1975

Uncovering "Nondiscernible" Differences: Empirical Research and the Jury-Size Cases

Richard O. Lempert
University of Michigan Law School, rlempert@umich.edu

UNCOVERING "NONDISCERNIBLE" DIFFERENCES: EMPIRICAL RESEARCH AND THE JURY-SIZE CASES

Richard O. Lempert

[643]

UNCOVERING "NONDISCERNIBLE" DIFFERENCES: EMPIRICAL RESEARCH AND THE JURY-SIZE CASES

*Richard O. Lempert**

I. REASONS FOR NOT DISCERNING DIFFERENCES

IN *Williams v. Florida*[1] the Supreme Court, relying on a "few experiments" that in the eyes of the majority indicated "no discernible difference" in the results reached by six- and twelve-member juries,[2] held that a decrease in the size of the criminal jury from twelve to six members was not inconsistent with the sixth amendment right to jury trial. In *Colgrove v. Battin*[3] the majority read four recent studies as providing "convincing empirical evidence"[4] confirming the conclusion in *Williams,* and so found a

1. 399 U.S. 78 (1970).

2. 399 U.S. at 101. The studies cited were Cronin, *Six-Member Juries in District Courts,* 2 BOSTON B.J. No. 4, at 27 (1958); Phillips, *A Jury of Six in All Cases,* 30 CONN. B.J. 354 (1956); Tamm, *The Five-Man Civil Jury, A Proposed Constitutional Amendment,* 51 GEO. L.J. 120 (1962); Wiehl, *The Six Man Jury,* 4 GONZAGA L. REV. 35 (1968); *New Jersey Experiments with Six-Man Jury,* 9 BULL. OF THE SECTION OF JUD. ADMIN. OF THE ABA (May 1966); *Six-Member Juries Tried in Massachusetts District Court,* 42 J. AM. JUD. SOC. 136 (1958).

3. 413 U.S. 149 (1973).

4. 413 U.S. at 159-60 n.15. The four cited studies were INSTITUTE OF JUDICIAL ADMINISTRATION, A COMPARISON OF SIX- AND TWELVE-MEMBER CIVIL JURIES IN NEW JERSEY SUPERIOR AND COUNTY COURTS (1972); Bermant & Coppock, *Outcomes of Six- and Twelve-Member Jury Trials: An Analysis of 128 Civil Cases in the State of Washington,* 48 WASH. L. REV. 593 (1973); Note, *An Empirical Study of Six- and Twelve-Member Jury Decision-Making Processes,* 6 U. MICH. J. L. REF. 712 (1973); Note,

[644]

similar decrease in the size of the civil jury to be not inconsistent with the seventh amendment.[5] In both cases the Court considerably overstated the implications of the empirical evidence that it cited, and in both cases Professor Hans Zeisel, one of the country's leading academic advocates of empirical research into law-related problems[6] and preeminent among students of the American jury,[7] responded with telling criticism.

In his article responding to *Williams*, Zeisel shows that the "experiments" on which the Court relied are in no scientific sense experimental; rather they are speculative or impressionistic reports based on limited or, in one case, no experience with the six-member jury.[8] Zeisel is gentle with the Court; he never emphasizes the majority's extreme disingenuousness in citing these reports as experiments and in relying on them as evidence of "no discernible difference." Certainly, a careful reader of the cited studies would have appreciated their nonexperimental nature and their limited bearing on the issue of whether jury size affects jury verdicts.[9]

Six-Member and Twelve-Member Juries: An Empirical Study of Trial Results, 6 U. MICH. J. L. REF. 671 (1973).

5. *Williams* was a state case, but the opinion appears to apply to the federal criminal jury as well. However, Justice Powell, not on the Court when *Williams* was decided, apparently would allow the states more leeway to vary traditional modes of jury trial than he would the federal government. *See* Johnson v. Louisiana, 406 U.S. 356, 366 (1972) (Powell, J., concurring in both *Johnson* and Apodaca v. Oregon, 406 U.S. 404 (1972)). If Justices Brennan and Douglas were to realize that many of the values that they saw threatened by allowing nonunanimous verdicts, *see* 406 U.S. at 380, 395 (dissenting opinions), also were undermined by diminishing the size of the jury, it is possible that the Court, if faced with the issue, would find that the sixth amendment continues to require twelve-member juries in federal criminal cases. *Colgrove* sustained federal district court rules providing for six-member juries in civil actions.

6. *See, e.g.*, H. ZEISEL, H. KALVEN & B. BUCHOLZ, DELAY IN THE COURT (1959); Zeisel, *Reflections on Experimental Techniques in the Law*, 2 J. LEGAL STUDIES 107 (1973); Zeisel, *Social Research on the Law: The Ideal and the Practical*, in LAW AND SOCIOLOGY 124 (W. Evan ed. 1962); Zeisel, *The Uniqueness of Survey Evidence*, 45 CORNELL L.Q. 322 (1960).

7. *See, e.g.*, H. KALVEN & H. ZEISEL, THE AMERICAN JURY (1966) [hereinafter THE AMERICAN JURY]; Zeisel, *The Jury and the Court Delay*, 328 ANNALS 46 (1960); Zeisel, *What Determines the Amount of Argument Per Juror?*, 28 AM. SOC. REV. 279 (1963).

8. Zeisel, . . . *And Then There Were None: The Diminution of the Federal Jury*, 38 U. CHI. L. REV. 710, 713-15 (1971). For a student piece making some of the same arguments, see Note, *The Effect of Jury Size on the Probability of Conviction: An Evaluation of Williams v. Florida*, 22 CASE W. RES. L. REV. 529 (1971).

9. This disingenousness, coupled with the mistaken way in which the Court reads findings from *The American Jury, supra* note 7, suggests that the majority simply assumed that jury size had no relation to jury verdicts and was searching for evidence to support this assumption. *See* Zeisel, *supra* note 8, at 719-20 (commenting on the Court's misreading of data from *The American Jury* in *Williams*, 399 U.S. at 101-02 n.49).

The response to *Colgrove*,[10] written with Shari Seidman Dia-
mond, focuses on the four studies cited by the majority[11] as further
support for the proposition that there is no discernible difference
between the results reached by six- and twelve-member juries. Zeisel
and Diamond demonstrate that significant flaws in the design of
each study preclude any cautious observer from basing conclusions
about differences between six- and twelve-member juries on the
reported results.[12]

The article responding to *Williams* closes with a statistical anal-
ysis of the relationship between jury size, composition, and behavior.
Zeisel demonstrates that, given certain plausible assumptions, a
change in jury size from twelve to six members should influence

---

10. Zeisel & Diamond, *"Convincing Empirical Evidence" on the Six-Member Jury*,
41 U. CHI. L. REV. 281 (1974).

11. *See* note 4 *supra.*

12. *See* Zeisel & Diamond, *supra* note 10, at 283-90. The New Jersey and Washing-
ton studies were confounded by what have been called "selection effects." Both studies
were based on trials in which the parties to a case determined jury size. Good experi-
mental technique demands random assignment. If those cases in which the parties
chose the twelve-member jury were systematically different from the cases in which
they chose the six-member jury, differences associated with the different-sized juries
might in fact be attributable to the differing nature of the cases heard. Similarly,
findings of no difference between different-sized juries might occur where real differ-
ences associated with jury size are canceled by differences associated with the
kinds of cases heard by the different-sized juries. It is clear that such a systematic
difference existed in the New Jersey study—twelve-member juries were chosen when
more money was at stake. INSTITUTE OF JUDICIAL ADMINISTRATION, *supra* note 4, at 7.
A similar difference in the Washington study is possible. *See* Zeisel & Diamond, *supra*,
at 284. The Washington study is further limited, although Zeisel and Diamond do
not point this out, because it concerned only workmen's compensation cases involving
appeals from administrative decisions under an unusual procedure whereby the at-
torneys read the record of the administrative hearing to the jury. Bermant & Coppock,
supra note 4, at 594. Given these circumstances, the results of even a well-designed
study could not be safely generalized to all civil trials.

The Michigan Before-and-After Study, Note, 6 U. MICH. J. L. REF. 671, *supra*
note 4, suffers from unfortunate "historical" effects. At about the time when Michigan
switched from twelve- to six-member juries in civil cases, the state also introduced a
mediation board procedure and began to allow discovery of insurance policy limits.
These changes had a substantial impact on the incidence of pre-trial settlements, so
that the cases heard by the six-member juries were not comparable to those heard
by the larger panels. *See* Zeisel & Diamond, *supra*, at 288-89.

The Michigan Laboratory Experiment, Note, 6 U. MICH. J. L. REF. 712, *supra*
note 4, also suffers from a number of defects that prevent one from accepting its
findings as evidence that there is no difference in the results reached by six- and
twelve-member juries. Chief among these are the small number of juries studied—eight
of each size—and the fact that the particular case chosen impressed most jurors as
favoring the defendant. *See* Zeisel & Diamond, *supra*, at 286-87. As discussed below,
see text at notes 19-37 *infra*, a strong case for either party masks any effects of jury
size. For a fuller critique of the Michigan Laboratory Experiment, see Diamond, *A
Jury Experiment Reanalyzed*, 7 U. MICH. J. L. REF. 520 (1974).

For a discussion of a variety of factors that often must be controlled for in legal-
impact studies, see Lempert, *Strategies of Research Design in the Legal Impact Study:
The Control of Plausible Rival Hypotheses*, 1 LAW & SOC. REV. 111 (1966).

jury verdicts.[13] The article criticizing the *Colgrove* studies suggests ways in which jury research ought to be conducted to test for size effects.[14] Taken together, the two articles imply that if the *Colgrove* studies did not have the methodological weaknesses for which they are criticized, verdict differences associated with jury size would be revealed. It is this implication that I wish to dispute in the first section of this article. With one possible exception, the likely effect of conducting methodologically more sophisticated versions of the *Colgrove* studies is the generation of further "convincing empirical evidence" to support the Court's position. This is so even if there are situations where jurysize undoubtedly affects jury verdicts.

A. *Problems of Jury-Size Research*

My point is not that verdict differences associated with jury size cannot be revealed through careful empirical investigation. Indeed, at several places in this article I will suggest research strategies likely to reveal such differences.[15] Rather, it is that typical strategies of legal-impact research, such as those utilized in the *Colgrove* real-world studies,[16] are unlikely to uncover differences associated with jury size however well they control for those plausible rival hypotheses that form the usual threats to the validity of impact research.[17] The reason lies in the unamenability of the jury-size problem to the usual techniques of aggregate data analysis.

The difficulties of the real-world approach are best illustrated if we look at the careful research designs that Zeisel and Diamond present in the latter part of their *Colgrove* article. Their "ideal" design would require a jurisdiction in which six-member juries were optional.[18] During the experimental period, cases in that jurisdiction would be tried simultaneously before two juries, one of six and one of twelve members. Although the court and the attorneys would know which jury would render the binding verdict, the two juries would deliberate without this knowledge. Effects of jury size presumably would be assessed by comparing the verdicts reached. Disagreement in a substantial proportion of the verdicts rendered in

---

13. Zeisel, *supra* note 8, at 715-20.

14. Zeisel & Diamond, *supra* note 10, at 291-92.

15. *See* Part III *infra.*

16. By "real-world studies" I mean studies of actual juries deciding real cases. These would include all of the studies discussed in *Colgrove* except the Michigan *Laboratory Experiment*, Note, 6 U. MICH. J. L. REF. 712, *supra* note 4.

17. For a discussion of common rival hypotheses, see Lempert, *supra* note 12.

18. *See* Zeisel & Diamond, *supra* note 10, at 291.

each case would be evidence that size effects existed. This conclusion would be strengthened if the differences were directional, since random factors could explain some nondirectional disagreement. If, on the other hand, the two juries disagreed in only a small number of cases, one would not be inclined to attribute the difference in verdicts to jury size.

In the abstract, this design and analytical framework might seem ideal. It has the particular virtue of revealing verdict differences on a case-by-case basis.[19] However, it holds a major trap for the unwary researcher: In any actual study, data analysis is likely to proceed on the implicit assumption that each trial provides an occasion on which any existing jury-size effects can, with a certain *constant* probability, be expected to appear. If, for example, in thirty out of one hundred trials the simultaneous verdicts were different, the researcher probably would conclude that there were substantial jury-size effects. If, on the other hand, differences arose in only five out of one hundred pairs of verdicts, the conclusion probably would be that jury-size effects were minimal.[20] But the probability of divergent verdicts in any particular case may be close to zero. A five per cent disagreement rate may reflect disagreement in all or a substantial percentage of those cases where jury size *reasonably* could be expected to influence jury verdicts.

B. *Most Cases Are Clear*

One may test this last point by attempting to estimate the fraction of cases in which jury size can be expected to have a reasonable probability of affecting the verdict. The best empirical starting point for making such an estimation is Kalven and Zeisel's *The American Jury*. In this book the authors perform a radical version of the experiment that Zeisel and Diamond propose. Instead of comparing the verdicts reached by juries of twelve with those reached by juries of six, they compare the verdicts of juries of twelve with those of "juries" of one, the judge. The difference between judges and twelve-member juries in socioeconomic status, legal sophistication, role conceptions, and cognitive processes of evidence evaluation are likely to exceed substantially the differences between six- and twelve-member juries with respect to these factors. Thus, one may

---

19. *See* text at notes 42-46 *infra.*

20. This difference might be better evaluated if one changed the experiment so that *two* juries of each size heard each trial. This would allow one to compare explicitly the amount of disagreement when different-sized juries judge the same cases with the amount of disagreement attributable to random differences among juries of the same size.

take the percentage of verdict agreement between judges and twelve-member juries as a minimum estimate of the percentage of cases in which jury-size effects could be expected to have virtually no influence on the ultimate verdict. In *The American Jury* the reported level of agreement is about seventy per cent.[21] If this analysis is accurate, different verdicts by six- and twelve-member juries in ten per cent of the Zeisel and Diamond experimental trials would reflect disagreement in thirty-three per cent of the cases in which disagreement was a realistic possibility. The former figure might strike some as minimal evidence of a size effect, while the latter figure would lead many to conclude that powerful size effects existed. The same difference might be reflected in formal determinations of statistical significance.

