must be drawn somewhere<sup>22</sup> does not support any constitutional distinction between six- and five-member juries that is not also applicable to *any* jury with less than twelve members. Part III analyzes the more elaborate statistical models cited by Justice Blackmun in *Ballew*. This section shows that the Justice's treatment of the statistical literature is, at best, careless, and that the statistical studies do not support the conclusion that juries of six are functionally superior to juries of twelve. Part IV considers the possibility that *Ballew* implicitly establishes a new test for the constitutionality of changes in jury size, and argues that such a new standard is unworkable and undesirable. The Article concludes that any diminution in the size of the jury from the traditional twelve should be rejected as unconstitutional.

## I

### THE CONSTITUTIONAL TESTS

#### A. *The Standards*

The principal argument for deviating from the historic norm of twelve-member juries is the reasonable belief that smaller juries cost less.<sup>23</sup> The constitutional question is, however, whether such smaller and presumably cheaper bodies are adequate to assure litigants their right to a jury trial.

In theory at least, there are a variety of standards for determining which characteristics of the common law jury are constitutionally indispensable. Perhaps the three most plausible standards are, briefly: (1) a purely historical test, which would insist on adherence to eighteenth century practice, since that was the framework on which the Constitution was crafted; (2) a no-functional-difference test, which would allow changes so long as they created no discernible differences in the jury's

---

22. 435 U.S. at 246 (Powell, J., concurring).

23. See Lermack, *supra* note 13, at 953. However, the savings produced by reducing jury size are probably not proportional to the magnitude of the reduction. A variety of studies comparing twelve- and six-member juries have found little or no savings in the time spent on jury selection or in trial time. Padawer-Singer, *supra* note 11, at 50. Other studies of the cost savings associated with six-member juries are reviewed by Lermack, who concludes that "although some economies are achieved when jury size is reduced from twelve to six, the savings are less, both in time and money, than what was predicted." Lermack, *supra* note 13, at 961. See also Sperlich, *supra* note 13, at 276; Zeisel, *supra* note 5, at 711-12.

functioning; and (3) a "concept-conception" test, which would countenance any change so long as the resulting decisionmaking body fell within the concept of a jury—a body structured to promote group deliberation and to provide a fair possibility that a cross-section of the community will be represented on it.

The first and most stringent standard would demand that the jury size at the time of the adoption of the Constitution be maintained. There are strong indications that in providing for a right to trial by jury, those who drafted and ratified the Constitution meant to mandate the use of a body of exactly twelve persons.<sup>24</sup> Consequently, this historical test would preclude *any* change in the size of the jury. In *Williams*, however, the Court rejected this rigidly historical analysis, disparaging it as "blind formalism."<sup>25</sup> Instead, Justice White's majority opinion insisted that the use of twelve jurors is sheer "historical accident" and "wholly without significance 'except to mystics.'"<sup>26</sup>

Although this treatment of the Constitution's history may be heavy-handed,<sup>27</sup> there are arguments that support the Court's view that the historical record is not determinative. These arguments suggest that other constitutional tests may be appropriate. Perhaps the most appealing alternative looks to jury functioning and is derived essentially from a "negative pregnant" textual argument. Starting from the premise that the framers believed that twelve-member juries would best protect certain values, this argument emphasizes that the framers did not, however, specify the exact number of jurors in the Constitution. It concludes that jury size can be modified provided it is demonstrated that the values associated with twelve-member juries are not thereby threatened.<sup>28</sup> This test is quite restrictive. It would allow jury size to vary, if at all, only within a range where it could be said with assurance that there was substantial functional equivalence between the traditional and smaller juries. This would ensure that the constitutional values grounded in the twelve-person jury were not impaired. Although somewhat literalistic, this argument preserves the substance of the right to trial by jury while accommodating some alterations in the details of

---

24. See Lempert, *supra* note 12, at 660 n.55. For further discussion of the history of the jury, see, e.g., Devlin, *Jury Trial of Complex Cases: English Practice at the Time of the Seventh Amendment*, 80 COLUM. L. REV. 43 (1980); Foster, *Trial by Jury: The Thirteenth Century Crisis in Criminal Procedure*, 13 U. BRIT. COLUM. L. REV. 280 (1979); Singley, *Ballew v. Georgia: Five is Not Enough*, 52 TEMP. L.Q. 217, 220-24 (1979); authorities cited in Sperlich, *Trial by Jury: It May Have a Future*, *supra* note 17, at 196 n.18.

25. 399 U.S. at 89, 102, *quoted in*, e.g., *Burch v. Louisiana*, 441 U.S. at 135.

26. 399 U.S. at 103.

27. See note 24 and accompanying text *supra*. Nor is this the first opinion penned by the Justice that can be faulted on such grounds. See Kaye, *Congressional Papers, Judicial Subpoenas and the Constitution*, 24 U.C.L.A. L. REV. 523, 542-43 (1977).

28. This argument is advanced by Lempert, *supra* note 12, at 661 n.55.

the jury structure. Thus, it avoids the charges of "blind formalism" and "mysticism" leveled in *Williams* against the purely historical test.<sup>29</sup>

The no-functional-difference standard that emerges from this analysis of the jury clauses in the Constitution is akin to, if not identical with, the emphasis on operational differences in *Williams*.<sup>30</sup> Having rejected the historical test, the *Williams* Court reasoned that a body that differs even markedly from the common law jury in its superficial appearance can nevertheless qualify as a "jury" as long as it is structured "to promote group deliberation . . . and to provide a fair possibility for obtaining a representative cross-section of the community."<sup>31</sup> Assuming that these values are substantially similar to those cherished by the framers, *Williams* was correct in searching for functional differences that might have endangered those values. However, the Court failed—dismally—in the application of its no-functional-difference test. As the critics skilled in the social sciences were quick to show, the Court's conclusion that there was "no discernible difference" in the functioning of twelve- and six-person juries was naive or worse.<sup>32</sup>

On the basis of the no-functional-difference test, most commentators have assumed that if probable discrepancies in the behavior of differently sized juries could be highlighted, the smaller juries would automatically be found unconstitutional.<sup>33</sup> But the no-functional-difference standard is not the only alternative to a strictly historical analysis, as subsequent interpretations of *Williams* have shown. *Williams* and its progeny do not clearly delineate the extent to which a body can operate differently than the common law jury, yet still qualify as a "jury" for constitutional purposes. In *Williams*, for example, the very

29. Indeed, it may well be what Justice White had in mind in claiming that "there is absolutely no indication in 'the intent of the Framers' of an *explicit* decision to equate the constitutional and common law characteristics of the jury." *Williams v. Florida*, 399 U.S. at 99 (emphasis added).

30. If there is a difference, it is that the Court in *Williams* seemed willing to permit alterations in jury size so long as no resulting effects could be shown. Under a pure no-functional-difference test, an initial affirmative burden would be placed on the proponents of the smaller juries to show functional equivalence. See Lempert, *supra* note 12, at 661 n.55. Yet, in form at least, the Court has adhered to the principle of resolving "substantial doubts" about the effects of changes in the jury structure against such changes. See, e.g., *Burch v. Louisiana*, 441 U.S. at 136-37; *Ballew v. Georgia*, 435 U.S. at 239.

31. 399 U.S. at 100. In *Duncan v. Louisiana*, 391 U.S. 145, 156 (1968), as well as in most subsequent cases, the Court has stated that the purpose of trial by jury is to provide "a safeguard against the corrupt or overzealous prosecutor and against the compliant, biased, or eccentric judge." There are, however, important functions of the jury system not recognized by this narrow formulation. See generally G. CALABRESI & P. BOBBITT, TRAGIC CHOICES 57-64 (1978); Lempert, *Jury Size and the Peremptory Challenge*, 22 LAW QUADRANGLE NOTES 8, 10 (1978); Singley, *supra* note 24, at 231-35.

32. See text accompanying notes 4-6; authorities cited note 12 *supra*; Sperlich, *Trial by Jury: It May Have a Future*, *supra* note 17, at 208-09.

33. E.g., Sperlich, *Trial by Jury: It May Have a Future*, *supra* note 17, at 218-22.

failure of the majority to perceive *any* "discernible" differences in the functioning of twelve- and six-member juries spared them from deciding whether *some* (presumably small) differences could be tolerated. Thus, it is unclear whether *Williams* even establishes a generally applicable no-functional-difference standard. In *Ballew*, several Justices were willing to concede that six-person juries *do* operate at least somewhat differently than their twelve-member counterparts.<sup>34</sup> Yet, rather than revise their opinions of the constitutionality of the six-member jury to conform to the strict no-functional-difference test, these Justices blandly asserted that the six-member jury met the functional test of *Williams* while the five-person jury did not. This implies that within certain bounds, functional differences, like superficial differences, will not render smaller juries unconstitutional.

This in turn implies that there are other possible rejoinders to the argument that the framers desired to "crystallize" the jury as a twelve-person body. The most plausible of these<sup>35</sup> builds on a distinction discussed in recent philosophical analysis. This approach treats the "jury"

---

34. 435 U.S. at 234-38 (separate opinion of Justice Blackmun), 245-46 (Powell, J., concurring) (the latter implicitly acknowledging that six-member juries function similarly to five-member panels).

35. A less persuasive response to the historical argument should also be noted. This very different argument for allowing deviations from the historical norm simply asserts that some distortion in jury functioning can be justified by substantial economies in the operation of smaller juries. Indeed, in *Ballew* and *Burch*, the Court indulged in the assumption that "countervailing" government interests such as fiscal and administrative savings could justify an otherwise unconstitutional departure from the common law jury. *Burch v. Louisiana*, 441 U.S. at 139; *Ballew v. Georgia*, 435 U.S. at 249. The Court's reference to bureaucratic efficiency, particularly in an area of criminal procedure, is disturbing. First, administrative costs are, to some degree, in the control of the government. For example, if the government chose to curtail plea bargaining, the number of jury trials would undoubtedly increase. Could the government then point to the enhanced potential savings in the use of even smaller juries to justify reducing jury size to a number previously held to be unconstitutional? Furthermore, even if some attention to administrative convenience is permissible, there is a point at which the jury system has been so altered that no further changes can be justified by considerations of efficiency. As the ensuing discussion of the jury as a contested "concept" will underscore, no analysis of the jury-size issue can afford to overlook the obvious point that the Constitution specifies a "jury." This right to trial by jury attaches whether or not it is inconvenient or expensive. Merely calling a decisionmaking body a jury does not make it one. If tinkering with the common law jury changes the institution to the point where it behaves grossly differently than the standard conception of the jury, the neoteric system is unconstitutional regardless of the countervailing interests the government can point to.

The jury size issue thus differs from the many cases in which the Court has considered whether procedural due process necessitates trial-type protections in contexts where the Constitution does not explicitly guarantee such protections. There the Court has weighed the importance of safeguards in achieving just results against the administrative costs of implementing the procedures. *E.g.*, *Goldberg v. Kelly*, 397 U.S. 254 (1970). See generally Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28, 49-52 (1976); Saphire, *Specifying Due Process Values: Toward a More Responsive Approach to Procedural Protection*, 127 U. PA. L. REV. 111 (1978); Note, *Specifying the Procedures Required by Due Process: Toward Limits on the Use of Interest Balancing*, 88 HARV. L. REV. 1510 (1975).

spoken of in the Bill of Rights as a "concept" and regards the twelve-member jury with which the framers were acquainted as a mere "conception" capable of being replaced by other conceptions included within the more general concept.<sup>36</sup> If adopted, this approach would permit a wider range of jury sizes. It would still require the proponents of smaller juries to establish that the particular structure they would substitute for the framers' conception is not so functionally dissimilar as to fall outside the contours of the broader concept embodied in the Constitution. However, the restructured version would not necessarily have to be the functional equivalent of the twelve-person jury, and therefore, the degree of protection afforded certain basic values might also differ.<sup>37</sup>

Because it does not appear that all the Justices are willing to read *Williams* as condemning any and all changes in the jury system that threaten the values associated with the sixth and seventh amendments, this Article will consider the Court's attempt to draw the line at six-member juries in terms of both the strict no-functional-difference test *and* the less restrictive concept-conception distinction. We shall thus ask whether functional differences affecting constitutional values are likely to emerge as jury size is reduced from twelve, and at what point these differences become large enough to remove the smaller bodies from the constitutional concept of the jury. In this way, we shall try to locate the respective limits beyond which the probable impairment in jury functioning can no longer be justified.