In fact, 30 per cent is probably far too high an estimate of the percentage of cases in which jury size could reasonably be expected to influence jury verdicts. The detailed data analysis in *The American Jury* provides a basis for a more refined estimate, although any final figure is necessarily imprecise.[22] Kalven and Zeisel begin their analysis with a total of 3576 cases.[23] In 1063 instances the judge and jury disagreed.[24] This constitutes 29.7 per cent of the sample cases and is a first estimate of the maximum number of cases in which jury-size effects might foster divergent verdicts. One may refine this estimate by eliminating those cases of disagreement in which size effects are very unlikely or by determining the percentage of cases in which the *possibility* of jury-size effects must be acknowledged.

Kalven and Zeisel report that the jury hangs in 197 cases, or 5.5 per cent of the total sample.[25] Jury size is likely to affect the probability that a jury will hang,[26] so these may be treated as cases in which jury size potentially affects the ultimate verdict.[27] In 65

---

21. The American Jury, *supra* note 7, at 109. This figure excludes cases in which the jury hangs as well as cases in which judge and jury disagree on a final verdict. Kalven and Zeisel deal only with criminal cases. The analysis that follows is limited by this restriction as well as by any methodological weaknesses that inhere in their study.

22. Any appearance of precision in a final estimate of the percentage of cases in which jury-size effects may be manifested is false. The estimate reflects any imprecision in the data presented in *The American Jury*, and it is weakened by the fact that those data are not presented in such a way as to facilitate this reanalysis. Particular problems are presented by the fact that in many cases more than one reason is given for judge-jury disagreement.

23. The American Jury, *supra* note 7, at 47.

24. *Id.* at 109, Table 23.

25. *Id.* at 110, Table 24.

26. *See* Zeisel, *supra* note 8, at 720 (citing data from Florida trials that suggest that six-member juries hang about half as often as twelve-member juries).

27. Treating all hung juries as cases in which one might expect a difference attribut-

additional cases, or 1.8 per cent of the total sample, Kalven and Zeisel find no reason for judge-jury disagreement.[28] Since these might be cases in which jury size could affect the verdict, the conservative procedure is to count them with the hung-jury cases for a minimum estimate of 7.3 per cent. This leaves 801 cases, or 22.4 per cent of the total, to be explained. In all but 7 of these cases disagreement is attributed in whole or in part to jury sentiments about the defendant, jury sentiments about the law, or evidentiary factors.[29] One subcategory under evidence, amounting to 2.1 per cent of the sample, is attributed to random or unknown factors.[30] Since these could be factors that differentially affect the verdicts of different-sized juries, these cases raise the minimum estimate to 9.4 per cent, leaving 20.3 per cent of the total sample unexplained.

The remaining subcategories under evidence and the categories involving jury sentiment all appear to reflect differences between the attitudes that jury members and judges bring to cases.[31]

---

able to jury size is a conservative procedure, that is, it leads to an overestimate of the percentage of cases in which one might expect jury-size effects to be manifested. Some cases are so close, confused, or divisive that one would expect six- as well as twelve-member juries to hang.

28. THE AMERICAN JURY, *supra* note 7, at 110-11, Tables 24, 26. There are actually 101 cases where Kalven and Zeisel cannot explain reported disagreement, but 36 of these involve hung juries and so have been already counted.

29. The 7 exceptions are cases in which disagreement is attributed solely to facts only the judge knew or to disparity of counsel. It would appear that these factors would not interact with jury size, so these cases might be subtracted from the ceiling figure. However, it is conservative to conclude that all of these cases resulted in hung juries and thus to make no correction. In 123 cases involving jury sentiment or evidentiary factors there is disparity of counsel or facts only the judge knew. *Id.* at 113, Table 28. Since there is no reason to expect these factors to affect different-sized juries in different ways, there is no reason to single them out for analytical purposes.

30. *Id.* at 390, Table 59. In calculating this percentage a correction had to be made so that hung juries would not be counted twice. For this purpose the weighted figures presented in *id.* at 456, Table 120, were used. First, 7 cases were subtracted from the total number of cases, and the categories of "facts only the judge knew" and "disparity of counsel" were eliminated. *See* note 29 *supra*. Then the percentages attributed to the remaining categories were adjusted so they would total 100 per cent. This resulted in 74.7 per cent or 115 of 154 remaining hung juries being attributed to "evidence factors." In the absence of further information it was assumed that these hung jury cases were distributed proportionately across the subcategories of the "evidence factors" presented. These hung-jury cases were subtracted from the total number of "evidence factor" cases, leaving 638 cases. The proportion of evidence disagreements attributed to random or unknown factors among these cases equaled 77 cases, or 2.1 per cent of the total sample.

31. This appears obvious where judge-jury differences turn on the jury's sentiments about the law or the defendant. It is less obvious where disagreement is attributed to evidence factors, because here the cognitive differences between individual and group evaluations might be expected to lead to different verdicts totally apart from any attitudinal differences. In fact, however, with the exception of the "random" subcategory discussed in the text, all of the divergent verdicts that Kalven and Zeisel attribute to evidence factors reflect attitudinal differences between judge and jury rather than

Differences in verdicts of different-sized juries in these circumstances are unlikely unless the selection of six jurors results in a panel with a substantially different array of attitudes than that found in a jury of twelve. The problem is to estimate the proportion of cases in which the characteristic attitudes of different-sized juries will be so different that divergent outcomes are likely. One generally reasonable behavioral assumption makes this problem statistically manageable: In those situations in which jury verdicts reflect juror attitudes, the final verdict will reflect the attitudes held by an initial majority of the jury.[32] This assumption implies that in the remaining cases of judge-jury disagreement panels were drawn in such a way that seven or more members held attitudes toward the crime, the defendant, or the fact-finding process that the judge did not share or considered inappropriate.[33] Accordingly, the verdict of a six-member jury will diverge from that of a larger jury if four or more panel members do not share key attitudes held by seven members of the twelve-member panel. Thus, the problem as refined is as follows: Assume that a twelve-member jury has been drawn and that seven or more of its members hold attitude X. If one dismisses

cognitive differences. In 56 per cent of the evidence disagreements, the closeness of the evidence gave the jury a sense that it was at liberty to indulge sentiments about the law or the defendant that might otherwise have been inappropriate. In 21 per cent of these cases juries apparently saw witnesses, particularly witnesses without records, as much more credible than did more experienced judges. In 11 per cent of these judge-jury disagreements, the jury apparently set a higher standard of reasonable doubt than did the judge. *Id.* at 390, Table 59. In all of these situations it appears that the difference between judge and jury was influenced more by the relatively set attitudes that jurors brought to their deliberations than by differences in the way in which individuals and groups evaluate facts.

32. This assumption appears to be a reasonable description of how a jury typically behaves. Kalven and Zeisel report on the basis of juror interviews that where the initial ballot contains seven or more votes of "guilty," a final verdict of "guilty" is rendered 90 per cent of the time, a result of "not guilty" is reached about 4 per cent of the time, and the jury hangs in the remaining 6 per cent of the cases. Where the initial majority votes "not guilty," a final verdict of "not guilty" is rendered 94 per cent of the time, a verdict of "guilty" 1 per cent of the time, and the jury hangs 4 per cent of the time. *Id.* at 488, Table 139 (combining the first and second and fourth and fifth columns). One problem with this analysis is that the first ballot may not occur until after some discussion. However, Broeder, commenting on the same study, says that in most cases the ballot was taken immediately. Broeder, *The University of Chicago Jury Project,* 38 NEB. L. REV. 744, 747 (1958). *See* Note, *supra* note 8, at 542-45. Working with real jurors, Simon reports that in only 9 per cent of all cases and 11 per cent of cases reaching a consensus did a jury decide against an initial majority. R. SIMON, THE JURY AND THE DEFENSE OF INSANITY 117 (1967). In similar experiments, Hawkins reports that in only 15 per cent of all cases did the jury hold against an initial majority. C. Hawkins, Interaction and Coalition Realignments in Consensus-Seeking Groups: A Study of Experimental Jury Deliberations, August 17, 1960 (unpublished doctoral dissertation in University of Chicago Library).

33. Some judges felt that the jury's role entitled them to consider certain factors that a judge should not. THE AMERICAN JURY, *supra* note 7, at 450, Table 1B.

that jury and draws a jury of six members from the same popula-
tion, what is the probability that four or more members of the
smaller jury will hold attitude X? The best estimate is about 77 per
cent;[34] thus, where a jury's decision may be traced to the attitudes

34. For the solution of this problem I am indebted to my research assistant, Dan
Russell, and to the consulting services of The University of Michigan's Statistical Lab-
oratory.

Take a population of characteristic X of unknown proportion, p, such that $0 \leq p \leq$
1. Because p is unknown, assume it may fall with equal probability anywhere between
0 and 1 (In other words, we assume a uniform distribution on p, $f(p) = 1$. If we had
some knowledge of p, we could assume a beta distribution for p that approximated
our knowledge of p.)

Let $B =$ the event that 7 or more persons out of a sample of 12 have characteris-
tic X.

Let $A =$ the event that 4 or more persons out of a sample of 6 have characteris-
tic X.

Assume a large enough population that one can use a binomial approximation (or
assume sampling with replacement). The conditional probability of A given B is ex-
pressed as follows:

$$\text{Prob. } (A|B) = \frac{\text{Prob. } (A\&B)}{\text{Prob. } (B)}.$$

To find Prob. (B) and Prob. $(A\&B)$ Bayesean methods can be used, since both prob-
abilities are dependent on p, which is unknown.

Let $k_1(b)$ be the probability density function (pdf) for the 12-member sample, which
we want to find to determine Prob. (B).

$g(b|p)$ is the conditional pdf of the 12-member sample, given p.

$g(b|p) = \binom{12}{b} p^b(1-p)^{12-b}$.

$k(p,b)$ is the joint pdf for p and b.

Thus, $k(p,b) = g(b|p)f(p)$ $\left[\text{since } g(b|p) = \dfrac{k(p,b)}{f(p)}\right]$.

$k_1(b)$ is the marginal pdf of b, which is given by $k_1(b) = \displaystyle\int_0^1 g(b|p)f(p)dp$.

Hence, if n is the required number of successes with respect to B, and m is the
required number of successes with respect to A, Prob. $(B) = k_1(B) =$

$$\int_0^1 \sum_{n=7}^{12} \binom{12}{n} p^n(1-p)^{12-n}f(p)dp.$$

Similarly,

$$\text{Prob. } (A\&B) \int_0^1 \left[\sum_{m=4}^{6} \binom{6}{m} p^m(1-p)^{6-m}\right]\left[\sum_{n=7}^{12} \binom{12}{n} p^n(1-p)^{12-n}\right]f(p)dp.$$

Since we let $f(p) = 1$, we have Prob. $(A|B) = \dfrac{\text{Prob. } (A\&B)}{\text{Prob. } (B)} =$

$$\frac{\displaystyle\int_0^1 \left[\sum_{m=4}^{6} \binom{6}{m} p^m(1-p)^{6-m}\right]\left[\sum_{n=7}^{12} \binom{12}{n} p^n(1-p)^{12-n}\right]dp}{\displaystyle\int_0^1 \sum_{n=7}^{12} \binom{12}{n} p^n(1-p)^{12-n}dp}$$

$$= .7692.$$

with which its members view a problem, the verdict of a six-member jury drawn randomly from a population will differ from the verdict rendered by a twelve-member jury drawn from the same population about 23 per cent of the time.[35] Adding these cases of expected divergency (23 per cent of the 20.3 per cent of cases in which judge-jury disagreement was attributable to additudinal differences, or 4.7 per cent of the total sample) to the previous minimum estimate of 9.4 per cent produces a final estimate of 14.1 per cent as the proportion of cases in which jury size has a reasonable probability of affecting jury verdicts.

This final estimate explains why real-world research, even research as carefully designed as Zeisel and Diamond's "ideal" experiment, is likely to lend apparent support to the Supreme Court's conclusion that there is no discernible difference in the verdicts rendered by different-sized juries. There are simply too few cases in which it is reasonable to suppose that size effects will be manifested in divergent verdicts. If, for example, size affects verdicts in one third of the cases where such effects appear possible, the aggregate data collected in a Zeisel and Diamond ideal experiment would reveal fewer than five divergencies in every hundred trials.[36] Since some divergent verdicts might be expected to occur by chance, such a low figure would lead most to conclude too hastily that the *Williams* decision was right and that jury size has little or no impact on jury verdicts.[37]

See generally A. Hogg & A. Craig, Introduction to Mathematical Statistics, 61, 63, 208 (1970)

35. It should be noted that much of this difference in expected verdicts is not due to differences in jury size. Rather, it is due to the fact that two random samples have been drawn from the same population. If, for example, C were the event that 7 or more persons out of a second independent sample of size 12 would have characteristic X, then

$$p(C|B) = .8358.$$

36. Size effects would not be anticipated in all, or indeed most, of those cases in which the preceding analysis suggests they are possible. Many of the cases included in the 14.1 per cent estimate were included simply because lack of information meant that the possibility of a size effect could not be ruled out. In addition, while size can be expected to affect the rate of jury hanging, many cases in which juries hang may reflect the closeness of the case rather than size-related aspects of jury dynamics. Including all of the hung-jury cases in the 14.1 per cent figure was therefore extremely conservative. *See* note 27 *supra.*

37. It could be argued that, even if the absolute number of divergent verdicts was small, the sophisticated researcher could spot trends reasonably attributable to jury size. For example, if there were 10 divergent verdicts out of 140 trials and 8 of these were cases in which the twelve-member jury hung while the six-member jury reached a verdict, one might conclude that twelve-member juries were more likely to hang than six-member juries. There are two problems with this argument. First, the

C. *Selection Effects and Misleading Aggregates*

As an alternative to their ideal design, Zeisel and Diamond describe a "next best" design that in their view approximates a

---

researcher might still conclude that the difference is so small in magnitude as to be unimportant or show only weak size effects. Second, there appears to be no reason to expect size effects to manifest this amount of directionality. In some cases size effects might lead one jury to favor the plaintiff, in some they might lead it to favor the defendant, and in some they might lead it to reach a verdict where the other jury hangs. Of course, the possibility of spotting trends associated with jury size increases with the number of trials (and, hence, the number of divergent verdicts) examined. If a sufficiently large number of cases were examined, some of the objections that I have made to Zeisel and Diamond's ideal experiment would be weakened. However, given the expense involved in providing two juries for all cases, the desire of researchers to finish their work in a reasonably short time, and general expectations about what can be proved with relatively small samples, I would guess that reports of an "ideal design" experiment would be based on between 100 and 300 trials.

Professors Zeisel and Diamond suggest that this problem can be dealt with by the following "easily drawn, but difficult to implement" modification of their ideal design:

Judge's opinion of case

	Clear for:		Not clear, leaning toward:	
	Acquittal	Conviction	Acquittal	Conviction
Percentage of six-member jury verdicts agreeing with judge				
Percentage of twelve-member jury verdicts agreeing with judge				

Letter from Hans Zeisel to the author, November 4, 1974 [hereinafter Letter]. This modification does separate those cases in which the evidence is so clear that size effects are unlikely to be reflected in verdict differences. However, it raises the practical problems of securing judicial cooperation and spending enough time and money to acquire an adequate sample of "not clear" cases.

Ironically, the analysis in the text suggests that the Court may have been more justified in relying upon the nonexperiments cited in *Williams* than it would have been in relying upon apparently well-designed replications of the real-world studies cited in *Colgrove*. At least some of the studies noted in *Williams* were based upon the intuitions of experienced observers that there were no differences in the results reached by six- and twelve-member juries. These observers did not systematically count observations; they reported a *gestalt*, a general perception that a group of cases tried to six-member juries produced results similar to those that would have been expected from an analogous group of cases tried to twelve-member juries. Had there been even one case in which an observer felt that twelve would have decided differently than six, the conflict with the over-all *gestalt* might well have suggested further investigation. The typical "real-world" researcher, however, will examine only those aspects of the case that he has chosen to measure. At the end there will be only numerical differences of a greater or lesser magnitude; there is unlikely to be any detailed appreciation of the processes that produced these differences. If the differences are small, the researcher is likely to be very wary in attributing them to jury size.

I do not mean to suggest that the Court in *Williams* was on particularly solid ground in relying on impressionistic data. All of the cited reports apparently were written by advocates of the six-member jury, and the advocate is often blind to facts that contravene his position. To the extent that one can rely on soft, uncontrolled data, the New Jersey study provides the most important evidence: Lawyers in New Jersey disproportionately chose the twelve- rather than the six-member jury when larger amounts of money were at stake. *See* note 12 *supra*.

controlled experiment by randomly assigning cases to six- and twelve-member juries.[38] It seems likely that research will follow the "next best" rather than the ideal design, because the former, by avoiding duplicate juries, promises to be both less expensive and more consistent with ordinary court procedure than its ideal counterpart.[39]

The next best design requires a jurisdiction in which jury size is optional. Attorneys in such a jurisdiction would express a preference for six- or twelve-member juries. If the attorney had no preference or preferred the smaller jury only because of the lower jury fee,[40] the case would be assigned to a panel of either twelve or six members by some random "lottery."[41] Presumably one would compare the verdicts of the twelve-member panels with those of the six-member panels to determine any size effect. Although this type of design might be appropriate for some legal-impact research, it has flaws not shared by the ideal design that further increase the likelihood that experimentation will reveal no association between jury size and jury verdicts.

The first flaw is that only cases in which both attorneys are indifferent to jury size will be examined. If attorneys generally have any intuitive appreciation of situations in which jury size might affect verdicts, the proportion of experimental trials with a reasonable probability of size effects will be even smaller than was the case with the ideal experiment. Indeed, the experiment might prove impossible in some circumstances. If in criminal cases defendants' attorneys regard hung juries as victories, one might expect them always to insist on twelve-member juries.

The second and more basic flaw of this next best design turns on the fact that the data will have to be aggregated for analytical purposes.[42] If, for example, the six-member jury trials result in forty per cent defendants' verdicts while the twelve-member trials result in sixty per cent defendants' verdicts, the experimenter presumably

---

38. Zeisel & Diamond, *supra* note 10, at 291.

39. *But see* Zeisel, 2 J. LEGAL STUDIES 107, *supra* note 6, at 123-24, for an example of court cooperation with a research design similar to Zeisel and Diamond's ideal design. The major difference is that the "mock" juries were labeled and seated in the spectator section of the courtroom. In the "ideal design" the mock jury would be unlabeled and presumably would sit with the trial jury.

40. All experimental litigants would pay only the six-member jury fee regardless of the size jury assigned. Zeisel & Diamond, *supra* note 10, at 291.

41. *Id.*

42. This flaw is more basic because it means that even if a court—perhaps in a jurisdiction where six-member juries have been required—randomly assigned cases to six- and twelve-member juries, size effects would be less likely to be revealed than with the ideal experiment.

would conclude that a size effect exists. If, on the other hand, both sets of trials resulted in forty-five per cent defendants' verdicts, the likely conclusion would be that there was no size effect. Yet the latter finding would disguise a twenty per cent disagreement rate between the two types of juries if the twelve-member panels decide ten per cent of the cases for the defendant when the six-member panels would have found for the plaintiff and vice versa. In other words, to the extent that size effects are not directional, differences between the verdicts of six- and twelve-member juries will, in the aggregate, cancel out.[43] Such nondirectionality appears particularly likely in civil litigation. Kalven and Zeisel point out that in the civil cases they have studied, the judge finds for the plaintiff fifty-seven per cent of the time and the jury finds for the plaintiff fifty-nine per cent of the time, hardly a great difference.[44] But an examination of case-by-case statistics reveals a twenty-two per cent rate of judge-jury disagreement.[45] Zeisel and Diamond's next best design would show only the two per cent disagreement rate; their ideal design would report twenty-two per cent disagreement. The flaws inherent in the next best design appear so severe that the results of such research can never substantially support the proposition that jury size has no effect on jury verdicts.[46]

D. *Testing Statistical Significance*

Totally apart from the design used, one characteristic of most empirical analysis enhances the probability that research will fail to discern actual differences in verdicts rendered by different-sized juries. This feature is the routine reliance on significance tests with conventional levels of statistical significance to determine whether

---

43. This cancellation effect would not be expected with respect to certain statistics, such as the percentage of hung juries and the variance of the amounts of damages awarded in civil cases. *See* text at notes 184-86 *infra*. Hence these statistics might be used to measure size effects in a "next best design" experiment. I do not feel, however, that the possibility of using such statistics renders the next best design a viable form of research. Differences in the percentage of hung juries, particularly in civil cases, are likely to be quite small. The courts and other less scientifically sophisticated "consumers" of jury-size research are likely to pay more attention to differences in the direction of jury verdicts and in the average amount of damage awards (for which no difference is expected) than they are to the less familiar statistic, the variance.

44. THE AMERICAN JURY, *supra* note 7 at 63, Table 16.

45. *Id.*

46. One might conclude on the basis of next best design research showing no association between jury size and party verdicts that in the long run reducing the size of the jury does not systematically favor one party over the other. However, unless one knows the situations in which jury size affects jury verdicts, the finding that there is no systematic jury-size bias does not mean that one should be indifferent to the question of jury size. *See* Part II *infra*.

differences deserve to be treated as discernible. To understand the problem one must understand some basic characteristics of significance tests.

Statisticians distinguish two types of errors that one may make in evaluating hypotheses.[47] "Type I" error is the mistake of rejecting a true hypothesis; "type II" error is the mistake of failing to reject a false hypothesis. In a typical jury-size experiment the researcher tests the "null hypothesis"[48] that jury size as an independent variable has no effect on jury verdicts as a dependent variable. A type I error would be to conclude that jury size affects jury verdicts when in fact it does not; a type II error would be to conclude that jury size does not affect jury verdicts when in fact it does.

Empirical research almost always reveals some association between the values of independent and dependent variables. A common research problem is to determine whether observed association reflects a possible causal relationship between the two variables. Tests of significance are designed to deal with one recurrent situation where association between two variables may be observed in the absence of any relationship of interest, that is, where the association results from chance.

An example may clarify the problem. Consider the researcher who wishes to examine the relationship between jury size and the frequency of hung juries, and who has access to records of thousands of cases, some of which were tried to twelve-member juries and some of which were tried to six-member juries. Assume that the researcher cannot afford to collect data on all of the cases in the files. Instead, he decides to start with the first case in the file and examine every fifth case thereafter. If at the end of this investigation he finds that five per cent of the twelve-member juries were hung compared to three per cent of the six-member juries, he may nonetheless be reluctant to conclude that twelve-member juries hang more frequently than six-member juries. It is possible that his results simply reflect the luck of the draw. For no particular reason, twelve-member jury trials ending in hung verdicts might have been disproportionately represented in the cases he sampled. Had he chosen other cases from the file, he might have found that five per cent of the six-member juries were hung as compared to three per cent of the twelve-member juries and, had he examined every trial, he might have found that four per cent of all juries regardless of size were

47. *See, e.g.,* H. BLALOCK, SOCIAL STATISTICS 92-96 (1960).

48. A null hypothesis generally is the hypothesis that there is no association between two variables.

unable to reach verdicts. How is he to know if the results from the cases he did examine are representative of the entire group? How is he to know whether a conclusion that jury size affects propensity to hang is justified? Significance tests aid in solving this problem[40] by indicating the probability that the selected cases would show an association as large as the one observed if there was no association in the larger population from which the sample was drawn. If such tests revealed that the difference was significant at the .05 level, the researcher would be reasonably confident that the observed association was not due to chance. Only one out of twenty times would his random selection procedures lead him to find a difference as large as the one he observed if there was no association between size and hanging among all of the cases in the file. A significance level of .01 would indicate that only one out of one hundred random samples would reveal a difference as large as the observed difference if there was no association in the population from which the sample was drawn.

These significance tests measure the probability of type I error— the probability of rejecting a true null hypothesis.[50] In the above example in which the significance level associated with a two per cent difference was .05, the researcher would reject the null hypothesis; he would decide that jury size was associated with rates of hanging, knowing that the odds were only one in twenty that he was being fooled by chance.

Knowing these odds presents the researcher with another problem. Should he proclaim or publish his results? Is a ninety-five per cent probability that one's results are not due to chance sufficient to justify release of results that may influence statistically unsophisticated judges and legislators in their resolution of important legal issues? There is no certain answer. The decision depends on the costs of making type I and type II errors.[51] The conventions of social science, however, do suggest an answer: At a significance level of .05 one is generally justified in concluding that the relationships

---

49. The example should make clear that regardless of the results of the significance tests, difficulties of inference remain. For example, there might be an association among all cases between jury size and the proportion of hung juries, but this association might reflect the fact that attorneys typically choose larger juries for more complicated cases and that such cases are likely to hang regardless of jury size. This is one reason why findings that are of statistical significance may be of no significance in any practical sense.

50. *See* note 48 *supra.*

51. For a given sample size the probability of type II error increases as the probability of type I error decreases. For a given significance level, the probability of type II error decreases as sample size increases.

in the data are not due simply to chance. More importantly, the conventions implicitly assume that significance levels much above .05 justify or indeed require the decision not to reject the null hypothesis. Thus, relationships significant at the level of .1 or greater are often not reported in the literature or, if reported, are mentioned as not significant.[52] Too often there is little evidence that the researcher has thought about the suitability of the conventions for his particular problem.

The difficulty with accepting conventional significance levels as a guide in jury-size research stems from the fact that these levels are selected to be very conservative with respect to type I error. This conservatism reflects the values of social science. Data-based theory must rest on relatively firm foundations. Multiple replications of research to ensure the reliability of findings are often expensive, difficult to accomplish, and rarely accompanied by the prestige accorded original studies. Hence original research must carry facial guarantees that the results reported are not artifacts of chance. Relatively stringent significance levels are therefore required before a relationship may be assumed to exist and a theory's support acknowledged.[53]

The values of social science, however, are not the values of the law. When the Supreme Court rejects a constitutional attack on six-member juries partly on the ground that such a shift will not change trial results, surely the Court ought to be more concerned with type II error, the possibility that available research has failed to reveal true differences between the verdicts rendered by different size juries, than with type I error, the possibility that reported size effects do not in fact exist.[54] Legally, the argument is that the

---

52. Although the conventions of social science describe a behavioral norm, researchers may reject them in some circumstances. Hans Zeisel has suggested that researchers interested in generating hypotheses be more venturesome in interpreting data and less captivated by conventional levels of statistical significance. *See* Zeisel, *The Significance of Insignificant Differences,* 19 Pub. Opinion Q. 319, 320-21 (1955). One problem is that it takes a certain degree of sophistication and confidence in what one is doing before one feels justified in disregarding what is conventional about a discipline. To judge by the published literature, much law-related research is done by individuals lacking sophistication or self-confidence. The danger of giving undue respect to conventional significance levels is not one that inheres in the research problem, but one that is very likely to occur in practice.