### B. The Criteria

To facilitate this inquiry into the effects of jury size on jury func-

---

36. See generally Gallie, *Essentially Contested Concepts*, 56 PROC. ARISTOTELIAN SOC'Y 167 (1956); Perry, *Contested Concepts and Hard Cases*, 88 ETHICS 20 (1977); J. RAWLS, A THEORY OF JUSTICE 10 (1971). This distinction between concept and conception has been applied to justify judicial recognition of changing moral standards in construing constitutional provisions giving legal status to "moral rights." See, e.g., R. DWORKIN, TAKING RIGHTS SERIOUSLY 147 (1977); O'Fallon, *Adjudication and Contested Concepts: The Case of Equal Protection*, 54 N.Y.U.L. REV. 19 (1979); Radin, *The Jurisprudence of Death: Evolving Standards for the Cruel and Unusual Punishment Clause*, 126 U. PA. L. REV. 989, 1031-32 (1978). There is some doubt about the applicability of the concept-conception distinction to disputations about the nature of the jury. For example, it is not clear that the concept of a jury is essentially evaluative or is "persistently vague." See Gallie, *supra*, at 172 n.1.

37. The concept-conception test allows the courts greater freedom in selecting and valuing the objectives and values of the jury system than does the no-functional-difference approach. Of course, using this freedom, a court could simply employ the pure historical test or the no-functional-difference test to delimit the jury concept. At the other extreme, a court could approve a restructured jury system that serves in the most minimal way possible values associated with an independent factfinding body. Because the test is basically without form as an initial matter, serious problems arise when courts attempt to implement it. See text accompanying notes 110-11 *infra*.

tioning, the values protected by the jury system must be delineated with some care. Given the purposes of the jury identified in *Williams* and its progeny, any change in the size of the jury should be scrutinized for possible adverse effects on the quality of jury decisionmaking and factfinding.<sup>38</sup> Inasmuch as the "quality" of jury performance is an amorphous term, the constitutional inquiry should be focused on more specific components: representativeness, accuracy, and reliability.

Any functional analysis must consider the impact a modification in size might have on the ability of the jury to secure a fair cross-section of the community. This may be called the question of representativeness, and it is logically distinct from the issues of accuracy and reliability.<sup>39</sup> A system of "blue ribbon" or "expert" juries, for example, might reach more accurate verdicts<sup>40</sup> and produce more predictable results. Yet, such a system would hardly satisfy the demand for representativeness. Representativeness as an element of quality must be studied if a deviation from the traditional jury system is to be upheld as constitutional.

Of course, the constitutional tests that look to jury quality must also focus on whether a particular transformation of the common law jury threatens to detract from decisional accuracy and reliability. In most cases, a good decision is one that is accurate and reliable.<sup>41</sup> An accurate verdict or finding is one that reflects the true state of affairs.<sup>42</sup> For example, a jury determination that a warranty was breached when in fact the product functioned as warranted would amount to an inaccurate or erroneous decision.<sup>43</sup> Reliability, on the other hand, relates to

38. Factfinding can be viewed as a species of decisionmaking. *See, e.g.*, V. BARNETT, *COMPARATIVE STATISTICAL INFERENCE* 201 (1973). When not desiring to distinguish between jury verdicts and jury findings of fact, I shall usually use the broader term jury decisions.

39. Although representativeness, accuracy, and reliability are distinct at the conceptual level, I shall argue that these features tend to be highly correlated, especially when one is restricted to more or less random selection of jurors. *See, e.g.*, text accompanying notes 60-71 *infra*.

40. The assumption that "blue ribbon" juries return more accurate verdicts can be challenged. *See, e.g.*, Note, *The Case for Black Juries*, 79 *YALE L.J.* 531 (1970). Indeed, heterogeneity may well enhance accuracy in many situations. *See* text accompanying notes 63-71 *infra*.

41. In some instances, a "good" jury decision may be one which departs from the law and is founded instead on the jury's sense of what is just in the circumstances. *See, e.g.*, Lempert, *supra* note 31, at 9. Obviously, it is difficult to devise, even in principle, a measure of the quality of a jury's nullification decision. *See, e.g.*, R. WASSERSTROM, *THE JUDICIAL DECISION: TOWARD A THEORY OF LEGAL JUSTIFICATION* ch. 5 (1961). Still, it seems fair to insist that a jury which invokes its implicit power to nullify should base its actions on an accurate assessment of the facts of the case. Moreover, once it correctly appreciates the factual and legal situation, the jury's decision to depart from the judge's instructions should not be aberrant, and it should be representative of the community's sense of justice. Ensuring accuracy, reliability and representativeness—the important features of jury decisionmaking discussed in the text—thus should promote the proper exercise of the jury's power to nullify.

42. *See, e.g.*, Lempert, *Modeling Relevance*, 75 *MICH. L. REV.* 1021 (1977).

43. Of course, there is typically no practicable, independent way to ascertain the truth of the

stability, or reproducibility, of results, whether or not correct.<sup>44</sup> A jury verdict can be said to be reliable if it is likely that a second randomly selected jury of the same size, observing the same voting protocol, and presented with identical evidence, argument, and instructions would return the same verdict. Any structural change that is likely to increase the incidence of mistaken verdicts or to undermine verdict reliability ought to trigger careful scrutiny under the constitutional standards grounded in these values.

## II

### THE LINE-DRAWING METAPHOR

Does an inquiry into representativeness, accuracy, and reliability support the result that a six-person jury is constitutional whereas a five-member body is not? The Court's recent approach, as shown above, is essentially to maintain that a line must be drawn somewhere, and that some arbitrariness in its exact location is tolerable. In this general form, however, the line-drawing metaphor blurs two distinct arguments together.

On the one hand, the metaphor might be a claim that differences in quality are minor as jury size is steadily reduced from twelve to six, but that a relatively large effect is experienced as one approaches the five-member jury. Graphically, this version of the line-drawing argument asserts that a plot of the "costs"—the untoward effects on quality—drawn as a function of jury size steepens sharply as one moves from a six- to a five-member jury. It is, in brief, an argument about marginal costs.<sup>45</sup> Construed in this fashion, however, the line-drawing argument seems misguided. As a constitutional matter, the Court

---

matter before the jury. See L. COHEN, *THE PROBABLE AND THE PROVABLE* 120 (1977). For analytic purposes, however, one can speak meaningfully of correct and incorrect verdicts and of accurate and inaccurate findings of fact. See, e.g., Kaye, *Probability Theory Meets Res Ipsa Loquitur*, 77 MICH. L. REV. 1456 (1979); *The Laws of Probability and the Law of the Land*, 47 U. CHI. L. REV. 34 (1979).

44. Theoretically, of course, if all juries always returned accurate verdicts, verdict reliability would be assured. Unfortunately, juries sometimes err, although there is generally no independent method for ascertaining the accuracy of jury verdicts. There are reasons for entertaining an independent analysis of the relationship between jury size and reliability under these circumstances. For example, to the extent that one assumes that unreliable verdicts must also signify potential inaccuracies, reliability itself becomes a badly needed measure of accuracy. Cf. Justice Blackmun's use of reliability statistics in *Ballew v. Georgia*, 435 U.S. at 234, 237-38 (reliability equated with accuracy; more unreliable juries also assumed to be less accurate).

45. However, it is not founded on the putative desirability of marginal cost pricing. Certainly, it might be economically efficient to choose a size for which the marginal costs (in terms of the loss in accuracy, reliability, and representativeness) equaled marginal benefits (in dollar savings attributable to smaller panels). But economic efficiency is not a constitutional value; the economically optimal size may lie below the zone of constitutionally permissible functional differences. See note 35 *supra*.

should be concerned, not with *marginal* costs, but with the *net* effect of the change in jury size, since *each* marginal increase reflects an incremental sacrifice in the underlying values. Moreover, even taken at face value, the argument fails, for there is reason to think that the cost curve does not steepen sharply as jury size diminishes from six to five.<sup>46</sup>

Consequently, it seems more correct to interpret the Court's allusion to line-drawing as an argument about total costs. That is, the metaphor is probably meant to express the Court's judgment that the adverse impact of reductions in jury size is consistently small as size decreases from twelve to eleven, from eleven to ten, and so on, but that the cumulative effect becomes intolerable when the small loss in moving from six to five is added to all the other small losses experienced in the transition. The "total cost" argument, then, does not pretend that the marginal cost of the reduction from six to five jurors is especially disturbing. It merely asserts that at about that point the net effect becomes large enough to tip the balance against further size reductions, either because at about this size important functional differences become discernible, or because functional differences become so dramatic that the body no longer falls within the constitutional concept of the jury as the Court perceives it.

At first blush, this version of the argument is appealing, perhaps because it is so familiar. Every first year law student soon perceives that there are few naturally "bright lines" to separate what is lawful from what is not. Confronted with one variation after another on a set of hypothetical facts, the student quickly learns that one way to arrest a precipitous descent down the omnipresent "slippery slope" is to declare that the position being defended admits of some arbitrariness, but to insist that the line must be drawn somewhere.<sup>47</sup> Yet, at this point the argument is incomplete. To justify drawing the line at the admittedly arbitrary point selected, the student should also demonstrate that this point lies within the broad zone where it is arguably proper to draw the

---

46. The conclusion follows from the statistical reasoning discussed at text accompanying notes 52-77 *infra*. The analysis appropriate to the "marginal costs" version of the line-drawing argument is presented in Tables 2 and 4, and Figures 1 and 2 of the Appendix. These materials suggest that generally the effects on representativeness, accuracy, and reliability increase slightly but smoothly as jury size is reduced in uniform increments.

47. *Cf., e.g., Burch v. Louisiana*, 441 U.S. at 137 ("[L]ines must be drawn somewhere . . . . This line drawing process, . . . cannot be wholly satisfactory, for it requires attaching different consequences to events which . . . actually differ very little."); *Louisville Gas Co. v. Coleman*, 277 U.S. 32, 41 (1928) (Holmes, J., dissenting) ("When a legal distinction is determined, as no one doubts it may be, between night and day, childhood and maturity, or any other extremes, a point has to be fixed or a line has to be drawn, or gradually picked out by successive decisions, to mark where the change takes place. Looked at by itself without regard to the necessity behind it, the line or point seems arbitrary. It might as well or might nearly as well be a little more to one side or the other.").

line.<sup>48</sup> Only if one is working within this "zone of reasonableness" can one afford to be arbitrary.<sup>49</sup>

Similarly, the argument that a jury of six can be sustained while a jury of five cannot, simply because "a line must be drawn somewhere," is insufficient. If the line-drawing metaphor is being used in conjunction with the no-functional-difference test, then it must also be shown that the deleterious effects associated with the reduction in size from twelve to six are at once negligible and of the same magnitude as the further impact of the move from six to five. If the former losses overshadow the latter, then *both* five- and six-member panels are functionally distinguishable from twelve-person juries, and the zone in which the Court should reasonably be worried about drawing the line is closer to twelve than to six. On the other hand, if the Court is speaking of drawing a different line—the one that demarcates the boundary of the constitutional concept of a jury—some nontrivial losses in the transition to smaller juries might be acceptable. However, the question of whether the differences in degree are by that point so large as to amount to a difference in kind still must be addressed. If that question cannot be answered on a principled basis, then the workability of the concept-conception standard itself is thrown into doubt.