53. Ironically, the social science professions are structured so as to increase the probability that a study reporting a significance level will be the aberrant one of twenty. Researchers are unlikely to publish, or be able to publish, their results unless significance levels of about .05 are attained. Yet, if enough researchers work on the same problem, it is likely that at least one will achieve results significant at about .05 by chance. *Cf.* Good, *Fallacies, Statistical,* 4 INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES 292, 296 (1968).

54. Professors Zeisel and Diamond comment:

framers of the Bill of Rights contemplated a body of twelve when they used the term "jury." Despite the Supreme Court's equivocation on this point in *Williams*, the history seems clear.[55] Therefore,

> We do not quite share your views on "significance." Whatever conventions exist, and whatever the logical explicit rationale—the issue which you claim to address (but do not focus on) is: whether or not the Court should act, in one direction or the other, following the significance calculus. But this is a practical question which depends on a cost-benefit accounting of what would be lost (in justice, money, time, etc.) if the decision were in error. It is the combination of the significance calculus and the cost-benefit analysis that should form the proper basis for the decision.

Letter, *supra* note 37. I do not disagree with the conclusions that Zeisel and Diamond reach. *Cf.* text following note 181 *infra*. However, my point in this discussion is somewhat different from the one they suggest. I am not stating how a court "should" act; rather, I am commenting on the kinds of data that researchers less sophisticated than Zeisel and Diamond are likely to present to courts and the way in which many jurists are likely to react to the statistical results presented to them.

55. Justice White, writing for the majority in *Williams*, briefly traced the history of the jury in an effort to suggest that the framers of the Constitution and the Bill of Rights were not *necessarily* contemplating a body of twelve when they used the term "jury." Justice White's discussion concludes:

> We do not pretend to be able to divine precisely what the word "jury" imported to the Framers, the First Congress, or the States in 1789. *It may well be that the usual expectation* was that the jury would consist of 12, and that hence, the most likely conclusion to be drawn is simply that little thought was actually given to the specific question we face today. But 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.

399 U.S. at 98-99 (emphasis added) (footnote omitted). The use of the equivocating words in italics, like the use of the term "experiments" to refer to the research cited in *Williams*, *see* text at notes 9-10 *supra*, is disingenuous at best. A more accurate statement would have been that the framers almost certainly contemplated a body of ˙ twelve when they used the term "jury." Jury size was infrequently and incidentally mentioned in the debates surrounding the adoption of the Constitution and Bill of Rights, but that is probably because the fact that juries were bodies of twelve was so universally assumed as to go almost without saying. Thus, most references to the jury by the framers do not mention size. There was, however, one explicit comment. Governor Randolph, speaking in the Virginia ratifying convention about the article III guarantee of jury trial in criminal cases, stated: *"There is no suspicion that less than twelve jurors will be thought sufficient."* 3 ELLIOTS DEBATES 467 (emphasis added). For further remarks indicating that the framers implicitly assumed that the petit jury would be made up of twelve, see the remarks by Governor Randolph, *id.* at 469, Patrick Henry, *id.* at 544, and Edmund Pendleton. *Id.* at 547. In the North Carolina ratifying convention there were some remarks that suggested that the mode of trial by jury mandated by the Constitution would be somewhat different in the different states, but there is no suggestion that the size of the jury might be other than twelve. *See* the remarks by Archibald Maclaine, 4 *id.* at 176, and Richard Spaight. *Id.* at 208.

James Wilson, one of six men to sign both the Declaration of Independence and the Constitution, a contributor to the Constitutional Convention second only to Madison, one of the original appointees to the Supreme Court, and one of the nation's first law professors, began a lecture on the jury by stating: "When I speak of juries, I feel no peculiar predilection for the number twelve: a grand jury consists of more, and its number is not precisely fixed." 2 THE WORKS OF JAMES WILSON 503 (R. McCloskey ed. 1967). However, his subsequent comments make it clear that, whatever leeway he would allow grand juries, he conceived of the petit jury as a twelve-member body. He writes, "What is a verdict? It is the joint declaration of twelve jurymen upon their oaths," *id.* at 524, and, "My theory is shortly this. To the conviction of a crime, the undoubting and the unanimous sentiment of the twelve jurors is of indispensable necessity." *Id.* at 528. *See also id.* at 506, 515.

those who argue that jury size is not defined constitutionally, be-
cause size does not affect verdicts and hence has no relationship to
sixth and seventh amendment values,[56] should have the burden of

---

Further evidence that the framers assumed that the term "jury" referred to a body
of twelve is found in the writing of Blackstone. Blackstone, in the most influential
legal treatise in colonial America, wrote: "[N]o man should be called to answer . . .
for any serious crime, unless . . . the truth of every accusation . . . should . . . be
confirmed by the unanimous suffrage of twelve of his equals and neighbors . . . ." 4
BLACKSTONE, COMMENTARIES *349-50. Blackstone also wrote that "[w]hen the trial is
called on, the jurors are to be sworn . . . to the number of twelve . . . ." *Id.* at *352.

The majority in *Williams* supports its argument for a distinction between the com-
mon-law and constitutional conceptions of juries by arguing that "even though the
vicinage requirement was as much a feature of the common-law jury as was the 12-
man requirement, the mere reference to 'trial by jury' in Article III was not inter-
preted to include that feature." 399 U.S. at 96. A major difficulty with the majority's
argument is that, unlike the situation with respect to the twelve-member feature, it
does not appear that there was substantial agreement among the states as to what a
vicinage requirement would mean, because the states differed in the degree to which
they required jurors to be drawn from the neighborhood or county in which the
crime had occurred. Whatever the common law of England, it does not appear that
there was any uniform rule in the original states with respect to vicinage at the time
the Constitution was written or the Bill of Rights enacted. Consider, for example,
Madison's comments in a letter to Edmund Pendleton:

> [The Senate] are . . . inflexible in opposing a definition of the *locality* of Juries.
> The vicinage they contend is either too vague or too strict a term, too vague if de-
> pending on limits to be fixed by the pleasure of the law, too strict if limited to
> the County. It was proposed to insert after the word juries—"with the accustomed
> requisites"—leaving the definition to be construed according to the judgment of
> professional men. Even this could not be obtained. *The truth is that in most of
> the States the practice is different, and hence the irreconcilable difference of ideas
> on the subject.*

5 DOCUMENTARY HISTORY OF THE CONSTITUTION OF THE UNITED STATES OF AMERICA 211
(1905) (emphasis in last sentence added). The majority in *Williams* reproduces this
portion of the Madison letter. 399 U.S. at 95-96. However, the italicized sentence is
replaced with ellipses. The majority in *Williams* does cite in a footnote an earlier let-
ter of Madison to Pendleton, in which Madison wrote: "In many of the States, juries,
even in criminal cases, are taken from the State at large; in others, from districts of
considerable extent; in very few from the County alone. Hence a dislike to the re-
straint with respect to *vicinage*, which has produced a negative on that clause . . . ."
399 U.S. at 95 n.39, *quoting* Letter from James Madison to Edmund Pendleton, Sept.
14, 1789, in 1 LETTERS AND OTHER WRITINGS OF JAMES MADISON 491 (1865).

Significantly, the Supreme Court in prior decisions had always found that the term
"jury" as used in the sixth amendment referred to a jury of twelve. *See, e.g.,* Patton
v. United States, 281 U.S. 276 (1930); Thompson v. Utah, 170 U.S. 343 (1898).

Although I feel that White's historical analysis in *Williams* was colored so as to
lend undue support to the majority's opinion, my conclusion, based on what I believe is
a fairer reading of history, is not far from the Court's conclusion. Even if the framer's
did conceive of juries as bodies of twelve, I see no reason, given that the framers chose
not to specify the number of jurors, why jury size cannot be constitutionally reduced
*so long as the values protected by the sixth and seventh amendments are not thereby
threatened.* My reading of history does lead me to disagree with White in one im-
portant particular, however. Since it was almost certainly contemplated at the time
the sixth and seventh amendments were written that the juries to which they referred
would be made up of twelve, those who argue for a reduction in jury size should be
required to prove by the weight of the available evidence that a change in jury size
does not threaten any of the interests that those amendments were designed to pro-
tect. If the evidence is so ambiguous that one cannot honestly reach such a conclu-
sion, no change should be allowed at this time.

56. This is obviously a non sequitur, although I think it reflects the logic of many

empirically proving the lack of relationship. Instead, the uncritical use of significance tests in jury-size research puts a heavy burden of proof on partisans of the status quo.[57]

Researchers may compound the problem by reporting results that do not reach conventional levels of statistical significance as "not significant." Jurists or legislators may be understandably misled. They may mistakenly read the term "significance" in its ordinary rather than statistical sense and conclude that it has been affirmatively demonstrated that there is no difference in verdicts rendered by different-sized juries.[58]

---

who argue that jury size is, within limits, constitutionally irrelevant. For a discussion of other values that the sixth and seventh amendments are arguably designed to promote, see note 209 *infra*.

57. If, for example, research reveals that verdict differences associated with jury size are significant at the .20 level, the odds are four to one that the observed differences are not due to random factors. Yet from this data many researchers would uncritically conclude that there is no significant difference in the verdicts of different-sized juries.

58. Two other problems with the use of significance tests in jury size (and other) research are worth mentioning, although they do not favor findings of "no discernible difference." The first is whether significance tests are appropriate at all absent random sampling procedures. Strictly speaking, for example, none of the real-world studies discussed in *Colgrove* were based on a random sample of cases. They all involved populations of cases defined by jurisdiction and time period. Some would argue that this means that tests of statistical significance are simply inappropriate. Others would present two justifications for using these tests: First, cases across jurisdiction or time period or both are in the aggregate so similar that the cases studied resemble a random sample of this larger aggregate. Second, the significance tests are merely being used to assess the probability that random factors other than sampling might lead to differences as large as the difference observed. The second argument is clearly tenable, the first more doubtful. Regardless of the justification, one should be aware of exactly what the significance test is testing. For a discussion of these issues, see D. MORRISON & R. HENKEL, THE SIGNIFICANCE TEST CONTROVERSY (1970).

A second general problem has to do with the "significance" of significance tests. Significance levels depend directly on sample size. With large enough samples one usually can find a statistically significant relationship between any two variables. Most such associations will be completely unimportant for any scientific or practical purpose. Given a statistically significant relationship, importance is better judged by the strength of the association revealed. Except where correlation coefficients are involved, research published in law reviews rarely includes measures of strength of association. One problem with such measures is that except at the extremes, many of the most common nonparametric measures have no clear intuitive meaning. Whether or not measures of association are presented, the important issue in determining practical significance will often be the strength of the relationships revealed. One must be sensitive to this question and not accept a result as important simply because it is statistically significant. In many cases, a percentage distribution will provide an adequate means for assessing the importance of results. If, for example, in a very large sample it was discovered that twelve-member juries hung 5.001 per cent of the time while six-member juries hung 5.0 per cent of the time, the difference of .001 per cent might attain a high level of statistical significance yet be reasonably dismissed as unimportant. On the other hand, had the difference been 3 per cent it might have been thought important, although, if the sample were sufficiently smaller, the attained significance level might not be as high as it was in the first case. For a discussion of measures of strength of relationship, see, e.g., H. BLALOCK, *supra* note 47, at 225-34.

E. *The Real Issue*

The argument thus far is that for the following reasons the results of apparently sound methods of empirical research are likely to suggest that jury size has no discernible effect on jury verdicts, even if size does in fact determine verdicts in some cases: First, unless the researcher is able to identify *a priori* those few cases that are good candidates for size effects, differences attributable to size are likely to be so diluted in the sample studied as to appear non-existent or unimportant; second, verdict differences attributable to size are likely to be disguised if comparisons are made on an aggregate basis; third, conventions of statistical analysis will lead researchers to be overly cautious in interpreting findings that indicate possible size-associated differences in jury verdicts.

Of these three potential problems the first is in a sense the most fundamental, for it infects the other two. Findings of no aggregate size effects or of statistical insignificance become more likely when the cases studied include many in which no size effect can be expected. Yet one might reasonably object that the labored analysis suggesting that only fifteen per cent of all cases are candidates for size effects proves too much. That analysis, if accepted, proves that jury-size differences will affect at most about one sixth of all verdicts rendered. If this is so, does it not support the Court's decision in *Williams?* If even the most sensitive research is likely to show that the proportion of cases in which verdicts are affected by jury size is small, why should not the system opt for the sure savings engendered by smaller juries?[59]

This would be a strong argument if size effects, where they exist, are unrelated to other values. If one cannot say that in cases of size-linked disagreement the verdicts of twelve are likely to be "better" with reference to some value than the verdicts of six, what apart from history argues in favor of twelve?[60] If, on the other hand, the divergences between six- and twelve-member juries are systematically related to other values, there may be important reasons to continue with the larger jury.[61] If, for example, verdicts of six are more likely

---

59. Zeisel and Diamond estimate that the federal court system would save about four million dollars annually if all civil juries were reduced from twelve to six. Zeisel & Diamond, *supra* note 10, at 294. If federal criminal juries and all state juries were reduced to six, these savings would be multiplied several times over.

60. I suggest later that some other factors might exist. *See* note 209 *infra.*

61. It may be that the cases of divergence are disproportionately those that motivated the framers to guarantee a right to jury trial in the first place. For example, one might ask why there is a jury system in criminal cases at all. Why not just use the judge? Verdicts would be different in fewer than one third of the cases heard, and, given plea bargaining, this represents only a small fraction of the cases reaching the

to reflect prejudice against minority groups than verdicts of twelve, a strong argument exists for retaining the larger jury despite the extra costs involved. A small percentage of cases may still be an absolutely large number.[62] In the next section I will draw on statistical theory and group dynamics research to speculate on the ways in which verdicts of six are likely to differ from those of twelve. This should provide some insight into the values affected by changes in jury size and will suggest designs for jury research that should reveal size effects that do in fact exist.

II. Verdict Differences and Jury Size: The Direction of Effects

A. *Bases for Analysis*

In this section I will try to pinpoint some of the reasons why size differences may be expected to lead to verdict differences and suggest the directions that these differences are likely to take. In this effort I draw on common sense, sampling theory, and the social-psychological literature on behavior in small groups. Reliance on the last two of these poses problems. The sampling model postulates that jurors are selected by some independent random process. Although one may assume that a particular venire is selected from a given population at random,[63] it is clear that chance does not en-

_____

courts. One answer is that the differences between judge and jury are highly directional. The tendency of juries to be more lenient and to interpret laws from the public's perspective may be an important reason why the framers thought it necessary to guarantee a right to jury trial in the Constitution. *Cf.* Sax, *Conscience and Anarchy*, 57 Yale Review 481 (1968). The discussion thus far suggests that the twelve-member jury is more likely than the six-member jury to promote important values in its decision-making. This is, of course, an empirical question. It is possible that the six-member jury has important advantages over the twelve-member jury with respect to important values. To the extent it does, the case for six is considerably strengthened.

62. The authors of The American Jury, *supra* note 7, at 12, noted that in 1955 alone 60,000 criminal jury cases were tried to a verdict. A 5 per cent rate of difference in decisions by six- and twelve-member juries would produce 3000 cases in which size affects verdicts; a 15 per cent rate would produce 9000 such cases. Thus, over a period of years and including civil cases, a small percentage difference in results due to jury size will reflect an absolutely large number of cases.

63. Discrimination against a group in jury selection may occur at two stages. The group may be underrepresented in the population from which the jury is drawn or the selection of jurors from a particular population may be conducted in such a way that members of certain groups serve less frequently than their proportion in the population would indicate. (The word "population" is used here and in this paper to refer to the list of names from which the court official selects the venire for a term; it does not refer to the general population of an area, although the characteristics of prospective jurors as a group should be more or less congruent with the characteristics of the area's adult population.) Discrimination in the first sense would be illustrated by procedures that excluded all blacks in the district from the population from which jurors were drawn. Discrimination in the second sense would be illustrated by a procedure in which a court clerk failed to summon any blacks whose names were

tirely determine the final composition of a sitting jury. Judges may exclude individuals for hardship or for cause, and attorneys may exclude potential jurors on a peremptory basis. Nevertheless, I believe that the arguments based on sampling theory remain sound.

Exclusions by the judge for cause or at the prospective juror's initiative may be thought of as refinements in the population from which jurors are selected[64] rather than distortions of a random sampling process. Excuses by the judge for cause promote the ideal of the unbiased jury,[65] and excuses for hardship are in no obvious systematic way related to the likely attitudes of the excused individuals. The allowance of peremptory challenges poses more difficulty since lawyers challenge those jurors whom they view as likely to decide against them. This suggests that challenges might nullify an important difference between six- and twelve-member juries— the higher probability that the latter will contain individuals with viewpoints held by a minority of the population. If the

---

drawn. The statistical model to be developed makes certain assumptions about the population or jury list from which a venire is drawn; hence the argument based on the model is unaffected by whether the population with the assumed characteristics represents a fair or biased selection from some larger population of individuals. The model does assume that selection for actual jury service from the jury list will be by some independant random procedure. Conclusions drawn from the model are threatened to the extent that deviation from such procedures with respect to relevant characteristics occurs. It does not matter whether deviations from chance are the results of invidious discrimination, such as the failure to call any blacks, or permitted procedures, such as the peremptory challenge. For a discussion of the statistical analysis of jury discrimination, see Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases,* 80 HARV. L. REV. 338 (1966).

64. If one could identify in advance those individuals so involved in or opinionated about a case that they could not decide fairly or those so inconvenienced by jury duty that it would be a hardship for them to serve, one would exclude these individuals from the list of potential jurors. This happens with certain classes of individuals, such as convicted felons, who are excluded for reasons unrelated to their ability to decide fairly. While such exclusions make the jury less representative of the larger community, they do not remove a sufficient percentage of those representing relevant minorities to invalidate assumptions of substantial minority representation in the population. *See* Tables One & Two *infra,* where it is assumed that relevant minorities may constitute from zero to fifty per cent of a population. If one were to assume, for example, that all blacks would be excused on request because they could not afford transportation expenses to court, then arguments based on these tables with respect to a black point of view would not hold.

65. If the biased or involved individual is more likely to sway a jury or to be unaffected by the facts, his removal should lead to greater verdict similarity between different-sized juries, since a twelve-member jury is more likely than a six-member jury to contain such individuals. To the extent that jurors challengeable for cause are discovered and excluded, a potential normative advantage of the six-member jury is eliminated. However, it is clear that not all such jurors are discovered and removed. *See, e.g.,* Broeder, *The Impact of the Vicinage Requirement: An Empirical Look,* 45 NEB. L. REV. 99 (1966); Broeder, *Voir Dire Examinations: An Empirical Study,* 38 S. CAL. L. REV. 503 (1965).

"minority"[66] is so small that it is unlikely to be represented on the six-member jury, the twelve-member jury will rarely have so many minority representatives that they could not all be peremptorily removed.[67] However, so long as the ratio of permitted challenges to jury size remains constant, the effect of peremptories is mitigated by their availability to both parties and by the fact that challenged jurors will be replaced randomly. Furthermore, members of certain recognizable minorities may not be challenged, either because no connection is perceived between the juror's minority status and his likely decision or for other tactical reasons.[68] Most importantly, many factors likely to affect juror decision-making are not or cannot be discovered during voir dire, and so cannot be the basis of a peremptory or for-cause challenge.[69] These nonrandom elements in the final selection of the trial jury do mean, however, that the apparent precision of a statistical analysis based on the assumed random sampling of jurors is false; hence the statistical analysis should be taken as no more than an indication of the likely direction of jury-size effects and a loose estimation of their magnitude.

Different problems exist in drawing on small-group research for clues as to likely differences in the dynamics of different-sized

---

66. The term "minority" is used herein to refer to all sorts of minorities (*e.g.*, racial minorities, opinion minorities, the minority who have been in an auto accident).

67. The distorting effect of peremptory challenges on the statistical arguments with respect to minority representation increases with the absolute number of allowed challenges. However, if the same number of challenges is allowed before juries of six as is allowed before juries of twelve, reported size effects are likely to be exaggerated in certain circumstances.

68. Broeder, 38 S. CAL. L. REV., *supra* note 65, at 505, reports that some lawyers felt that challenging veniremen would irritate those finally selected as jurors, and that some jurors disapproved of challenges to blacks on the apparent basis of race. *Id.* at 526. Personal injury lawyers apparently face particular difficulty when they must decide whether to challenge the obviously handicapped.

69. Voir dire may fail to uncover information that might be revealed by more skilled questioning or honest answering, *see, e.g.*, Broeder, 45 NEB. L. REV. 99, *supra* note 65; Broeder, 38 S. CAL. L. REV. 503, *supra* note 65, and the process may not be well-suited for the discovery of character traits likely to affect juror decision-making. For example, authoritarianism or sex-role identification, traits unlikely to be revealed except through psychological testing, have been found to affect the decisions of individuals faced with trial-type problems. *See, e.g.*, Boehm, *Mr. Prejudice, Miss Sympathy and the Authoritarian Personality: An Application of Psychological Measuring Techniques to the Problem of Jury Bias*, 1968 WIS. L. REV. 734; Lipsitt & Strodtbeck, *Defensiveness in Decision Making as a Function of Sex-Role Identification*, 6 J. PER. & Soc. PSYCH. 10 (1967); Mitchell & Byrne, *The Defendant's Dilemma: Effects of Jurors' Attitudes and Authoritarianism on Judicial Decisions*, 25 J. PER. & SOC. PSYCH. 123 (1973); Rokeach & Vidmar, *Testimony Concerning Possible Jury Bias in a Black Panther Murder Trial*, 3 J. APP. SOC. PSYCH. 19 (1973). Interestingly, a team of lawyers and social scientists collaborating on jury selection in the trial of the "Harrisburg 7" identified as generally favorable two jurors who together hung the jury against acquittal on several counts. Schulman, Shaver, Colman, Emrich & Christie, *Recipe for a Jury*, 6 PSYCH. TODAY 37 (May 1973).

groups. Subjects of such research are typically college students[70] motivated by small monetary payments or course requirements. They are given limited time to solve problems quite different from those faced by jurors. Furthermore, much of the literature relevant to size effects deals only with gross differences between group and individual decision-making, while research designs that do vary size are frustratingly often limited to groups of between two and six or seven members. Size effects reported in such research may not hold for groups of six and twelve members. Finally, inconsistencies among studies raise questions about the reliability of certain findings.

Although the tentative nature of any conclusions based on small-group research must be emphasized, these studies nevertheless form the basis for the most educated guesses one can make about likely differences between the dynamics of six- and twelve-member groups. Some findings in particular have been replicated in enough different circumstances that they appear to provide relatively reliable information about the ways in which individuals act in groups;[71] heavy reliance is placed on such findings in the discussion that follows.

At a conceptual level, one may distinguish between the way in which numbers affect the membership of different-sized groups[72] and the way in which numbers affect interaction in different-sized groups.[73] In practice the distinction is difficult to maintain, because there is no neat way to separate the extent to which the dynamics of different-sized groups turn on composition from the extent to which they turn on other factors that are a function of numbers. Neverthe-

---

70. Such groups pose particular problems because they are likely to be relatively homogeneous in social status, and some are sexually homogeneous as well. Some of the best available studies, involving individuals actually called for jury duty who listened to tape-recorded rather than actual trials, indicate that differences of social status and sex are strongly associated with the roles that individuals play in jury deliberations. *See* R. SIMON, *supra* note 32, at 113-19; C. Hawkins, *supra* note 32, at 30-41; James, *Status and Competence of Jurors*, 64 AM. J. Soc. 563 (1959); Strodtbeck, James & Hawkins, *Social Status in Jury Deliberations*, 22 AM. Soc. REV. 713 (1957); Strodtbeck & Mann, *Sex Role Differentiation in Jury Deliberations*, 19 SOCIOMETRY 3 (1956).

71. *E.g.*, the body of work on conformity inspired by the classic studies of Sherif, *Group Influence upon the Formation of Norms and Attitudes*, in READINGS IN SOCIAL PSYCHOLOGY 77 (I. Newcomb & E. Hartley eds. 1947), and Asch, *Effects of Group Pressure upon the Modification and Distortion of Judgments*, in GROUP DYNAMICS 189 (2d ed. D. Cartwright & A. Zander 1960).

72. For example, numbers may result in differences in the incidence of minority representation.

73. For example, a larger group must be arranged in a physically different way than a smaller group, and such differences might have ramifications for patterns of interaction. *See, e.g.*, Strodtbeck & Hook, *The Social Dimensions of a Twelve-Man Jury Table*, 24 SOCIOMETRY 397 (1961).

less, this discussion will loosely follow such a distinction. First, I will discuss differences in jury composition associated with size and their dynamic implications. Then I will focus on size effects as they more generally affect the quality of group decision-making.

B. *The Likelihood of One Minority Juror*

For present purposes, the most important distinctions between six- and twelve-member groups drawn randomly from the same population are (1) that the members of the larger group are likely to be more broadly representative of the population from which the groups are drawn, and (2) that the average characteristics of the larger group are likely to approach more closely the average characteristics of the population from which the groups are drawn. The greater representativeness of the larger group results from the obvious fact that an individual with a given characteristic is more likely to appear in a random collection of twelve individuals than in a random collection of six.[74] Thus, the presence of jurors with viewpoints, abilities, quirks, or racial identities that characterize only a minority of the population is more likely with larger juries.[75]

Table One presents the probabilities for juries of six and twelve that no jury members will have a characteristic found among a specified percentage of the population. Over most of the range of population percentages the differences between probabilities for the two sizes are substantial.[76]

The meaning of figures in the Table can be illustrated by an example. Let us assume the minority characteristic to be race and the population figures to represent the percentage of black individuals

---

74. This greater probability is a function of the fact that an individual with a given characteristic has twelve chances to appear in a group rather than six.

75. For the same reason, factors that characterize a majority of the population are more likely to be found among members of a twelve-member jury than among those on a six-member jury; an all-black jury in a jurisdiction that is 90 per cent white is more likely if the jurisdiction seats juries of six than if it seats juries of twelve. However, since such unrepresentative juries are very unlikely with juries as large as six, the textual discussion will focus on factors that characterize only a minority of the population.

76. The calculations presented in Table One and in Tables Two and Three *infra* assume independent random sampling with replacement. (The fact that the replacement assumption does not hold with respect to juries is not important because the population from which jurors are sampled is large enough so that lack of replacement does not affect the calculations.) The formula used is the binomial expansion:

$$(p + q)^n = \sum_{k=0}^{n} (n^n - k)p^k q^{n-k},$$

where p = probability of a member of the population having the characteristic; $q = 1 - p$ = probability of a member not having the characteristic, and n = sample size.

TABLE ONE

PROBABILITIES FOR JURIES OF SIZES SIX AND TWELVE THAT NO JUROR WILL BE SELECTED WITH
A CHARACTERISTIC SHARED BY GIVEN PERCENTAGES OF INDIVIDUALS IN THE POPULATION FROM
WHICH THE JURY IS DRAWN

Probability of no juror being chosen with particular characteristic		
Percentage of individuals sharing characteristic in the population	Six-member juries	Twelve-member juries
55-100%	0.00	0.00
50	.016	0.00
45	.028	.001
40	.047	.002
35	.075	.006
30	.118	.014
25	.178	.032
20	.262	.069
15	.377	.142
10	.531	.282
5	.735	.540
0	1.00	1.00

in the population from which the jury is drawn.[77] In a population forty-five per cent black, one would expect about twenty-eight six-member juries in every thousand to have no black members; only one in a thousand twelve-member juries would contain no blacks. If blacks constitute twenty-five per cent of the names on the jury list, about one out of six six-member juries would have no black members; this compares to about one out of thirty for juries of twelve. Finally, if blacks constitute only ten per cent of the population, slightly more than half of the six-member juries would be without blacks, compared to a little more than a fourth of the twelve-member panels.