To gauge how seriously jury functioning is likely to be affected as jury size changes, one can look to elementary statistical reasoning, to studies of juries and other small groups, and, of course, to common sense. In the remainder of this section, I argue on these bases that the probable impact on representativeness, accuracy, and reliability in the change from six to five is negligible in comparison to the already substantial losses suffered in the transition from twelve to six. I conclude that to the extent that a no-functional-difference test is being used, this demonstration renders the line-drawing metaphor pointless. However, I do not maintain that the analysis necessarily precludes drawing the line between six and five on the basis of the looser concept-conception

---

48. *Cf., e.g., Louisville Gas Co. v. Coleman*, 277 U.S. at 41 (Holmes, J., dissenting) ("But when it is seen that a line or point there must be, and that there is no mathematical or logical way of fixing it precisely, the decision of the legislature must be accepted unless we can say that it is very wide of any reasonable mark."); *Pearce v. Commissioner*, 315 U.S. 543, 558 (1942) (Frankfurter, J., dissenting) ("But the fact that a line has to be drawn somewhere does not justify its being drawn anywhere.").

49. And even within this zone, there may be some plausible, though not compelling, basis for narrowing the field of choice. See W. TWINING & D. MIERS, *HOW TO DO THINGS WITH RULES: A PRIMER OF INTERPRETATION* 125-26 (1976). This consideration is explored at note 111 and accompanying text *infra*.

In some ways, the claim that there is a "zone of reasonableness" outside of which the line-drawing argument is inapposite is analogous to the commonly held view that legal rules have some "open texture," but that outside this region of "hard cases" the proper application of the rule is clear. See generally, *e.g., H.L.A. HART, THE CONCEPT OF LAW* (1961); R. DWORKIN, *supra* note 36.

test. But it raises serious doubts about whether the concept-conception distinction provides an appropriate standard by which to judge the constitutionality of panels that may be functionally inferior to the common law jury.<sup>50</sup> Once this analysis is complete, we shall be in a position to evaluate the statistical models cited by Justices Blackmun and Stevens for the proposition that juries in the size range of six to eight are "optimal."<sup>51</sup>

### A. Representativeness

The extent to which jury size influences the representativeness of juries has been carefully analyzed by several commentators who have argued forcefully that the differences between twelve- and six-member juries are likely to be substantial.<sup>52</sup> It is not difficult to extend these results to show that the incremental loss in representativeness due to the reduction from six to five is so small that if, under the no-functional-difference test, the reduction to five creates discernible functional differences, then so does the change to six. Consequently, the point at which the Court should be concerned with the effect of size reduction clearly occurs in some range well above the line between six and five. On the other hand, if the concept-conception standard is the applicable test, the impact of these results is less clear. The problem is not that the analysis shows no clear jump in the representativeness of twelve- as opposed to five- or six-person juries, but rather that under the concept-conception standard the question of what falls within the constitutional concept of a jury is not an empirical question, but a more complex value judgment.<sup>53</sup>

Statistical models of the jury selection process almost invariably assume that jurors are selected independently and at random.<sup>54</sup> Given this assumption,<sup>55</sup> it is easy to show that members of larger juries are more likely to be broadly representative of the population from which

---

50. See text accompanying notes 110-11 *infra*.

51. See note 17 and accompanying text *supra*.

52. E.g., Lempert, *supra* note 12; Zeisel, *supra* note 5.

53. See Note, *supra* note 5; Lermack, *supra* note 13, at 972.

54. E.g., M. FINKELSTEIN, *QUANTITATIVE METHODS IN LAW* (1978); Penrod & Hastie, *supra* note 13; Fabian, *On the Effect of Jury Size*, 72 J. AM. STAT. A. 535 (1977); Saks & Ostrom, *Jury Size and Consensus Requirements*, 1 J. CONTEMP. L. 163 (1975); authorities cited notes 57-58 *infra*. The Supreme Court has used such models in cases alleging racial discrimination in grand jury selection. E.g., *Castaneda v. Partida*, 430 U.S. 482, 485-99 (1977). See also de Cani, *Statistical Evidence in Jury Discrimination Cases*, 65 J. CRIM. L. 235 (1974).

55. The actual selection process is by no means completely random. The arguably non-random elements are exclusions by the judge for cause or hardship and peremptory challenges by attorneys. Peremptory challenges impose the most serious limitation on the statistical model, but neither this factor nor judicial exclusions invalidate the model. Indeed, the statistical predictions correspond reasonably well with the observations reported by the Columbia Jury Study Project. See Padawer-Singer, *supra* note 11. Nevertheless, the assumption of randomness suggests that the

juries are drawn, and that the average characteristics of the larger juries are more likely to approach the average characteristics of the population.<sup>56</sup> In particular, one can calculate the extent to which representativeness, suitably defined, is lost as jury size diminishes. As a measure of representativeness, one can look to the probability that no jury members will have a certain characteristic found among a specified percentage of the population.<sup>57</sup> To see how the relevant probabilities can be found, suppose that ten percent of the population from which jurors are drawn are Hispanic.<sup>58</sup> The probability that the first juror drawn will be Hispanic is then ten percent, or .10. The probability that this juror will not be Hispanic is simply  $1 - .10 = .90$ . If the population from which the jurors are drawn is large, then the probability that the second juror will not be Hispanic is also .90. The probability that both the first *and* the second will not be Hispanic is  $(.90)(.90) = .81$ . By an extension of this logic, the probability that five jurors in a row will not be Hispanic is  $(.90)^5 = .59$ , and the probability that all six persons on a six-member jury will not be Hispanic is  $(.90)^6 = .53$ . In other words, out of a hundred juries of six members each, one would expect to find fifty-three in which Hispanics are not represented. For the same number of five-member juries, the expected number<sup>59</sup> devoid of Hispanics increases by six, to fifty-nine. In contrast, the expected number of juries with no Hispanic jurors in a group of one hundred randomly selected twelve-member juries is twenty-eight.

In short, this example illustrates that the effect on representativeness of size changes in the vicinity of six-member juries is small in comparison to the loss suffered in going from the twelve- to the six-person jury. By the time one reaches the six-member jury, an average of twenty-five out of every one hundred juries are already less representative than twelve-member juries. The observation that the line must be drawn somewhere does not support fixing it in the region where, according to our example, the representative quality of a full one-quarter

---

analysis presented here should be understood as a loose estimate of the magnitude of the effect of jury size on representativeness. See Lempert, *supra* note 12, at 665-66.

56. Lempert, *supra* note 12, at 668.

57. Lempert discusses the effect of jury size on this quantity, although he does not explicitly use it as a measure of representativeness. *Id.* at 668-73. See also Zeisel, *supra* note 5, at 416-21.

58. Obviously, ethnic status is used only as an illustration. We could speak, with equal facility, of the probability that no juror will be predisposed to favor a police officer's testimony over that of an accused's, that no juror will be a corporate executive, that no juror will be a man, that no juror will have anti-authoritarian attitudes, and so on. Much of the value of the jury system inheres in its ability to bring persons with conflicting biases (and varied expertise) together to deliberate as a group. See generally Lempert, *supra* note 31.

59. According to a corollary of the statistician's "law of large numbers," (better known as the law of averages) the actual number approaches this "expected number" as the number of cases approaches infinity. See, e.g., H. BRUNK, AN INTRODUCTION TO MATHEMATICAL STATISTICS 153 (3d ed. 1975) (Corollary C).

of all juries is seriously affected. The more comprehensive calculations tabulated and graphed in the Appendix reveal that for rare and moderately rare characteristics—those found in less than one-quarter of the population—substantial losses in the expected frequency with which these characteristics are manifested on juries are incurred well before jury size is reduced to five or six. If the debate is about the presence or absence of functional differences, the “zone of reasonableness”—the region in which it is fair to say that some arbitrariness in drawing the line is unavoidable—seems to lie in a much higher range of jury sizes.

### B. Accuracy

The picture with regard to the likely accuracy of jury verdicts as a function of jury size is harder to discern, but it has the same general outline. The complicating factor is, of course, jury deliberations. Because jury deliberations, as well as jury selection, frequently influence verdicts, estimating the effect of size reductions on accuracy demands some sensitivity to the nature of the interactions among jurors with different outlooks, insights, recollections, personalities, and prejudices. No precise mathematical model now exists that can relate meaningfully the likelihood of a correct decision to the number of jurors participating in the process.<sup>60</sup> Nevertheless, social-psychological research into group decisionmaking and problem solving has isolated two related factors—size and heterogeneity—that suggest that smaller juries are more prone to err than are larger bodies.

First, this research has tended to show that “both quality of group performance and group productivity [are] positively correlated with group size.”<sup>61</sup> For example, larger groups do better than smaller ones

60. Some authors have attempted to compute the probabilities that different sized juries will convict (or acquit) a defendant when the defendant is actually innocent (or guilty). Gelfand & Solomon, *A Study of Poisson's Models for Jury Verdicts in Criminal and Civil Trials*, 68 J. AM. STAT. A. 271 (1973); Gelfand & Solomon, *Modeling Jury Verdicts in the American Legal System*, 69 J. AM. STAT. A. 32 (1974); Gelfand & Solomon, *Analyzing the Decision-making Process of the American Jury*, 70 J. AM. STAT. A. 305 (1975); Gelfand & Solomon, *An Argument in Favor of 12-Member Juries*, in MODELING THE CRIMINAL JUSTICE SYSTEM 205 (S. Nagel ed. 1977); Gelfand & Solomon, *Considerations in Building Jury Behavior Models: An Argument in Favor of the Twelve-Member Jury*, 17 JURIMETRICS J. 292 (1977) [hereinafter cited as *Considerations*]; Grofman, *Not Necessarily Twelve and Not Necessarily Unanimous: Evaluating the Impact of Williams v. Florida and Johnson v. Louisiana*, in PSYCHOLOGY AND THE LAW 149 (G. Bermant, C. Nemeth & N. Vidmar eds. 1976); Grofman, *Jury Decision-Making Models*, in MODELING THE CRIMINAL JUSTICE SYSTEM, 191, 196 (S. Nagel ed. 1977); Nagel & Neef, *supra* note 17. Such computations should not be taken too seriously, since they are based on simplistic assumptions or parameters whose values drastically affect the calculated probabilities of these two types of errors. See text accompanying notes 93-106 *infra*. In the next section, I also show that this approach, even if valid, does not support the result in *Ballew*. In estimating the effect of size changes on accuracy, however, I adopt the analysis presented in Lempert, *supra* note 12, and do not rely on the more elaborate—and more questionable—models of verdict accuracy.

61. Thomas & Fink, *Effects of Group Size*, 60 PSYCH. BULL. 371, 373 (1963). Thomas and

where crucial inputs are needed, apparently because different members may make critical contributions at different stages.<sup>62</sup> In jury deliberations, the crucial inputs may involve recollection of testimony, recognition of inconsistencies, or comprehension of the court's instructions.

Second, the literature supports the common sense notion that up to a certain point, the more viewpoints, types of expertise, and mutually cancelling biases that are brought to bear on the task of reconstructing events and coming to a decision, the more accurate the decision will be.<sup>63</sup> Heterogeneity seems particularly important where strongly held preconceptions or prejudices are at work,<sup>64</sup> and heterogeneous groups have been found to solve more problems correctly than homogenous groups of equal size.<sup>65</sup>

These two effects—the enhanced availability of crucial inputs and heterogeneity—should coalesce in the jury setting to make larger juries more accurate factfinding instruments than smaller ones. This inference seems confirmed by the Columbia University Jury Study, which recently found from an experimental study of ninety-two twelve- versus six-person juries hearing the same videotaped case, that among the smaller juries, “a smaller percentage of statements indicating inaccurate recall were corrected,” and that the smaller juries “displayed fewer resources among their members, and looked for guidance in the form of instruction from outsiders more often.”<sup>66</sup>

Focusing on the factor of heterogeneity, one can estimate roughly the likely impact jury-size reductions can have on accuracy. In particular, it can be argued that the measure of representativeness previously considered also serves as a crude indicator of probable jury accuracy.<sup>67</sup>

---

Fink qualify this conclusion by stating that this correlation is present “under some conditions,” but they add that “under no conditions were smaller groups superior.” *Id.* For a more careful review of the group decisionmaking studies, see Lempert, *supra* note 12, at 684-89. See also *Ballew v. Georgia*, 435 U.S. at 232-33 (separate opinion of Blackmun, J.) (adopting Lempert's analysis of the social-psychological studies and their relevance to the jury-size question).