It may make both common and constitutional sense to state, as the Supreme Court has, that so long as the procedures by which juries are selected are fair and nondiscriminatory, minority groups are not entitled to any specific degree of proportional representation;[78] nevertheless, it should be recognized that a reduction in jury size threatens some of the same values as are threatened by overt discrimination.[79]

---

77. Because blacks are likely to be underrepresented in the sampling frame from which jurors' names are drawn (*e.g.*, voting lists), the degree of underrepresentation apparent to black parties before courts is likely to be greater than these figures suggest.

78. *See, e.g.*, Brown v. Allen, 344 U.S. 443, 471 (1952); Neal v. Delaware, 103 U.S. 370, 394 (1880).

79. Justice Black, writing for the majority in Smith v. Texas, 311 U.S. 128 (1940), stated: "It is part of the established tradition in the use of juries as instruments of public justice that the jury be a body truly representative of the community. For racial discrimination to result in the exclusion from jury service of otherwise qualified

While trials in which the presence of a single black juror would affect the verdict may be rare, common sense suggests at least two such situations. The first occurs when the presence of a black might inhibit other jurors from expressing prejudices that if expressed might influence the deliberative process.[80] A remark such as, "After all, he's black; he's probably committed some other crime if not this one" could 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. If the remark were not made, potential holdouts might not be influenced by the idea behind it, even if they privately believed that the defendant's race made it likely that he had committed other crimes. The possibility of resolving a difficult decision-making problem by this kind of rationalization might not occur to some jurors who would assent to the remark if it were made. Or, each juror, so long as he thinks he alone holds the belief in question, might be reluctant to act on it.[81]

The second situation in which the presence of at least one black might influence the jury verdict is where the black juror possesses expertise that allows him to give other jurors important information. Broeder, for example, reports that in one case he studied the ability of a black juror to explain why a black youth might flee from the police even if innocent may have influenced the jury's decision.[82] In other cases, the interpretation of argot may be important.[83]

---

groups . . . is at war with our basic concepts of a democratic society and a representative government." 311 U.S. at 130. As has been stated, the representative nature of the jury is decreased when it is reduced in size from twelve to six. This very fact makes it more difficult to prove statistically that a set of juries with little or no minority representation was chosen by a process that was discriminatory, because the probability that such a set of juries would be chosen by chance increases when jury size is reduced. For a discussion of the statistical proof of jury discrimination, see Finkelstein, *supra* note 63.

80. In a series of post-verdict juror interviews, Broeder found that the presence of a black woman on a jury did not prevent two jurors from voicing strong anti-black sentiments. However, the trial referred to occurred in 1954 in an Illinois city that was segregated in some respects; it appears much less likely that racial prejudices would be expressed so freely in front of blacks today. Even in the trial studied, the presence of the black woman may have been crucial. She berated the men because of their statements, and some jurors, including at least two who were strongly prejudiced, voted for acquittal because they did not want to be associated with these men. Broeder, *The Negro in Court*, 1965 DUKE L.J. 19, 23.

81. Sociologists term a situation in which each member of a group mistakenly believes that he alone has a particular belief or engages in a particular kind of activity a situation of "pluralistic ignorance." Action in such situations differs from action that would be expected if the individuals were aware of the beliefs or acts of others.

82. Broeder, *supra* note 80, at 30.

83. In the trial of Huey Newton, the meaning within the black community of the phrase "take care of business" was apparently of some importance. Address by Herman

While there is a danger that the presence of a single black would cause other jurors to defer too much to one who dishonestly or mistakenly purports to have cultural expertise, this danger is probably less than that of misunderstanding a peculiar feature of black culture. Furthermore, the potential danger of deference may well increase with smaller juries. The best counter to an unwarranted prima facie claim to expertise is the presence of another individual with a similar prima facie claim but with a different viewpoint. The twelve-member jury is more likely than the six-member jury to have two or more blacks to correct or corroborate each other. Indeed, so long as black representation in the jury population is much above ten per cent, the chances of finding only one black on a jury are greater with six than with twelve.[84]

If my argument about the role that blacks might play in jury deliberations is correct, there may indeed be differences in the results reached by different-sized juries, and in situations of difference the retention of the twelve-member jury would promote informed, unbiased fact-finding. But my choice of blacks as an example may strike some as biasing the value conclusion. Just as juries of twelve are more likely than those of six to have black members, so are they more likely to have bigoted members; just as they are more likely to have individuals with helpful personal experiences, so are they more likely to have individuals who are beyond all reason.[85] Clearly any scheme that increases the probability of minority representation on juries is a two-edged sword with respect to values important to the legal system. It does not

---

Blake, black sociologist who testified at the Huey Newton trial, National Conference of Kent Fellows, Columbia, Mo., August 1968.

84. The probabilities of finding *only* one minority group member on six- and twelve-member juries are as follows:

Proportion minority representation in population	Probability of one and only one minority juror	
	Six-member juries	Twelve-member juries
.50	.093	.003
.45	.136	.007
.40	.186	.018
.35	.244	.036
.30	.302	.071
.25	.356	.126
.20	.393	.211
.15	.399	.301
.10	.355	.377
.05	.232	.342

85. For example, an individual who believed that God had ordained a conviction apparently prevented acquittal of the "Harrisburg 7" and hung the jury on certain counts. *See* Schulman, Shaver, Colman, Emrich & Christie, *supra* note 69.

follow, however, that the edges are equally sharp. For several reasons, the benefits of the increased representation of minorities with favored viewpoints that the choice of twelve entails should far outweigh any detriment associated with the increased representation of minorities holding disvalued positions.

First, many of the most disvalued minority positions are likely to be so uncommon that they will rarely be represented even on a jury of twelve. While six-member juries would be even less likely to contain such members, the absence of widespread stories about juries being hung by bigots or irrational persons, for example, suggests that few such individuals are ever seated even on twelve-member juries.[86] Not only is the incidence of such individuals likely to be low in the general population, but if called for jury duty such individuals, if spotted, may be challenged for cause on voir dire.

Second, an individual holding a normatively disvalued position, such as a bigot, might be reluctant to voice his views to strangers, probably would have his position refuted if stated, and might well not vote in accord with his predilections either because of pressure from his fellow jurors or because he realizes that in the context of a jury trial such a vote is inappropriate.[87] In addition, the mere presence of such an individual is unlikely to dissuade others from voicing generally approved positions, while the mere presence of other visible minorities, such as blacks, might keep individuals from voicing normatively disapproved positions.

Finally, the twelve-member jury is less likely to be completely aberrant than the six-member jury in that it is less likely to be entirely or overwhelmingly composed of those who represent disfavored positions in the community.[88] Since one would expect those whose opinions or attitudes are well anchored in community values to be particularly tenacious,[89] their presence should guarantee at

---

86. *But see* Hunter, *Law in the Jury Room*, 2 OHIO ST. L.J. 1, 18-19 (1935); note 85 *supra*.

87. Broeder, *supra* note 80, at 23, suggests that a number of prejudiced jurors consciously suppressed their biases in voting to acquit a black defendant.

88. *See* note 75 *supra*.

89. Nahemow and Bennet conducted a relevant study in a New York home for the aged. They found that those who generally conformed most to the informal residential norms of the homes were most resistant to counternormative persuasion on certain issues. These individuals generally had a greater stake in the home as a place to live. They were not disproportionately resistant to persuasion, however, with respect to matters not anchored in home norms, such as political issues or the question of which of two pictures was preferable. Nahemow & Bennet, *Conformity, Persuasibility and Counter-normative Persuasion*, 30 SOCIOMETRY 14 (1967). Similarly, Newcomb found that Bennington women who resisted the college's liberal norms tended to have strong roots in reference groups that also rejected the college's norms. Newcomb, *Attitude Development as a Function of Reference Groups: The Bennington Study*, in READINGS

least a hung jury in a situation in which most jurors are motivated by considerations that the larger community would regard as improper. A jury in which few or no members represent basic community values, though unlikely at either size, is far less likely with twelve than with six.

C. *The Likelihood of Two Minority Jurors*

It should be obvious that the argument that minorities are more likely to be represented on juries of twelve than on juries of six applies to minorities of all kinds, not just to such visible minorities as racial groups. In particular, the argument applies to attitudinal and perceptual minorities. The twelve-member jury is more likely to have one or more members who believe that high damage awards raise insurance rates or who have caught a subtle contradiction in the defendant's testimony. Individuals with unique perceptions or attitudes may disproportionately influence jury verdicts if other jurors do not hold conflicting attitudes or perceptions. However, where conflicts of attitude or perception do exist, the psychological literature provides substantial evidence that an individual in the minority is unlikely to resist group pressures to conform unless he becomes aware that at least one other member shares his position.[90]

The research supporting this conclusion has its roots in Asch's famous experiment, which became a paradigm for further research.[91] Asch required a subject to state which of three lines matched a criterion line. Subjects responding alone made virtually no errors. However, when naïve subjects responded in a group after seven others, each of whom made the same incorrect response, almost one third of the subjects responded incorrectly as well. In these conditions even those giving correct responses evidenced considerable agitation. Post-experiment interviews revealed that many of the conforming subjects had conformed only on the surface; they

---

IN SOCIAL PSYCHOLOGY 265 (3d ed. E. Macoby, T. Newcomb & E. Hartley eds. 1958). These findings also suggest that normative deviants may belong to minority groups that provide relatively strong anchorage for their values or attitudes.

90. It apparently is not unusual for jurors to vote in ways contrary to their convictions. *See generally* Note, *On Instructing Deadlocked Juries,* 78 YALE L.J. 100 (1968). Hawkins and Simon found that 12.8 per cent and 10 per cent of their jurors respectively were willing to admit that they voted against their preferred positions. C. Hawkins, *supra* note 32, at 102; R. SIMON, *supra* note 32, at 64. These figures may underestimate the percentage of unconvinced jurors because some may not have wanted to admit that they voted contrary to their beliefs. On the other hand, knowledge that they were participating in mock trials may have led some to capitulate when they would not have done so in a real trial.

91. *See* Asch, *supra* note 71. Asch was building on research reported by Sherif, *supra* note 71.

had continued to believe throughout the experiment that the group was wrong but had changed their answers because they did not want to appear different. Others actually had been persuaded; they became convinced that the group's choice, although different from theirs, was correct. To the attorney, an actual shift in perspective is worth the same as simple conformity; only overt behavior is important. Once jurors have publicly affirmed their verdicts, later attacks on the ground that one or more jurors did not agree with the verdict pronounced are unlikely to be successful.[92]

Asch refined his early results in a series of experiments that led to several findings of particular relevance to the jury-size question. First he varied the size of the unanimous group majority and found that increases in the number of majority members beyond three produced no substantial increment in conformity rates.[93] Asch also found that if he added just one true respondent to the group, the conformity rates of the naïve subjects fell off dramatically, even though the subjects continued to face absolutely large majorities against them.[94]

Of course, the decision-making problems faced by an actual jury differ substantially from those faced by Asch's subjects. In general the jury will face problems that will be more difficult to solve, in which the evidence will be more ambiguous, and in which the very

---

92. *See* Note, *supra* note 90, at 109 n.33 (collecting cases).

93. *See* Asch, *supra* note 71, at 197. Other research has generally corroborated these results. *See, e.g.,* Kiesler, *Group Pressure and Conformity,* in EXPERIMENTAL SOCIAL PSYCHOLOGY 233, 257 (J. Mills ed. 1969); Rosenberg, *Group Size, Prior Experience, and Conformity,* 63 J. AB. & SOC. PSYCH. 436 (1961). Some reported research is somewhat inconsistent. Gerard and his associates found that conformity increased in a generally linear fashion among groups of sizes two through eight, but inspection of his data reveals that the bulk of the increase in conformity occurred by the time the majority reached three. In one of the two conditions Gerard studied there was no apparent consistent linear increase after the majority reached three. Gerard, Wilhemy & Conolley, *Conformity and Group Size,* 8 J. PER. & SOC. PSYCH. 79 (1968). Goldberg, *Three Situational Determinants of Conformity to Social Norms,* 49 J. AB. & SOC. PSYCH. 323 (1954), using groups of two and four, and Kidd, *Social Influence Phenomena in a Task-Oriented Group Situation,* 56 J. AB. & SOC. PSYCH. 13 (1958), using groups of two, four, and six, both reported statistically insignificant trends for conformity to increase linearly with size. The insignificance of the trend might call Asch's results into question, since Asch would predict a sharp increase in conformity between two and four. However, both of these studies used very ambiguous tasks (judging intelligence from pictures and estimating light flickers), and, more importantly, both involved a situation in which the experimenter reported an average group judgment to an individual repeating a task alone. In neither experiment was the subject confronted with fellow group members expressing their opinions and listening to his.