62. See Faust, *Group Versus Individual Problem-Solving*, 59 J. ABNORMAL & SOC. PSYCH. 68 (1959); Kelley & Thibaut, *Group Problem Solving*, in 4 HANDBOOK OF SOCIAL PSYCHOLOGY 64, 69 (2d ed. G. Lindzey & E. Aronson 1969); Lempert, *supra* note 12, at 686-87.

63. See Lempert, *supra* note 12, at 685.

64. For a striking demonstration of this point, see Barnlund, *A Comparative Study of Individual, Majority, and Group Judgment*, 58 J. ABNORMAL & SOC. PSYCH. 55 (1959).

65. See authorities cited in Lempert, *supra* note 12, at 687 n.127.

66. Padawer-Singer, *supra* note 11, at 54.

67. This view is developed more fully in Lempert, *supra* note 12, at 670. It may be worth emphasizing that these estimates pertain to relative rather than absolute values; they do not translate readily into statements about the actual number of mistakes that can be expected with differently sized juries. Because the calculations look solely to the “heterogeneity effect” (and ignore the “crucial input effect”), they may understate the effect of size reductions. On the other hand, many litigated cases have fairly clear outcomes. Kalven and Zeisel, for example, found that judges expressed disagreement with the verdict returned by the jury in less than 30 percent of the criminal cases sampled. H. KALVEN & H. ZEISEL, *THE AMERICAN JURY* 109 (Table 23) (1966).

With various minority viewpoints at work, individual prejudices are more likely to cancel out, and individuals with valuable specialized knowledge or particularly astute insights are more likely to be represented.<sup>68</sup> For example, the complete absence of black jurors could interfere with accuracy in a criminal case involving a black defendant. The involvement of at least one black juror might give other jurors important information, such as the interpretation of argot.<sup>69</sup> In addition, the very presence of a single black might inhibit other jurors from voicing prejudices that, if expressed, could influence the deliberative process. Thus, a black's presence might inhibit a white juror arguing for conviction from making a remark to the effect that the defendant, being black, has probably committed some crime, if not the one with which he is charged.<sup>70</sup>

The representation of other minorities—not merely racial or ethnic—can have similar salutary effects on jury decisionmaking. In sum, to the extent that juries devoid of certain important but not ubiquitous characteristics are likely to be less able factfinding and decisionmaking bodies,<sup>71</sup> the representativeness analysis also reveals that the line-drawing argument with respect to accuracy is more appropriate when considering the relative merits of, for example, ten- or eleven-member juries. Where the issue is whether differences are discernible, the metaphor is not persuasive by the time one has descended into the region of the six-person jury.

### C. Reliability

For juries randomly selected from a heterogeneous population, reliability is also closely connected with size. To see how large the difference in reliability as a function of jury size can be, imagine that a damage action in which liability is not contested is somehow tried to all members of a community. Some people would feel large damages would be appropriate, many would opt for moderate damages, and

---

*See also* Lempert, *supra* note 12, at 648-53. Consequently, in most cases even a "better" heterogeneous panel could not be expected to return a different verdict than a more homogenous group. For this reason, the quantitative effect of jury size on heterogeneity, which is used here as an indirect measure of accuracy, will overstate the absolute incidence of incorrect verdicts.

68. Lempert, *supra* note 31, at 10.

69. For other examples of such expertise, see Broeder, *The Negro in Court*, 1965 DUKE L.J. 19, 30, and Lempert, *supra* note 12, at 670.

70. Lempert, *supra* note 12, at 670. Such a remark might influence potential holdouts for acquittal, "particularly if made at the end of an exhausting deliberation process when a minority was looking for reasons to renounce their earlier commitments." *Id.*

71. Of course, a smaller jury is also more likely to be devoid of unreasonable or bigoted individuals, but there are good reasons to think such prospects pose a less serious problem than might first appear. *See id.* at 672-73.

others would be inclined to give only a small damage award.<sup>72</sup> More specifically, suppose that a graph of the number of persons who would vote for each possible dollar award forms the familiar bell-shaped curve known to statisticians as a normal, or Gaussian, distribution, whose mean, or average, is the dollar amount at which the curve reaches its peak.<sup>73</sup> The conventional statistical measure of how much the damage estimates vary from person to person is called the standard deviation. If, as seems plausible, damage estimates are roughly "normally distributed,"<sup>74</sup> then approximately two-thirds of the community would vote for damages lying within one standard deviation to either side of the mean. For example, for a mean of \$3500 and a standard deviation of \$1000, only one-third of the community would favor awarding more than \$4500 or less than \$2500 to the plaintiff, and the average figure would be \$3500.

Now imagine that the case is heard instead by two juries of the same size, each unaware of the other. To say that jury decisions are reliable is to assert that the two damage awards returned are likely to be very similar. Indeed, if the system of ascertaining damages were perfectly reliable, then the difference in the awards would be zero. If one assumes that the members of each jury compromise their disagreements by averaging the figures initially suggested by every member of the jury,<sup>75</sup> one can calculate the maximum amount by which the

---

72. I know of no systematic data on how often civil jurors initially split on the issue of damages. However, it is known that about two-thirds of all juries in criminal cases find their vote split on the first ballot. H. KALVEN & H. ZEISEL, *supra* note 67, at 487. The supposition that the community would be divided thus seems quite realistic.

73. To portray the differences between six- and twelve-member juries, Ziesel, *supra* note 5, at 716-18, and Lempert, *supra* note 12, at 680-81, use a similar hypothetical. However, they posit a flat or uniform discrete distribution in which one-sixth of the community favors \$1000, another sixth favors \$2000, etc. Calculations based on such a flat distribution may overstate the effect of jury size on representativeness and reliability. It seems more likely that any actual distribution would peak somewhere in the intermediate range of preferred damage awards. The normal curve used here has this feature. To the extent that initial damage estimates are not consistent with a normal curve, other models must be employed. See note 77 and accompanying text *infra*.

74. It happens that for many continuously quantifiable characteristics of human and other populations, a display of the number of individuals plotted against the quantity measuring the characteristic has the shape of the "normal" curve. The use of the term "normal" to describe such frequency distributions is a historical artifact. It is not "abnormal" to find characteristics whose frequency distributions are not consistent with the shape of the "normal," or Gaussian curve.

75. The quotient verdict is forbidden, but not unheard of. More often, an "averaging" process seems to be employed, but it is not precise and does not weight every juror's opinion equally. With larger juries, more jurors may effectively be excluded, so that computations based on the simple averaging assumption may overstate the reliability problem somewhat. See Lempert, *supra* note 12, at 681 n.108.

Of course, these calculations are generally inappropriate for measuring the reliability of decisions returned by criminal juries. The decisions to be reached by criminal juries generally do not allow for an averaging process, except to the extent that the jury can bargain over the number of counts or a recommended sentence. Since the range of options open to a criminal jury is far more

awards of the two equal-sized juries would differ in, for example, 95 percent of the instances in which our two-jury experiment involving the same case was repeated.<sup>76</sup> For twelve-member juries, the two awards would differ by no more than \$800 in 95 percent of these instances. With eleven-member juries, the maximum difference at this "confidence level" of 95 percent is \$836. By the time the juries are down to size six, maximum variations of \$1132 would be expected. The further reduction to five-member juries produces the figure of \$1240. Thus, the difference in reliability between the six- and the five-member juries is notable, but it is dwarfed by the effect experienced in the transition down to six. This latter loss is three times the size of the former. Similar results follow even where the opinions in the community at large are not normally distributed.<sup>77</sup> Once again, if the mere presence or ab-

restricted than with a civil jury awarding damages, much more complex statistical models would have to be employed to measure reliability. However, because reliability is closely connected with representativeness, smaller criminal juries can also be said to be less reliable than larger ones. See note 39 and text accompanying notes 52-59 *supra*.

76. The means derived from samples of size  $n$  drawn from a normal population of mean  $\mu$  and standard deviation  $\sigma$  are normally distributed with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ . The standard deviation of the differences between the means of two independent samples of sizes  $n_1$  and  $n_2$  (also called the standard error) is given by

$$\sigma_{\bar{d}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

With samples in the form of two equally sized juries,  $n_1 = n_2 = n$ , so that the standard error reduces to

$$\sigma_{\bar{d}} = \sqrt{\frac{2\sigma^2}{n}}$$

Standard errors calculated according to this equation are listed in Table 3 of the Appendix.

Given the assumption of a normal population, the difference in the sample means is also normally distributed about a mean difference of zero. Since 95 percent of the area of the normal curve lies within 1.96 standard deviations of the mean, at the .95 confidence level the maximum difference in the jury means is given by

$$D_{95} = 1.96\sigma_{\bar{d}}$$

This quantity is also shown in Table 3.

The 95 percent confidence level is, of course, an arbitrary one. Changing the confidence figure obviously changes the number of standard deviations used in the equation for the confidence interval.

77. The exact confidence intervals will depend on the precise shape of the distribution in the community from which the jurors are drawn. As indicated in the text, such confidence intervals are easily calculated in the case of a normal distribution. For other distributions, the procedure is more involved. See J. PATEL, C. KAPADIA & D. OWEN, *HANDBOOK OF STATISTICAL DISTRIBUTIONS* 175-95 (1976). It is not necessary, however, to calculate confidence intervals as a function of sample size to indicate that the differences in the reliability of verdicts of six- and five-member juries are small compared to the disparity between six- and twelve-person juries. It can be shown that whenever a sample of size  $n$  is drawn at random from a population whose mean is  $\mu$  and whose standard deviation is  $\sigma$ , the sample mean will have a distribution whose mean is also  $\mu$  and

sence of discernible functional differences is determinative, the need to draw the constitutional line somewhere hardly justifies drawing it at, or even near, the point the Court has selected.

### III

#### THE OPTIMAL JURY SIZE

The sort of statistical modeling I have described is, of course, suggestive rather than definitive. Nevertheless, the indications it provides as to the likely impact of jury size on representativeness, accuracy, and reliability are clear and consistent. The statistical reasoning suggests that every one of these crucial values associated with the twelve-member jury is likely to be compromised significantly as the jury is diminished in size to six members. To the extent that the relevant standard is the existence *vel non* of functional differences, attempts to draw the line at six, as opposed to five, seem misdirected.

Yet, the Supreme Court was painfully aware of this type of analysis when it tenaciously clung to *Williams* in deciding *Ballew* and *Burch*. Perhaps the Powell group in *Ballew* and the majority in *Burch* scrupulously avoided references to the statistical studies because these Justices realized all too well that such studies seriously undermined the Court's opinions in *Williams* and *Colgrove*.<sup>78</sup> Or these Justices may have implicitly promulgated a new standard grounded in the concept-conception distinction. This possibility and its ramifications are discussed in Part IV.

But what of Justice Blackmun's opinion in *Ballew*? As previously noted,<sup>79</sup> Justices Blackmun and Stevens canvassed the literature in meticulous detail. Surely they had some reason to think the studies they cited supported rather than undercut their position. In one sense, these Justices were correct. As we have seen, there is likely to be some degradation in jury functioning when size is reduced from six to five.<sup>80</sup> Thus,

---

whose standard deviation is  $\sigma/\sqrt{n}$ . This result holds for any distribution, not merely those that are normal.