94. *See* Asch, *supra* note 71, at 195. This result has been corroborated by others and is generally accepted by social-psychologists. *See* Edmonds, *Logical Error as a Function of Group Consensus: An Experimental Study of the Effect of Erroneous Group Consensus upon Logical Judgments of Graduate Students,* 43 SOCIAL FORCES 33 (1964); Kiesler, *supra* note 93. *But cf.* Gorfein, *The Effects of a Nonunanimous Majority on Attitude Change,* 63 J. SOC. PSYCH. 333 (1964).

existence of a single correct answer might appear problematic.[95] Social-psychological literature suggests, however, that at least the first two of these differences should lead to stronger conformity effects than are found when subjects respond to simpler problems.[96] This is important if Asch's findings are to be generalized to the jury situation; astonishment that anyone should conform in the Asch situation sometimes leads readers to overlook the fact that only one third of the responses in Asch's experiment were conforming ones. When problems become more difficult or ambiguous, conformity rates can be double that or more.[97]

Columns one and two of Table Two present the probabilities that randomly selected six- and twelve-member juries will contain at least two individuals who share a specific perception or attitude

---

95. For a study of how jurors continuously revise their judgments as they receive new evidence in a close case, see Weld & Roff, *A Study in the Formation of Opinion Based upon Legal Evidence*, 51 AM. J. PSYCH. 609 (1938).

96. *See* Allen, *Situational Factors in Conformity*, 2 ADVANCES IN EXPERIMENTAL SOC. PSYCH. 133, 167-68 (1965); Asch, *supra* note 71, at 199; Blake, Helson & Mouton, *The Generality of Conformity Behavior as a Function of Factual Anchorage, Difficulty of Task, and Amount of Social Pressure*, 25 J. PERSONALITY 294 (1957); Coleman, Blake & Mouton, *Task Difficulty and Conformity Pressures*, 57 J. AB. & SOC. PSYCH. 120 (1958); Deutsch & Gerard, *A Study of Normative and Informational Social Influences upon Individual Judgment*, in GROUP DYNAMICS, *supra* note 71, at 211; Luchins & Luchins, *Social Influences on Judgments of Descriptions of People*, 60 J. SOC. PSYCH. 231 (1963).

One theoretical explanation for these results is the suggestion made by a number of writers that motivation to conform to group opinion comes from two distinct sources, one being normative and the other informational. *See, e.g.*, Allen, *supra*; Penner & Davis, *Conformity and the "Rational" Use of Unanimous Majorities*, 78 J. SOC. PSYCH. 299 (1969); Schulman, *Asch Conformity Studies: Conformity to the Experimenter and/or to the Group?*, 30 SOCIOMETRY 26 (1967). An individual looks to the judgment of others both to ascertain how others expect him to respond and because the reactions of others typically provide information about the nature of problems; indeed, much learning involves little more than learning how others categorize objects and events. When a problem is clear, individuals probably feel that there is little to gain from knowing how others would resolve it. When a problem is difficult, such as the problem of whether admittedly incriminatory evidence suggests guilt beyond a reasonable doubt, individuals may be more concerned with how the group expects them to behave. When evidence is ambiguous, individuals may be more willing to rely on the judgments of those who appear to find the evidence not ambiguous than would ordinarily be the case. Some theorists seeking to explain the Asch results suggest that individuals are perhaps conforming to what they believe to be the wishes or judgments of the experimenter, in addition to or instead of those of the group. *See* Luchins & Luchins, *supra* (such a perception can motivate substantial conformity to positions that would not otherwise be adopted); Schulman, *supra*. It seems probable, however, that the experimenter, like the group, is looked to only as a source of information on what responses are proper and as an expert on the reality behind the experiment.

97. For an extreme example of both ambiguity and conformity, see Jacobs & Campbell, *The Perpetuation of an Arbitrary Tradition Through Several Generations of a Laboratory Microculture*, 62 J. AB. & SOC. PSYCH. 649 (1961). Conformity may have been enhanced in this experiment, which involved the apparent movement of a stationary pinpoint of light in a dark room, because the subjects did not realize that the stimulus was almost completely ambiguous and the true answer was contrary to their perceptions.

when the population contains a given percentage of individuals who share the same perception or attitude. Generalizing uncritically from Asch's results, these are the probabilities of drawing juries in which an individual will find the support needed to maintain a minority viewpoint in the face of majority opposition.[98] At most population levels, the differences between different-sized juries are greater than they are when the probability that a single minority juror will be selected is at issue. If Asch's theory is correct, these data suggest that juries of twelve should hang substantially more often than juries of six, because minority members will more frequently find support for their dissenting viewpoints. Zeisel reports that this is in fact the case; 2.4 per cent of 290 six-member criminal juries hung compared to 5 per cent of the twelve-member juries in a larger national sample.[99]

This is a difference that even the most ardent advocate of the six-member jury should be able to discern. However, in terms of general policy the implications of this difference are far from clear. It is not obvious that the forces that cause juries to hang generally support values important to the legal system. Whether they do would depend on which minorities hang juries and why. Without knowing this it is impossible to say whether hung juries are worth the expense, delay, and uncertainty they involve. If minorities that hang juries tend to be irrational, prejudiced, or corrupted, the differential rates of jury hanging argue in favor of smaller juries. If, on the other hand, minority holdouts represent an element of community opinion particularly concerned that guilt be established

---

98. This does not mean that the minority will necessarily maintain their dissenting viewpoint permanently. They may be persuaded by rational arguments or capitulate under the pressure of having to defend a distinctly unpopular position over a long period of time. *See* C. Hawkins, *supra* note 32, at 122-50; Note, *supra* note 90, at 107-17. What it does mean is that without initial support there is little or no chance that a single minority juror will hold out against the majority. Kalven and Zeisel report that, of 155 actual jury trials they studied, in no case did a jury with only one initial dissenter fail to reach a verdict. THE AMERICAN JURY, *supra* note 7, at 462.

99. Zeisel, *supra* note 8, at 720. The 290 cases are post-1968 cases from the Miami Circuit Court in Florida; the national sample reports data collected much earlier for THE AMERICAN JURY, *supra* note 7. It does not appear likely that the time difference in data collection calls these results into question because there is no reason to believe that the rate of hung juries has decreased over the years. The nonrandom nature of the sample of cases reported in *The American Jury* might distort the national hung-jury rate, but there is no reason to expect that this is so. There is one puzzle in Zeisel's reported figure for the national sample, however. It is cited to page 56 of *The American Jury*, but on this page the hung-jury rate is reported to be 5.5 per cent, and not 5 per cent as reported in Zeisel, *supra*. If this is simply the result of a misprint in the article, the actual differences in hung-jury rates probably are greater than those reported in the text. Indeed, Kalven and Zeisel suggest that even the 5.5 per cent hung-jury rate they have calculated is likely to be an underestimate. THE AMERICAN JURY, *supra*, at 57 n.2.

TABLE TWO

PROBABILITY THAT AT LEAST A CERTAIN NUMBER OF JURORS WILL BE CHOSEN WITH A
PARTICULAR CHARACTERISTIC WHEN THAT CHARACTERISTIC IS SHARED BY GIVEN
PERCENTAGES OF INDIVIDUALS IN THE POPULATION FROM WHICH JURORS ARE DRAWN

Percentage of individuals sharing characteristic in the population	Probability of 2 or more out of 6	Probability of 2 or more out of 12	Probability of 4 or more out of 12
85-100%	1.00	1.00	1.00
80	.998	1.00	1.00
75	.995	1.00	1.00
70	.989	1.00	.998
65	.978	1.00	.994
60	.959	1.00	.985
55	.931	1.00	.964
50	.891	.997	.927
45	.836	.992	.866
40	.767	.980	.775
35	.681	.953	.653
30	.580	.915	.507
25	.466	.842	.351
20	.345	.725	.205
15	.224	.557	.092
10	.114	.341	.026
5	.033	.118	.002
0	0.00	0.00	0.00

beyond a reasonable doubt, or if they represent elements that do not share popular prejudices or misguided perceptions, their presence fosters the integrity of the trial process, and the data on hanging argues in favor of larger juries. My own hunch is that hanging minorities are of the latter sort, because I regard the perspectives of the minorities most visible in the community (*e.g.*, black people) as legitimate ones to bring to the jury room. Clearly there is a need for systematic analysis of the way minorities behave in the deliberative process.

There can be no doubt, however, that the interests of the parties to a case are differentially affected by the probability of a hung jury. Hung juries generally aid defendants. In a criminal case tried to a jury that hangs, the defendant remains unconvicted, has a chance that the prosecution will be dropped (a result that seems particularly likely if the jury favored acquittal), has had discovery of the prosecution's case, and will often be in a better position to plea bargain. In civil actions, defendants typically have "deeper pockets" than plaintiffs and are less in need of an immediate decision; thus, they are likely at a minimum to be placed in a substantially stronger bargaining position because of the delay that a new trial entails. One may argue the relative merits of giving defendants such ad-

vantages, but it appeared until *Williams* and *Colgrove* that the sixth and seventh amendments had foreclosed such arguments. These amendments apparently were intended to preserve the benefits that arise from having one's case tried to a jury rather than to a judge. One such benefit is the possibility of no verdict, an impossible result in completed bench trials. The Supreme Court ignored the matter of the hung jury in *Colgrove* and assumed away any difficulties in *Williams*.[100]

Column three of Table Two gives the probability of selecting at least four members of a twelve-member jury who share a particular perception or attitude from a population in which a given percentage of individuals share the same perception or attitude. In general, the probability of drawing at least four minority members out of twelve is somewhat less than the probability of drawing at least two out of six. Misreading Kalven and Zeisel, the Court in *Williams* thought that four dissenters would be needed on a twelve-member panel to provide the same probability of holding out or reversing the majority as two dissenters on a panel of six.[101] In his response to *Williams*, Zeisel pointed out that a key finding of the Asch experiments was that a second nonconformer was sufficient to break the hold of even a very large majority.[102] Zeisel's statistics

---

100. It is true . . . that the "hung jury" might be thought to result in a *minimal* advantage for the defendant, who remains unconvicted and who enjoys the prospect that the prosecution will eventually be dropped if subsequent juries also "hang." . . . But when the comparison is between 12 and six, the odds of *continually* "hanging" the jury seem slight, and the numerical difference in the number needed to convict seems unlikely to inure perceptibly to the advantage of either side.

Williams v. Florida, 399 U.S. 78, 101 n.47 (1970) (emphasis added).

In this quote the Court, after grudgingly acknowledging the defendant's advantage, misstates the issue. The defendant does not have to hang the jury *continuously* to gain advantages from a jury's failure to decide. Once is enough for some advantages, and very few cases are tried more than twice. Moreover, the Court, lacking data, apparently assumes that the difference between the probabilities of hanging a jury of six and a jury of twelve is slight. In fact, if Zeisel's data can be accepted, a jury of twelve is more than twice as likely to hang as a jury of six. In absolute terms, out of every 100,000 criminal trials there will be 2600 more hung juries if juries of twelve are used rather than juries of six. Kalven and Zeisel present some data that indicate that in almost two thirds of hung-jury cases the jury will have hung with a majority for conviction, THE AMERICAN JURY, *supra* note 7, at 460, but their sample size is so small that the reliability of these findings standing alone is questionable. The findings are corroborated, however, by the further finding with a much larger sample that in 80 per cent of the cases in which the jury hangs the judge would vote to convict, and in 79 per cent of the cases (hung juries excluded) in which the judge would convict the jury agrees. *Id.* at 56. Thus, it would seem that in a substantial majority of criminal cases in which the jury hangs, the defendant has avoided an immediate conviction.

101. 399 U.S. at 101-02 n.49.

102. Zeisel, *supra* note 8, at 719-20.

on hung juries[103] are predictable from Asch's findings,[104] but they would be surprising if the Court's "proportionality hypothesis" held.[105]

D. *Interverdict Stability*

Thus far the statistical analysis indicates that twelve-member juries are more likely than six-member juries to contain individuals who represent various minority groups and viewpoints. To the extent that the contributions of such individuals to the deliberative process are valuable, one would expect that where verdicts conflict, those of twelve will be "better." The statistical analyses presented below will suggest that the decisions of twelve-member juries are also likely to be "better" in two other respects. The decisions of twelve are likely to be more consistent across similar cases, and are more representative of the community in that they are more likely to reflect the decisions that would prevail if the entire community could judge the trial for itself.

---

103. *See* text at note 99 *supra.*

104. *See* text at notes 91-94 *supra.*

105. In an important recent article, Moscovici and Faucheux indicate that the conformity effects reported by Asch and others might depend on the consistency with which the experimenter's confederates report their results, rather than on the fact that they are in the majority. Moscovici & Faucheux, *Social Influence, Conformity Bias, and the Study of Active Minorities,* 6 ADVANCES IN EXPERIMENTAL SOC. PSYCH. 149 (1972). In support of this proposition, they present their own findings with a color perception problem in which a consistent minority of two were able to influence the views of four naïve subjects on the location of the border between blue and green. These results should not be seen as surprising; to some extent they were anticipated in Asch's early experiments in the less ambiguous line-judging context. Asch found, in a result that is cited only infrequently, that when one confederate in a group of sixteen made consistently wrong judgments on the line problem he was ridiculed; when as many as three of sixteen made consistent mistakes the ridicule changed to agitation, as the naïve subjects took them quite seriously. Asch, *supra* note 71, at 198. It is also interesting to note that Moscovici & Faucheux, *supra,* at 194-95, report that the degree of influence of their consistent minority was about the same as the influence of Asch's majority when one confederate broke ranks and gave true reports.