As mentioned in the text, the statistic that is commonly used to quantify dispersion or variability of a set of measurements is the standard deviation (or the standard deviation squared, which is called the variance). The awards that would be generated by simple averaging in the case of twelve-member juries would have a standard deviation of  $\sigma/\sqrt{n} = 1000/\sqrt{12} = \$289$ . The awards of six-person juries would be 1.4 times as variable, having a standard deviation of  $1000/\sqrt{6} = \$408$ . The awards of five-member juries would fluctuate with a standard deviation of  $1000/\sqrt{5} = \$447$ , which is one and one-half times as large as the standard deviation for twelve-member juries, but is not much different from the figure for six-person juries.

78. The American Civil Liberties Union, as *amicus curiae*, stated as much in its brief in *Burch*. See note 21 *supra*.

79. See text accompanying notes 16-17 *supra*.

80. See text accompanying notes 56-77 *supra*.

it is true that "the assembled data raise substantial doubt about the reliability and appropriate representation of panels smaller than six."<sup>81</sup> But these same "data" also create grave doubt about the proper functioning of the six-member jury. Consequently, it is hard to see how the statistical literature permits Justices Blackmun and Stevens to "understand the basis for the conclusion of Mr. Justice Powell that 'a line has to be drawn somewhere [in the vicinity of the six-member jury]'"<sup>82</sup>

However, the Blackmun opinion cannot so readily be labelled a case of judicial reluctance to follow simple, statistical logic to its ultimate conclusion. One paragraph in the opinion argues that juries of six to eight members may in fact be preferable to juries of all other sizes, at least with regard to accurate decisionmaking. If this claim is correct, then there is at least an arguable basis for the opinion's insistence on "[reaffirming] our holding in *Williams*."<sup>83</sup> Because this conclusion is so remarkable and the language in which it is cast is so opaque to persons not schooled in statistical terminology, it is worth setting out in its entirety:

[T]he data now raise substantial doubts about the accuracy of the results achieved by smaller and smaller panels. Statistical studies suggest that the risk of convicting an innocent person (Type I error) rises as the size of the jury diminishes. Because the risk of not convicting a guilty person (Type II error) increases with the size of the panel, an optimal jury size can be selected as a function of the interaction between these two risks. [One study] concluded that the optimal size, for the purpose of minimizing errors, should vary with the importance attached to the two types of mistakes. After weighting Type I error as 10 times more significant than Type II, perhaps not an unreasonable assumption, [the researchers] concluded that the optimal jury size was between six and eight. As the size diminished to five and below, the weighted sum of the errors increased because of the enlarging risk of convicting innocent defendants.<sup>84</sup>

Research to date, however, does not justify the sweeping claim that juries of six to eight members will produce the best balance of erroneous convictions and false acquittals.<sup>85</sup> Indeed, the two studies cited in

---

81. 435 U.S. at 239 (separate opinion of Blackmun, J.).

82. *Id.* at 232 n.10 (separate opinion of Blackmun, J.).

83. *Id.* at 239 (separate opinion of Blackmun, J.). The argument would not be sufficient, however. Even if a jury of approximately six were more accurate than a jury of twelve, it should not be constitutional if it is substantially inferior in the dimensions of representativeness and reliability. See text accompanying notes 21-38 *supra*. Nothing in the Blackmun opinion denies that with respect to these values—values that the opinion itself holds are of constitutional significance—the smaller juries are likely to be inferior.

84. 435 U.S. at 235 (separate opinion of Blackmun, J.) (footnotes omitted).

85. I use the phrases "false convictions" and "false acquittals" because the meaning of these terms is more apparent than that of "Type I" and "Type II" error. On the other hand, the phrase

the footnotes to the passage above,<sup>86</sup> read in their entirety, make no such claim.

The first study is a three page paper which appeared in *The American Statistician*. The author, Herbert Friedman, makes no calculations of the probability of mistaken convictions or acquittals as a function of jury size. Instead, ignoring any possible interactions among jurors and assuming that the probability that a juror will consider a defendant guilty is the same for every juror, he computes the probability that twelve- and six-member unanimous and nonunanimous juries would find a defendant guilty.<sup>87</sup> For example, if, in a case presented to a six-member unanimous jury, the probability that each juror would consider the defendant guilty was .6, and if all jurors voted only once without discussion, the probability of conviction would be the chance that the first juror would vote to convict (.6) times the chance that the second would do so (.6), and so on for all six independent jurors. This simple model thus gives a probability of conviction for the jury as a whole of .6<sup>6</sup>, or roughly .05. Whatever such calculations may reveal,<sup>88</sup> they provide no quantitative estimates of the number of erroneous verdicts for differently sized juries.<sup>89</sup>

The second study was undertaken by two political scientists, Stuart Nagel and Marion Neef. It does purport to calculate the likelihood of false convictions and false acquittals, and Nagel and Neef deduce on this basis that seven-member juries will produce the best mix of these two types of errors.<sup>90</sup> However, this result rests on a series of question-

---

"false acquittal" is slightly misleading, since I mean to include all instances, including hung juries, in which the jury fails to convict when the defendant is guilty.

86. See note 17 *supra*. Justice Blackmun apparently cites the Friedman study only to support the plausible view that with more and more persons required to convict a defendant, the risk of a false acquittal rises while the risk of a false conviction falls. *Contra*, studies cited note 60 *supra*. I argue below that an accurate evaluation of the relative magnitude of these effects is not currently available.

87. Friedman did not consider the performance of juries of other sizes.

88. The author concludes that rather than modifying the traditional jury, the legal system should strive for "[b]etter data resulting from more reliable and clearcut evidence combined with speedy and fair trials." Friedman, *supra* note 17, at 23.

89. Friedman's conclusions as to the relative likelihood of the two types of errors are entirely qualitative and emphasize the serious differences between juries of twelve as opposed to six or five:

[W]hile the innocent-appearing defendant is unlikely to be convicted under any circumstances, the individual who appears to be nearly guilty or can present only a weak defense runs a much greater risk of conviction with the [nonunanimous] or smaller jury. . . . [T]hose with . . . poor legal help . . . would lose much of the protection of the "reasonable doubt" restriction if conviction were made easier. Therefore, relaxing the jury procedure . . . will serve to markedly increase the probability of convicting those who are actually innocent.

*Id.* at 22-23.

90. See Nagel & Neef, *supra* note 17, at 957. Justice Blackmun describes the research as selecting the range six to eight, and in fact the numbers computed for six- and eight-member juries are virtually identical with the figure provided for the seven-person jury. *Id.* at 956 (Table 4).

able assumptions and dubious procedures.

Nagel and Neef begin with the model described by Friedman. They too pretend that every juror has the same independent likelihood of voting to convict in every case. This uniform probability of a ballot to convict ( $p$ ) is deduced from Kalven and Zeisel's 1966 finding that about 64 percent of criminal juries return guilty verdicts.<sup>91</sup> Assuming that group deliberation has no effect on a juror's decision, Nagel and Neef solve the equation  $p^{12} = .64$  to obtain  $p = .96$ . Next, they use this number to derive the probability that a six-member jury will convict by computing  $p^6 = .80$ . They obtain the corresponding probabilities for other jury sizes in like manner.

Estimating the incidence of mistaken verdicts requires still another assumption, this one pertaining to the proportion of persons brought to trial who are actually guilty. With little or no evidence, Nagel and Neef suppose that exactly 95 percent of all defendants who stand trial are guilty. On these assumptions, it follows that for a six-person jury, 80 percent of the 95 percent (making 76 percent) of the defendants who are really guilty will be found guilty, leaving 19 percent of all defendants falsely acquitted. Likewise, 80 percent of the 5 percent of the defendants who are actually innocent also will be convicted, giving a false conviction rate of 4 percent. By calculating comparable figures for other sized juries, Nagel and Neef predict that changing the size of the jury will increase the incidence of one type of error while decreasing the incidence of the other. Finally, to arrive at a jury size that achieves the least offensive combination of the two types of errors, the researchers blithely postulate that a false conviction is worth exactly ten erroneous acquittals. They conclude that the seven-member jury minimizes this weighted sum of the errors.<sup>92</sup>

91. In their classic study, Kalven and Zeisel reported that 64.2 percent of 3,576 twelve-person juries found defendants guilty. H. KALVEN & H. ZEISEL, *supra* note 67, at 56.

92. In reaching this conclusion, Nagel and Neef modify their model to account for interaction among jurors. They argue that if "collective" effects were nonexistent, each juror would decide independently whether to convict and vote for conviction with the same probability. As we have seen, they find from Kalven and Zeisel's data that this probability  $p$  is  $(.64)^{1/12} = .96$ . At the other extreme, they argue that if "collective" effects were overwhelming, each juror would act in concert, and  $p$  would equal .64. Again referring to Kalven and Zeisel's data, Nagel and Neef reason that since 68 percent of all jurors voted for conviction on the final ballot (in the 225 cases for which voting records were available), this 68 percent figure can be taken as an empirically determined value for  $p$ . Since it is slightly higher than the  $p$  of .64 predicted by the "collective mind" model, and far lower than the .96 predicted by the "independent model," they conclude that the "independent probability propensities" of jurors account for only a small amount—11 percent—of jury behavior. Nagel & Neef, *supra* note 17, at 952. Predictions of error rates are then made using a composite model in which the  $p$  of .64 rather than the  $p$  of .96 discussed in the text is dominant.

This treatment of juror interaction is criticized in note 93 *infra*. See also Penrod & Hastie, *supra* note 13, at 484-85. The important thing to notice at this point is that whether or not it is valid, this refinement can have no effect whatever on the determination of the optimal jury size.

That Justices Blackmun and Stevens should give credence and legal effect to this sort of model is disappointing. Every step in the analysis invites challenge. First, neither all jurors nor all cases are identical. Second, it would be an extraordinary coincidence if the proportion of truly guilty defendants were exactly 95 percent. Third, when the effects of deliberation are considered,<sup>93</sup> the probability that each juror will vote to convict may itself depend on jury size.<sup>94</sup> Finally, the premise that ten guilty persons—no more and no less—should go free to prevent the acquittal of one innocent person is better rhetoric than mathematics.<sup>95</sup>

Of course, these assumptions might not be so troublesome if large departures from them would produce only small differences in the predicted error rates and the resulting "optimum" jury size. But even slight variations in the assumptions generate dramatically different conclusions. When it is recognized that the strength of the evidence against defendants varies from case to case, a jury of twelve can actually emerge as optimal.<sup>96</sup> When a figure of .90 for the proportion of guilty defendants is substituted for the .95 figure, the optimal jury size skyrockets to twenty-three.<sup>97</sup> As for the "trade-off weight" of ten false

---

To account for the effects of deliberation by throwing in a pure "collective mind" term that, by definition, does not depend on jury size is to ensure that the oversimplified "independent mind" calculations actually determine the optimum jury size. This is so because the "optimal" point is found by taking the derivative of the weighed sum of the expected errors. A constant term (whose derivative is necessarily zero) contributes nothing to the equation which yields the "optimal" jury size. Thus, Nagel and Neef recognize that their treatment of deliberative effects does not influence their finding that a jury of seven is "optimal." See Nagel & Neef, *supra* note 17, at 954.

93. Nagel and Neef treat the deliberative process in the remarkably simple way outlined in note 92 *supra*. Among other things, the 11 percent "collective mind" calculation presupposes that the predictions of the "independent" and "collective mind" models should be combined linearly with an additive constant of zero. There is, however, no particular reason to think that the predictions should not be combined according to any one of a literally infinite variety of nonlinear formulae. With only one data point (the observed values of .642 and .677), it is impossible to test the accuracy of Nagel and Neef's ad hoc model. Nagel and Neef are aware of this questionable feature of their model, for they frankly admit that "there is nothing in what we know about the true situation that dictates this form of expression." Nagel & Neef, *supra* note 17, at 951.

94. See text accompanying notes 60-70 *supra*.

95. Even as legal rhetoric the attempt at quantification can be troublesome. Judicial approval of any statement of the exact number of innocent persons who should be sacrificed to permit the conviction of a specified number of guilty persons—something which is implicit in the trade-off weight—may well be inadvisable. Such things, it can be argued, are best left unsaid. See G. CALABRESI & P. BOBBITT, *supra* note 31; Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1199 n.27 (1979); Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971); Underwood, *Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individual Judgment*, 88 YALE L.J. 1408, 1428 n.54 (1979).

96. See Penrod & Hastie, *supra* note 13, at 486.

97. Nagel & Neef, *supra* note 17, at 967. Nagel and Neef candidly acknowledge that their unverified estimate of the proportion of guilty defendants is an "unstabilizing element" in their model.

acquittals for every false conviction, this number was admittedly chosen in part to conform to the range of jury sizes the Supreme Court had previously held constitutional.<sup>98</sup> Justice Blackmun's reliance on the study to support the Court's prior holdings and to reject an extension of them is, therefore, distressingly circular. Had Justice Blackmun wished to depict a five-person jury as "optimal," he need merely have inserted a trade-off weight of 9.4 into the model. Had he desired to overturn *Williams* and argue for a reversion to the twelve-member jury, he need only have used the "perhaps not unreasonable" premise that thirteen rather than ten guilty persons should be freed to save one innocent person from conviction.

What makes this all the more disturbing is that these matters are not obscured by the authors of the study. On the contrary, they make it plain that the optimal jury size their model generates is frightfully sensitive to changes in the arbitrarily determined trade-off weight and the ad hoc estimate of the proportion of guilty defendants.<sup>99</sup> They are careful to point out that in some respects their model is crude, that the magnitude of the effects they predict depends on "untested premises," and that their calculations are merely "illustrative . . . , designed mainly to demonstrate the methodology of the model."<sup>100</sup> Accordingly, the study cannot fairly be understood as seriously claiming, let alone proving, that juries in the range of six to eight members are preferable to larger or smaller ones.

Thus, none of the studies of jury accuracy that Justice Blackmun cites provides a satisfactory basis for distinguishing between six- and five-member juries. In addition, the Blackmun opinion fails even to mention another, more sophisticated model of jury decisionmaking that contradicts the suggestion that six- but not five-member juries fall within the "optimal" range. This model was first developed and used by the renowned mathematician Siméon Poisson to study the operation of the French jury system in the second quarter of the nineteenth century. In this century, it has been refined and extended by contemporary statisticians.<sup>101</sup> The model is remarkable in that it permits calculations

---

98. According to Nagel and Neef, "[s]ince a twelve-person jury implies a trade-off weight of 13, our tentative assumption of ten . . . is probably a reasonable one." Nagel & Neef, *supra* note 17, at 965. Moreover, they report, "Experimenting with different trade-off weights ( $W$ ) shows that when  $W$  is less than 9.5, the optimum jury size falls below six, which is the smallest jury the Supreme Court has permitted in criminal cases; when  $W$  is greater than 13, the optimum jury size is larger than twelve, which is probably the maximum jury size that is politically feasible. This experimentation both confirms the reasonability of our choice of  $W=10$  and demonstrates that other choices would not greatly change the outcome of our calculations." *Id.* The final point seems puzzling.

99. See notes 97-98 *supra*.

100. Nagel & Neef, *supra* note 17, at 975, 978.

101. Poisson's model is described, extended, and defended in a series of papers by Gelfand

of the probabilities of false convictions and false acquittals without requiring any assumptions about the ratio of innocent to guilty defendants.<sup>102</sup> Unlike the more simplistic models relied on by Justices Blackmun and Stevens, the Poisson model suggests that both six- and five-member juries can be expected to convict more innocent persons and to acquit more guilty ones than twelve-member juries.<sup>103</sup>

It is doubtful, however, that these calculations provide any hard information about the relative accuracy of juries of various sizes. Even with the improvements suggested by modern statisticians, Poisson's analysis seems incapable of predicting error rates with tolerable precision as long as it depends on the existing data on jury voting patterns. As currently employed, the Poisson model suffers from many of the defects that plague the mathematically less elegant models. Like the cruder models mentioned by Justice Blackmun, it treats all jurors as identical, and it regards all defendants and all offenses as fungible.<sup>104</sup>

---

and Solomon. See note 60 *supra*. A political scientist, Bernard Grofman, has also constructed a more elaborate version of the Poisson model. See *id.*

102. The details of Poisson's solution to the problem are presented in the articles by Gelfand and Solomon, *supra* note 60. A simplified version is outlined in the Appendix.

Penrod and Hastie, *supra* note 13, describe some other techniques as producing calculations of "juror accuracy" and "the error rates in jury verdicts." However, these techniques—which include a model developed by Penrod and Hastie themselves—not only suffer from the defects enumerated at text accompanying notes 94 *supra* & 104-06 *infra*, but more fundamentally, their structure precludes any estimate of errors as defined at text accompanying note 43 *supra*. At best, these models generate estimates of the probability that a jury of any given size will reach a verdict that departs from that which would be returned if the case were tried to the entire community. They are, therefore, closer to estimates of reliability than of accuracy.

103. See Appendix (Table 5); Fabian, *On the Effect of Jury Size*, 72 J. AM. STAT. A. 535 (1977); Gelfand & Solomon, *Rejoinder*, 72 J. AM. STAT. A. 536 (1977); Gelfand & Solomon, *Considerations*, *supra* note 60, at 311 (Table 10).

104. This results from the use of some mean value of the crucial variables to characterize each juror. Nagel and Neef speak only of the "average juror" who has a single, numerical "propensity to convict." In contrast, the Poisson model is more subtle, and uses two variables to calculate the probability that  $i$  jurors will vote to acquit and the remaining  $n-i$  will vote to convict on the first ballot. The variable that pertains directly to the jurors is  $\mu$ , the mean probability that a juror will vote correctly on the first ballot. This probability that a juror will vote correctly is assumed not to vary from juror to juror, from crime to crime, or from case to case. Gelfand and Solomon recognize that such an assumption is artificial, and in their "three parameter model" they allow  $\mu$  to assume two values,  $\mu_1$  and  $\mu_2$  (denoting the probability that a juror will not err if the defendant is guilty and the probability that he will not err if the defendant is innocent). See articles cited note 60 *supra*. It should be clear, however, that there are a host of other factors which should cause  $\mu$  (or, in the three parameter model,  $\mu_1$  and  $\mu_2$ ) to vary about a mean value or values. Cf. Penrod & Hastie, *supra* note 13, at 478-83 (outlining a model in which each juror will vote to convict with a probability  $p$  which takes on different values in different cases, depending on the evidence in each case).

The other variable in the Poisson model is  $\theta$ , the proportion of defendants who are guilty, or as Gelfand and Solomon prefer to say, the probability that the evidence will be "sufficiently convincing to establish a very strong probability of guilt." *A Study of Poisson's Models for Jury Verdicts in Criminal and Civil Trials*, *supra* note 60, at 271. Deductions of  $\theta$  based on French data segregated into two broad categories of crimes (those against the person and those against property) suggest that  $\theta$  depends on the nature of the crime. See *id.* at 278 (Table 5). Moreover, there

While the modern version of Poisson's model does handle the deliberative process more plausibly,<sup>105</sup> this facet of the model nevertheless looks only to juror interactions as averaged over all types of jurors and cases. Furthermore, it ignores the possibility that when there are deliberations, the probability that a particular juror will err will itself be a function of the size of the jury.<sup>106</sup>

It is thus clear that the statistical models of jury accuracy that Justice Blackmun relied on in *Ballew* are not adequate to permit the Court to extricate itself from the quagmire its handling of the jury size issue

---

is a possible feedback effect between  $\theta$  and  $\mu$ . If  $\theta$  is very close to one, as opposed to, for example, one-half, jurors may be more inclined to reason that "the defendant is guilty, or else he wouldn't be here." Where  $\theta$  really is high, this surmise may lead to fewer mistakes and thereby raise the value of  $\mu$ .

Using single values for the parameters in the models—which is about all that can reasonably be done with the limited data now available on juror voting patterns—creates two problems. First, because each parameter has some non-zero dispersion about its mean, it makes calculations based on the models inexact. Second, it may obscure potentially important differences between differently sized juries. For example, even if it could be accurately predicted that within the size range of 6 and 12, juries would return the same total number of convictions with the same numbers of false convictions and false acquittals, we could not know whether there would be differential effects according to the type of case, the quantum of the evidence, or the character of the defendant. That is, the statistical model could not (without information not now known and unlikely to be obtained in the near future) tell us whether six-member juries, for instance, are more or less prone to err than their twelve-member counterparts in handling cases involving politically unpopular defendants, sexual offenses, large corporate defendants, and all sorts of other matters. The use of "average" figures thus serves to mask what may be legally crucial differences.

105. Using data from the Kalven and Zeisel study and from a mock jury study, Gelfand and Solomon calculate the probabilities that a jury with a given distribution of votes on the first ballot will ultimately convict, acquit, or hang. This technique is reasonable when confined to making predictions about twelve-member juries. However, in making predictions about the conviction and error rates of smaller juries, Gelfand and Solomon assume that these transition probabilities do not change. For example, they assume that the probability that a twelve-member jury which initially votes 10 to 2 in favor of conviction will reach a unanimous verdict for conviction on the final ballot is the same as the probability that a six-member jury which initially divides five to one will ultimately convict. *E.g.*, Gelfand & Solomon, *Considerations, supra* note 60. In so doing, they resurrect the discredited "proportionality hypothesis" floated by the *Williams* Court. *See, e.g.*, Lempert, *supra* note 12, at 678-79. *But cf.* Grofman, *Jury Decision-Making Models, supra* note 60, at 192 (asserting that the proportionality thesis seems to hold); Klevorick & Rothschild, *A Model of the Jury Decision Process*, 8 J. LEGAL STUD. 141, 150 (1979) (asserting on the basis of intuition and an unpublished paper by Grofman that "this issue remains an open question").

106. This problem cannot be solved by interpreting  $\mu$  as the probability that a juror will err after discussion but before the first ballot. If the first ballot shows significant division of opinion, argumentation is likely to persist. Finally, a feedback effect can make  $\theta$ , the probability that a defendant brought to trial is actually guilty, depend on the jury size. Many attorneys believe that it is easier to convince a smaller number of jurors than to prevail before a larger panel. ROSCOE POUND—AMERICAN TRIAL LAWYERS FOUNDATION, *THE AMERICAN JURY SYSTEM* 20 (1977). If many prosecutors share this view, they might be more willing to proceed with less compelling evidence, and a lower value of  $\theta$  might come to characterize a regime of smaller juries. To the extent  $\mu$  and  $\theta$  depend on the size of the jury, taking values of  $\mu$  and  $\theta$  deduced from data on twelve-member juries to make predictions about smaller juries is bound to be wrong. Moreover, other deliberative effects which turn on jury size may also influence conviction probabilities. *See* text accompanying notes 60-71 *supra*.

has created. Neither the simple models cited in the Blackmun opinion nor the more subtle Poisson model overlooked by the Court justifies the view that juries of six are "optimal" as compared to those of five or twelve.

#### IV AND THEN THERE WERE TWELVE

As long as *Williams* is read as enunciating a no-functional-difference test, the efforts of various members of the Court to provide a principled explanation for condemning the five-member jury while simultaneously reaffirming *Williams* seem futile. The functional differences between the smaller juries seem overshadowed by the constitutionally significant differences between the smaller juries and the traditional twelve-member jury. Hence, fixing the line at the point marked by *Williams* and *Ballew*, and reaffirmed in *Burch*, has the appearance of expediency rather than principle. The likely functional inferiority of all smaller juries is not contradicted by the statistical predictions of false convictions and false acquittals briefly mentioned in Justice Blackmun's opinion in *Ballew*. While the use of statistical models to predict jury accuracy is intriguing, the current state of the art warns against placing much confidence in the numbers generated. To the extent the models of accuracy are suggestive, however, they support the view that the twelve-member jury is likely to be superior.

In short, the statistical arguments developed in Parts II and III establish that if the governing principle of *Williams* is that virtual functional equivalence is required, then it follows that the *Williams* Court grossly misapplied its own standard. Seen in this light, *Ballew's* and *Burch's* reaffirmances of *Williams* represent judicial intransigence—a willful disregard or cynical distortion of the writings of social scientists.<sup>107</sup>

Using the concept-conception test, however, it may be possible to harmonize or at least rationalize the jury-size cases. If the Court is seen as probing at the fuzzy edges of the contested concept of the jury, then the line-drawing appears less forced. *Ballew* is then perceived as clarifying the weight to be given the functional criteria in the concept-conception test "implicitly" set out in *Williams*<sup>108</sup> or, less charitably, as substituting *sub silentio* a much weaker standard for the absolute no-functional-difference test of *Williams*.<sup>109</sup> Under this view, the func-

---

107. See, e.g., Sperlich, *Trial by Jury: It May Have a Future*, *supra* note 17, at 221; Sperlich, *supra* note 13, at 275 (describing but not endorsing this view); Tanke & Tanke, *supra* note 17, at 1133.

108. See text accompanying notes 33-36 *supra*.

109. See Sperlich, *Trial by Jury: It May Have a Future*, *supra* note 17, at 218-21.

tional differences between juries of twelve members and those short of twelve may be real enough to violate a strict "no discernible difference" interpretation of *Williams*, but they are arguably not so enormous as to remove the slightly smaller juries from the general constitutional "concept" of a jury. Of course, it is not at all clear that the differences between twelve- and six-member juries are this trivial, but the point is surely debatable.

But even if juries of six can be said, under the "concept" approach, to lie within the zone of reasonableness where it is plausible to draw a line, they should nevertheless be rejected as unconstitutional. Ultimately, the effort to implement the essentially unstructured concept-conception distinction creates more serious problems than it solves.<sup>110</sup> No court can demarcate the intrinsically numerical contours of the nebulous concept of jury size without fashioning what will seem to be an arbitrary line, particularly if the statistical reasoning discussed above reflects reality. This line may be drawn between eleven and twelve, or between ten and eleven, or elsewhere, but wherever it might be fixed it will appear artificial. This inherent arbitrariness is compounded as one moves toward smaller juries that clearly are functionally inferior. Such arbitrariness in turn militates in favor of using the one criterion that does not seem completely arbitrary, and which does not tolerate incremental erosions of important values—the clear mandate of history.<sup>111</sup> In sum, whether the functional analysis mandated by *Williams* is understood to be tantamount to a strict no-functional-difference test, or whether it is viewed as merely defining the concept of the jury, the final conclusion is the same: any departure from the traditional conception of the "jury" as a body of twelve persons should be held unconstitutional.<sup>112</sup>

---

110. Of course, in other contexts, such as the determination of probable cause, lines that cannot be discovered by empirical or theoretical research are discerned by intuitive processes. However, the inaccuracy and arbitrariness in the application of these vague concepts are masked by the difficulty in comparing different factual situations. In contrast, for the jury-size issue, the result is determined once and for all, since the issue is ineluctably quantitative. Cf. Nesson, *supra* note 95 (costs of quantifying the burden of persuasion in criminal cases); Tribe, *supra* note 95 (same).

111. See note 24 *supra*. Of course, the size of the common law jury can itself be viewed as an arbitrary "historical accident." See text accompanying note 26 *supra*. Many constitutional provisions are, in the same sense, arbitrary. But this seems the sort of "second order" arbitrariness that must be accepted if there is to be any fidelity to the intent of the framers. It is not an argument for introducing the kind of "first order" arbitrariness apparent in *Ballew and Burch*.

112. Although further adherence to *Williams* is difficult to justify, there may be some reticence to overrule the case because of a concern that any disavowal of the opinion at this late date would have retroactive effect and release thousands of criminals who were convicted by smaller juries in states relying on the Court's opinion in *Williams*. Although there is ample precedent to suggest that a decision overruling *Williams* could be given purely prospective effect, see, e.g., *DeStefano v. Woods*, 392 U.S. 631, 634 (1968), the better view is that since the use of six-member juries jeopardizes the integrity of the fact-finding process, a decision to overrule *Williams* should apply retroactively. See, e.g., cases cited in Note, *A La Recherche du Temps Perdu: Retroactivity*

In entitling his telling critique of the majority opinion in *Williams* "And Then There Were None,"<sup>113</sup> Professor Zeisel implied that the Court, having started down the slope of smaller juries, would find no intellectually defensible stopping point. The opinions in *Ballew* and *Burch* testify to the prescience of this observation. By judicial fiat the Court has composed the current lyric "And Then There Were Six." It remains for the Court to reconsider the problem of jury size—freed from the *idée fixe* that the holding in *Williams* must be preserved. If this is done, it may be that the next decade's jury size opinions can be subtitled "And Then There Were Twelve."

---

*and the Exclusionary Rule*, 54 N.Y.U.L. REV. 84, 84 n.2 (1979). The opinions in *Brown v. Louisiana*, 100 S.Ct. 2214 (1980), making *Burch* at least partly retroactive (*see note 2 supra*), cannot be said to dispose of this issue.

Of course, the practical effect of a retroactively applicable decision could be mitigated by a liberal resort to the harmless constitutional error doctrine. Because many litigated cases are fairly clear, the number of convictions that might have been affected by differential jury size probably is a small fraction of all the cases heard by smaller juries. *See note 67 supra*; Lempert, *supra* note 12, at 653 (suggesting that only 14 percent of the cases that go to trial are such that the differences between twelve- and six-member juries are likely to be reflected in the outcome). Arguably, it should be possible to screen for those close cases in which size might have made a difference. Such cases might include those where the smaller jury reported itself at least once as deadlocked, where deliberations lasted more than a certain length of time, or where a reading of the trial transcript reveals closely balanced evidence or a situation in which jury prejudices might have been activated.

Yet, even this stratagem may be unsatisfactory. First, the potential administrative costs are enormous. Second, the scheme would entail a dangerous extension of the harmless error rule. In every Supreme Court case in which a constitutional error has been held to be "harmless," the error related to a discrete item of evidence, testimony, or argument. *See, e.g.*, *Chapman v. California*, 386 U.S. 18 (1967) (prosecutor's references to petitioner's failure to take the stand); *Harrington v. California*, 395 U.S. 250 (1969) (introduction of confessions of codefendants who did not take the stand); *Chambers v. Maroney*, 399 U.S. 42 (1970) (admission into evidence of ammunition seized in search of petitioner's home). *But see* *Coleman v. Alabama*, 399 U.S. 1 (1970) (case remanded for determination of whether the failure to provide indigent with appointment of counsel at preliminary hearing was harmless). The use of a constitutionally deficient petit jury, on the other hand, is a structural error of the sort that the Court has held cannot be overlooked as harmless even though no prejudice in a criminal trial can be articulated. *See* *Rose v. Mitchell*, 443 U.S. 545 (1979) (racial discrimination in selection of grand jury cannot be considered harmless even where a defendant has been found guilty beyond a reasonable doubt by a properly constituted petit jury at a trial on the merits that was free from other constitutional error). *See also* Note, *Harmless Constitutional Error: A Reappraisal*, 83 HARV. L. REV. 814 (1970).

113. Zeisel, *supra* note 5.

## APPENDIX

This appendix presents the detailed calculations intended to give a rough sense of the relative magnitude of the effects of jury size ( $n$ ) on the three facets of jury decisionmaking—representativeness, accuracy, and reliability—identified in Part I of this Article as constitutionally significant. It also presents some calculations of accuracy made on the basis of the Poisson model mentioned in Part III.

Tables 1 and 2, and Figure 1 pertain to representativeness and, indirectly, to accuracy.<sup>1</sup> The entries in Table 1 are found from the equation  $P_n = (1-f)^n$ , and the table is thus an expanded version of the one presented in Lempert.<sup>2</sup> The entries in Table 2 are calculated from the values given in Table 1. Within errors attributable to the process of rounding off, they should follow the equation  $-\Delta P_n = f(1-f)^{n-1}$  for  $n > 1$ . Figure 1 is a graph of this function  $-\Delta P$  for selected values of the parameter  $f$ .

These equations can readily be generalized to arrive at the total and incremental probabilities that  $i$  out of  $n$  jurors will have the characteristic found in a proportion  $f$  of the population from which juries are drawn. The probability of this event is given by  $P_n = \binom{n}{i}(1-f)^i f^{n-i}$ . The expression  $\binom{n}{i}$  stands for the number of combinations of  $i$  jurors that can be formed from  $n$  jurors. It can be computed from the formula  $\binom{n}{i} = \frac{n!}{(n-i)!i!}$  where, for any positive integer  $x$ ,  $x! = x(x-1)(x-2) \dots 1$ . The corresponding incremental probability function is  $-\Delta P_n = \frac{n}{n-i}(1-f)P_{n-1}$ .

Tables 3 and 4, and Figure 2 show the effect of jury size on reliability. Table 3 assumes a normal distribution of estimates of the amount of damages on the part of potential jurors. Table 4 is derived from Table 3, and it presents the marginal differences in reliability. Figure 2 is a graph of the marginal differences in the standard error of the damage estimates, as derived in Table 4.

Table 5 displays the probabilities given by the Poisson model for false convictions and false acquittals (more precisely, nonconvictions) for juries of selected sizes.<sup>3</sup> The maximum likelihood estimates for  $\theta$  and  $\mu$  are used. Following Table 5, a sketch of the basic features of the Poisson model is presented.

---

1. See Part II *supra*.

2. Lempert, *supra* note 12, at 669 (Table 1).

3. These probabilities are derived from Table 4B of Gelfand & Solomon, *Analyzing the Decisionmaking Process of the American Jury*, *supra* note 60, at 308.





FIGURE 1  
 DECREASE IN PROBABILITY THAT NO JURORS WILL  
 BE SELECTED WITH A CHARACTERISTIC SHARED  
 BY GIVEN PROPORTIONS OF THE POPULA-  
 TION FROM WHICH JURORS ARE  
 DRAWN AS A FUNCTION OF DE-  
 CREATING JURY SIZE

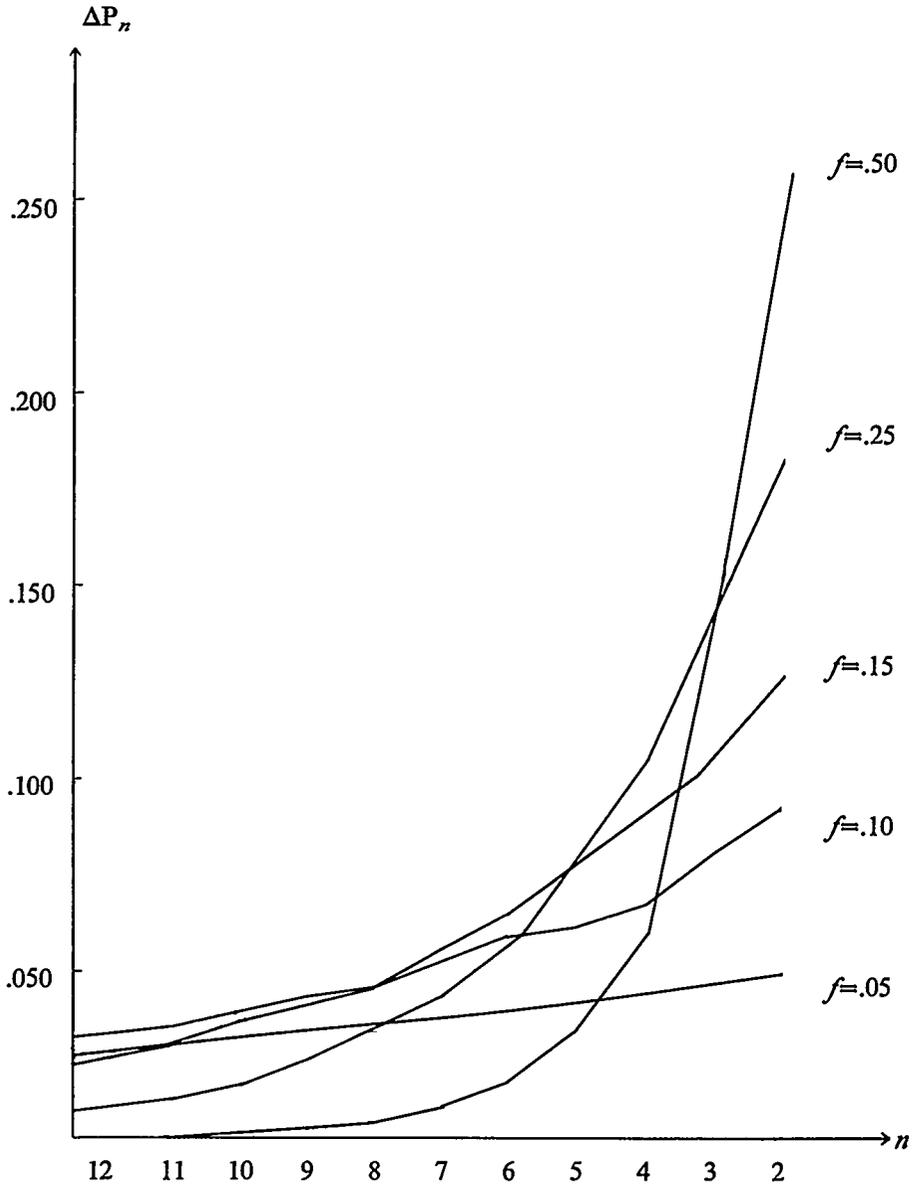


TABLE 3

STANDARD DEVIATION ( $\sigma_{\bar{d}}$ ) OF THE DIFFERENCE IN THE MEANS OF DAMAGE ESTIMATES (*i.e.*, STANDARD ERROR) FOR TWO JURIES OF THE SAME SIZE ( $n$ ) HEARING THE SAME CASE AND SELECTED AT RANDOM FROM A NORMAL POPULATION OF STANDARD DEVIATION  $\sigma = 1000$

Jury Size ( $n$ )	Standard Deviation ( $\sigma_{\bar{d}} = \sqrt{2\sigma^2/n}$ )	Maximum Difference at 95% Confidence Level ( $D_{95} = 1.96\sigma_{\bar{d}}$ )
12	408	800
11	426	836
10	447	877
9	471	924
8	500	980
7	535	1048
6	577	1132
5	632	1240
4	707	1386
3	816	1600
2	1000	1960
1	1414	2772

TABLE 4

INCREASE ( $-\Delta\sigma_{\bar{d}}$ ) IN THE STANDARD ERROR IN DAMAGE ESTIMATES FOR TWO JURIES OF THE SAME SIZE ( $n$ ) HEARING THE SAME CASE AND SELECTED AT RANDOM FROM A NORMAL POPULATION OF STANDARD DEVIATION  $\sigma = 1000$  AS JURY SIZE IS REDUCED IN INCREMENTS OF ONE

Jury Size ( $n$ )	$-\Delta\sigma_{\bar{d}} = (\sigma_{\bar{d}})_n - (\sigma_{\bar{d}})_{n-1}$
12	18
11	21
10	24
9	29
8	35
7	42
6	55
5	75
4	109
3	184
2	414
1	—

FIGURE 2  
INCREASE IN STANDARD ERROR WITH  
DECREASING JURY SIZE

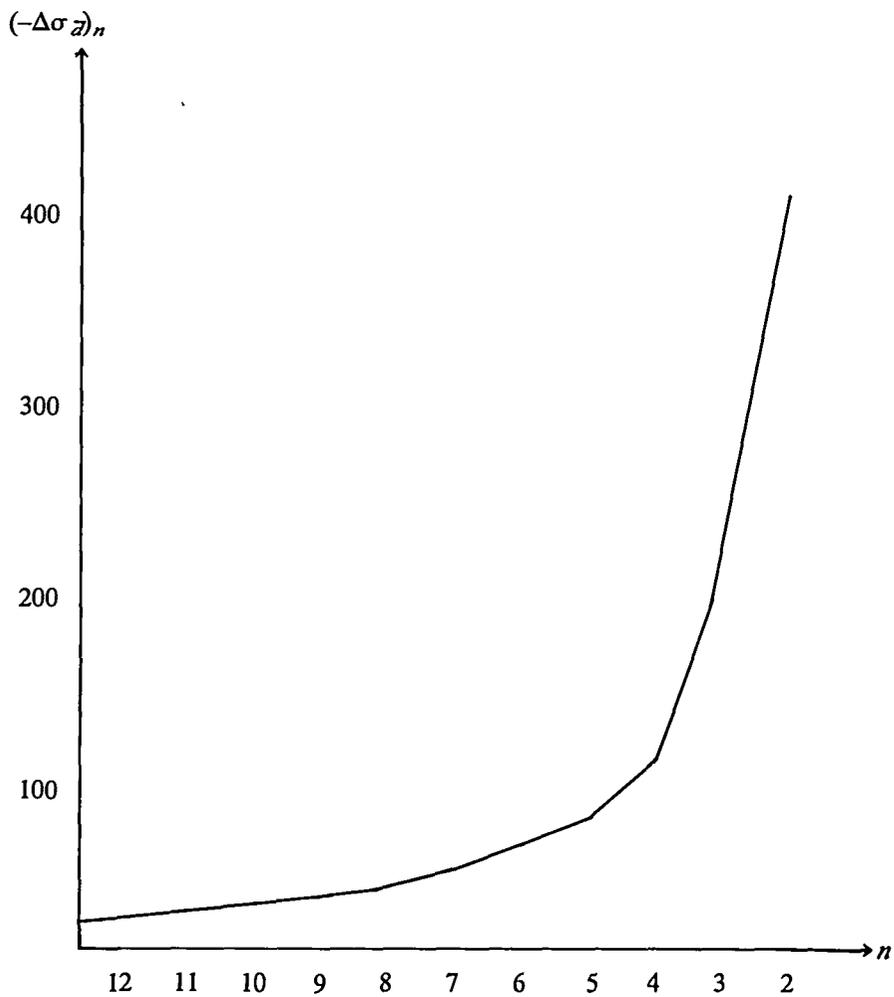


TABLE 5

PROBABILITIES FOR CONVICTION ( $P_C$ ), NONCONVICTION ( $P_{\bar{C}}$ ), GUILT GIVEN CONVICTION ( $P_{G/C}$ ), INNOCENCE GIVEN NONCONVICTION ( $P_{I/\bar{C}}$ ), CONVICTION AND INNOCENCE ( $P_{CI}$ ), AND NONCONVICTION AND GUILT ( $P_{\bar{C}G}$ ) ACCORDING TO POISSON MODEL ( $\theta = .69$ ,  $\mu = .88$ , simple model of majority persuasion)

Jury Size ( $n$ )	$P_C$	$P_{\bar{C}}$	$P_{G/C}$	$P_{I/\bar{C}}$	$P_{CI}$	$P_{\bar{C}G}$
12	.6897	.3103	.9997	.9984	$2.069 \times 10^{-4}$	$4.965 \times 10^{-4}$
6	.6842	.3158	.9937	.9695	$4.310 \times 10^{-3}$	$9.632 \times 10^{-3}$
5	.6842	.3158	.9940	.9696	$4.105 \times 10^{-3}$	$9.600 \times 10^{-3}$

## SIMPLIFIED DERIVATION OF POISSON MODEL

The Poisson model uses two variables to predict the probability of conviction and other, related probabilities. The mathematics is most easily understood by considering the simplest possible cases—that of a one-member “jury” and a two-member panel. The probability that a one-member “jury” will convict can be called  $P_C$  and expressed in terms of two parameters: the probability that the decision will be correct—denoted by  $\mu$ —and the probability that the defendant is guilty—denoted by  $\theta$ . To see how this is done, some further notation is helpful:

<u>symbol</u>	<u>meaning</u>
G	defendant is guilty
$\bar{G}$	defendant is not guilty
$P_G$	probability that defendant is guilty
$P_{\bar{G}}$	probability that defendant is not guilty
C	juror votes to convict
$\bar{C}$	juror does not vote to convict
$P_C, P_{\bar{C}}$	probabilities associated with C, $\bar{C}$
$C/G$	juror votes to convict given that defendant is guilty
$C/\bar{G}$	juror votes to convict given that defendant is not guilty
$P_{C/G}, P_{C/\bar{G}}$	probabilities associated with $C/G, C/\bar{G}$
CG	juror votes to convict and defendant is guilty
$C\bar{G}$	juror votes to convict and defendant is not guilty
$P_{CG}, P_{C\bar{G}}$	probabilities associated with CG and $C\bar{G}$

Now consider the case of a one-person “jury.” The probability that the “juror” will convict can be written as the sum of two terms: (1) the probability that he will convict when the defendant really is guilty and (2) the probability that he will convict when the defendant is not guilty.

In other words,

$$P_C = P_{CG} + P_{C\bar{G}} \quad (1)$$

Using the definition of conditional probability

( $P_{AB} = P_{A/B}P_B = P_{B/A}P_A$ ), we have

$$P_C = P_{C/G}P_G + P_{C/\bar{G}}P_{\bar{G}} \quad (2)$$

Now  $C/G$  is a correct decision and  $C/\bar{G}$  is an incorrect one. From the definition of  $\mu$  then,  $P_{C/G} = \mu$  and  $P_{C/\bar{G}} = 1 - \mu$ .

Similarly,

$P_G = \theta$  and  $P_{\bar{G}} = 1 - \theta$ . Consequently,

$$P_C = \mu\theta + (1-\mu)(1-\theta) \quad (3)$$

Equation (3) expresses  $P_C$  in terms of  $\theta$  and  $\mu$ , as promised, but is otherwise uninteresting. The two-person case is more revealing. It will be convenient to let  $\gamma_{n,i}$  represent the probability that exactly  $i$  jurors out of  $n$  will vote for acquittal on the first ballot. For the two-member jury, the probability that both jurors will vote to convict is  $\gamma_{2,0}$ . By an extension of the reasoning which led to Equation (3), it follows that

$$\gamma_{2,0} = \theta\mu^2 + (1-\theta)(1-\mu)^2 \quad (4)$$

The probability that the jury will split is

$$\gamma_{2,1} = \theta\mu(1-\mu) + (1-\theta)\mu(1-\mu) \quad (5)$$

The right hand side of Equation (5) can be simplified, of course, to  $\mu(1-\mu)$ . The remaining possibility is that both jurors vote to acquit on the first ballot. The probability of this event is given by

$$\gamma_{2,2} = \theta(1-\mu)^2 + (1-\theta)\mu^2 \quad (6)$$

The values of  $\theta$  and  $\mu$  can now be deduced—provided that we have measurements of the way two-member juries divide their first ballot votes in a representative cross-section of cases. With such measurements, specific values of  $\theta$  and  $\mu$  can be selected that generate the set of values for  $\gamma_{2,0}$ ,  $\gamma_{2,1}$ , and  $\gamma_{2,2}$  that most closely matches the observed set of values.

The generalization of this procedure to deal with other values of  $n$  and  $i$  is straightforward. One can show that, in general,

$$\gamma_{n,i} = \binom{n}{i} [\theta\mu^{n-i}(1-\mu)^i + (1-\theta)\mu^i(1-\mu)^{n-i}] \quad (7)$$

Using data on the first ballot voting patterns of twelve-member juries,<sup>4</sup> Gelfand and Solomon calculate values for  $\theta$  and  $\mu$  and use these values to make predictions about  $\gamma_{n,i}$  and related probabilities.<sup>5</sup> For the reasons stated in Part III of the text, these predictions may not be reliable.

---

4. See H. KALVEN & H. ZEISEL, *supra* note 67.

5. See articles cited note 60 *supra*.