The work of Asch and his followers and the research of Moscovici and Faucheux are inapposite to the jury situation in that the experimenter determines that some individuals will remain firm in their judgments despite the arguments of others. Jurors are not committed to consistency; consistency depends on the views they present, the views of others, and on their ability to persuade or their vulnerability to persuasion on the matter in issue. The most relevant research in this area is C. Hawkins, *supra* note 32. Hawkins, working with real jurors and tape-recorded trials, found that when jury deliberation revealed the existence of opposing factions, the factions as groups were expected to take about an equal amount of time in presenting arguments for the positions they espoused. Where a split is eight to four, for example, the four minority members have to do about the same amount of talking as the eight in the majority. In these circumstances one or more members of the minority often will be unable to keep up his end of the argument and may switch to the majority. A switch, of course, puts more pressure to participate on the remaining minority members and suggests that the minority position is less consistently held, which,

The proposition that the verdicts of a group of twelve-member juries will be more consistent than those of a group of six-member juries rests on the assumption that jurors often attempt to resolve their differences by compromise, a process that tends toward the average of individual judgments. The argument thus applies most directly where the matter in issue is easily averaged—where the case involves money damages, for example—but it might also apply where a jury must resolve multiple counts or choose the appropriate level of an offense.[106]

Professor Zeisel has provided a good illustration of size effects in this context.[107] Imagine a situation in which a damage action was tried to all the members of a community, and assume that the community was split on the damage issue: One sixth of the population felt that an appropriate award was $1000, one sixth felt that it was $2000, and so on by $1000 intervals up to the most generous one sixth, who felt that damages should be $6000. If differences were then compromised by averaging the individual judgments, the final verdict would be for $3500. If instead a jury was drawn from the community and their judgments averaged, it is likely that the final judgment would differ somewhat from $3500. Zeisel has calculated

---

following the thesis of Moscovici and Faucheux, should further increase the pressure on the minority to conform.

Although Hawkins worked only with twelve-member juries, one might argue that the pressure on two in six to talk approximates the pressure on four in twelve. In both cases the average minority member is expected to talk about twice as frequently as the average majority member. Thus, defections and fatigue should reduce the apparent consistency of the minority about as much in the smaller jury as in the larger, and the Court's proportionality hypothesis might hold.

There are at least two problems with this argument. One is that four-member minorities have certain advantages over two-member minorities even if they are opposed by majorities that are proportionately as large or larger. A four-member minority provides a greater opportunity for the members to relieve each other, so that each minority member will have more time to relax and think before it is again his turn to speak. Also, some individuals have personalities that make extended argumentation particularly distressful. While a four-member minority might tolerate such an individual, his presence as part of a two-member minority would place the entire burden of the defense on his colleague. The second difficulty with the argument is its failure to explain the figures that suggest that six-member juries hang about half as often as twelve-member juries. *See* note 99 *supra* and accompanying text. If the proportionality hypothesis held, for whatever reason, the hanging rates of six- and twelve-member juries should be about the same. This suggests that an approach such as Asch's, which emphasizes the absolute number of dissenters rather than the proportion of the dissenters to the majority, provides a better explanation of what goes on in the jury room.

106. *See* Zeisel, *supra* note 8, at 716-21. For an experiment that clearly shows that jurors attempting to reconcile conflicting damages estimates average their individual estimates, see C. Hawkins, *supra* note 32, at 57-66. Hawkins used actual jurors in three cities who were near the end of their term of service.

107. Zeisel, *supra* note 8, at 716-18.

that if six-member juries were drawn in these circumstances, fifty-one per cent of the verdicts would be for between $3000 and $4000, while sixteen per cent of the verdicts would be for more than $4500 or less than $2500. If twelve-member juries were drawn, sixty-eight per cent of the verdicts would fall between the first two figures while only four percent would fall outside the latter two amounts. The greater proximity of twelve-member jury verdicts to the average judgment of the community from which they are drawn means that damages awarded by two different juries to two plaintiffs suffering similar injuries are likely to be closer in amount if the juries each have twelve members than if they each have six. In a legal system that values the similar treatment of individuals in like circumstances, this result argues in favor of the twelve-member jury.[108]

E. *Representing the Community*

Where averaging processes are inappropriate, as with the choice between a verdict of guilt or innocence on a single count,[109] there is another sense in which the verdict of the twelve-member jury is more likely to be representative of the community judgment than that of the six-member jury. Assume that in the imaginary trial for damages described above, the community members are asked to decide liability or no liability by a preponderance of the evidence.[110] In the trial situation, where one can rarely be absolutely certain that

---

108. It is impossible to state, however, what this difference in proximity means in dollar terms. This would depend on the distribution of individual judgments before the deliberations proceed to an averaging process and on the weight that would be given to individual views. Zeisel comments that whatever the final distribution of awards, the variation in the judgments of the six-member juries as measured by the appropriate statistic, the standard deviation, will be about 42 per cent greater than the variation with juries of twelve. *Id.* at 718. But this probably overstates the difference because it assumes that all will contribute equally to the averaging process and that averaging procedures will be mathematically precise. Participation tends to be more equal in smaller groups, *see* text at notes 145-51 *infra*; thus, the average verdict of a six-member group may reflect the more or less equal input of five members with the views of one ignored, while the verdict of a twelve-member jury might reflect similar inputs from nine with three members effectively excluded. Even when all jurors contribute, the averaging process may be only a loose one. *See, e.g.,* C. Hawkins, *supra* note 32, at 57-66. Of course, if the jury adopts some formal means of assuring equal inputs, such as the forbidden but not unknown quotient verdict, Zeisel's analysis of magnitude applies and the argument holds precisely.

109. Even in this case there may be some room for averaging, as when one juror argues that 95 per cent certainty is necessary for conviction beyond a reasonable doubt, another argues in favor of 85 per cent certainty, and they compromise at 90 per cent certainty. This occurrence, however, seems unlikely. Hawkins found no indications of averaging on issues other than liability. *See* C. Hawkins, *supra* note 32, at 104-50. *But cf.* Zeisel, *supra* note 8, at 721.

110. The following analysis would also apply where the jury is deciding guilt or innocence and the standard is "beyond a reasonable doubt."

a particular verdict is correct, one might argue that the most appropriate jury verdict would be the one that reflects the considered judgment of a majority of the community. Part A of Table Three presents the probabilities by jury size that a majority of the jury will favor the plaintiff for given divisions of community opinion. Part B[111] assumes that when juries are divided equally on the first ballot, all ties will be subsequently broken, with defendants and plaintiffs having an equal chance of prevailing. Accepting this assumption, and assuming that juries always decide in the direction of their first majority, Part B presents the probability of plaintiff verdicts for given divisions of community opinion.[112]

TABLE THREE

PART A

PROBABILITIES FOR JURIES OF SIX AND TWELVE THAT A MAJORITY OF THE JURY WILL VOTE
INITIALLY FOR THE PLAINTIFF WHEN GIVEN PERCENTAGES OF THE COMMUNITY FROM
WHICH THE JURY IS DRAWN WOULD VOTE FOR THE PLAINTIFF

Per cent of community that would vote for plaintiff	0	10	20	30	40	50	60	70	80	90	100
Probability that at least 4 out of 6 would vote for plaintiff	.00	.00	.02	.07	.18	.34	.54	.74	.90	.98	1.00
Probability that at least 7 out of 12 would vote for plaintiff	.00	.00	.00	.04	.15	.39	.67	.88	.98	1.00	1.00

PART B

PROBABILITIES FOR JURIES OF SIX AND TWELVE THAT A MAJORITY OF THE JURY WILL VOTE
EVENTUALLY FOR THE PLAINTIFF WHEN GIVEN PERCENTAGES OF THE COMMUNITY FROM
WHICH THE JURY IS DRAWN WOULD VOTE FOR THE PLAINTIFF. (ASSUMES JURORS DO NOT
SWITCH VOTES WHEN THERE IS AN INITIAL MAJORITY, BUT WHEN THERE IS NO INITIAL
MAJORITY THE TIE IS EVENTUALLY BROKEN WITH AN EVEN CHANCE THAT IT WILL
BE BROKEN IN FAVOR OF THE PLAINTIFF)

Per cent of community that would vote for plaintiff	0	10	20	30	40	50	60	70	80	90	100
Probability that at least 4 out of 6 would eventually vote for plaintiff	0.00	.01	.06	.16	.32	.50	.68	.84	.94	.99	1.00
Probability that at least 7 out of 12 would eventually vote for plaintiff	0.00	.00	.01	.08	.25	.50	.75	.92	.99	1.00	1.00

111. Part B is taken from Note, *supra* note 8, at 546. The interpretation given to these figures, however, differs somewhat from the interpretation given by the author of the Note.

112. *See id.* at 542-44; note 32 *supra* and accompanying text. Of course, it is possi-

Part A reveals that in a situation in which only twenty per cent of the community would decide for the plaintiff, two out of one hundred six-member juries would have a pro-plaintiff majority compared to less than one out of one hundred twelve-member juries. Where thirty per cent of the community favors the plaintiff, six-member juries are almost twice as likely as twelve-member juries to have a pro-plaintiff majority. Where the community is equally divided or a small margin favors the defendant, the twelve-member jury is more likely to have a pro-plaintiff majority because the number of jurors needed for the majority, seven, is a smaller fraction of twelve than four is of six.[113]

Justice White, writing for the majority in *Williams*, found "little reason" to think that a six-member jury would, in any "meaningful" sense, be less representative of a given community than a jury of twelve.[114] But the results presented in Table Three, as well as the analysis of jury decision-making on the issue of damages, suggests that Justice White is either wrong in his intuitive analysis or that he used the word *"meaningful"* in a very different sense than I do.[115]

---

ble that some juries with initial majorities will hang; however, this possibility is more than counterbalanced in Part A of the Table by the fact that equal initial divisions are not partially credited to pro-conviction majorities, although initially split juries will sometimes decide to convict. For a justification of crediting equal splits to defendants and plaintiffs equally, see Note, *supra* note 8, at 545-46 nn.62, 63. Part B does not take account of the possibility of hanging, but hung juries are a relatively rare event with twelve and even rarer with six. *See* text at note 99 *supra*.

113. Figures associated with community proportions that add to 100 per cent do not themselves add to 100 per cent in Part A of Table Three because situations in which the jury is equally divided are excluded from these calculations. When these figures are included and divided equally, as in Part B, the results are symmetrical and do add to 100 per cent.

114. Williams v. Florida, 399 U.S. 78, 100 (1970).

115. In fairness to Justice White, it should be pointed out that the evidence from THE AMERICAN JURY, *supra* note 7, justifying the conclusion that the jury almost always decides in the direction of an initial majority may be based largely on cases that are relatively clear, in which 90 per cent of the community would decide the same way. The majority in such cases almost certainly coincides with community feeling, and this would probably be true if juries were reduced to six. It may even be that cases in which small initial minorities hang or in which juries reverse an initial vote are disproportionately cases in which the views of the jury's minority coincide with the views held by a community majority. However, as the discussion of Asch's research suggests, such an effective minority is more likely to be found on a twelve-member jury than on a jury of six. *See* text at notes 91-99 *supra*. Further limiting this defense of White's position is the finding from one experimental study using a case that apparently was close (35 per cent of the experimental juries hung and the initial pre-deliberation vote was 67 to 33) that 74 per cent of the cases with an initial majority that reached a verdict agreed with the initial majority view. In only 15 per cent of all cases did the jury hold against an initial majority position. *See* C. Hawkins, *supra* note 32, at 145. Simon, working experimentally with another apparently close case, reports that in only 9 per cent of all cases and 11 per cent of those cases reach-

Table Three illustrates one more important point. The representative advantage of the larger jury is, except at and slightly below a fifty per cent community division, unbiased; the twelve-member jury is more likely than the six-member jury to have a pro-plaintiff majority when a majority of the community would favor the plaintiff and it is more likely to have a pro-defendant majority when a majority of the community would favor the defendant. Thus, if the community's judgment is the standard, the quality of decisions rendered by twelve is likely to be higher than the quality of those of six regardless of which party has the better case.

This suggests that in *Williams* the Supreme Court missed an opportunity to promote the integrity of jury trials generally and ignored the perspective of the larger community. The Court apparently felt that if an alleged pro-defendant benefit from trial by twelve in criminal cases was offset by an equally likely pro-prosecution benefit, the benefits canceled and there was no cause for striking down the six-member jury. Thus, the Court never considered the important possibility that the allegedly offsetting benefits *both* worked to promote the quality of jury decisions generally.[116] Although the sixth amendment gives rights to the defendant, it does not follow that these rights were not in part given to enhance the interest of the larger community in correct verdicts. The Court might have considered the relationship of jury-size differences to the common good, but it did not.

F. *Research That Favors Twelve*

Thus far I have used the social-psychological literature only to justify certain assumptions or to form a basis for making inferences about the consequences of particular statistical distributions. But one might ask a more ambitious question: Is there reason to believe

---

ing a consensus did the jury decide in favor of an initial minority. R. SIMON, *supra* note 32, at 117.

116. Justice White responded to the suggestion that the twelve-member jury gives the defendant a greater advantage because he is more likely to find one juror who will insist on acquittal and prevent conviction with the argument that it is equally likely that the advantage will accrue to the state, because there is a greater chance that it will find one conviction-prone juror who will prevent acquittal. One might try to justify the Court's attitude by arguing that it was a defendant who was appealing in *Williams* and not the public generally; hence, if the defendant's interests are not hurt, he has no cause to complain and the Court no cause to act. This position, however, would be sensible only if the sole purpose of the sixth amendment was to promote the interests of defendants as a class, without regard to the public interest. Even if this interpretation were sound, it should not have been relied upon in *Williams* because there the defendant had already been convicted; had his trial been before twelve instead of six the additional jurors could have only helped him.

that decisions of twelve-member juries are likely to be of a generally higher (or lower) quality, in terms of the dynamics of decision-making, than decisions rendered by juries of six? Social-psychological literature on the relationship between group size and the quality of group decision-making bears on this question. Although the methodology of the group-size research limits its relevance for the jury problem,[117] the available studies do provide information on the ways in which size can affect the quality of group decisions, suggest tactics for experimentation in more realistic settings, and illustrate how laboratory research may bear on a practical legal problem.

Thomas and Fink summarize the findings of thirty-one studies in which group size was an important variable by stating: "[I]t appears that both quality of group performance and group productivity were positively correlated with group size under some conditions, and under no conditions were smaller groups superior."[118] This conclusion holds generally, but the literature does not appear to justify Thomas and Fink's confidence that this will *always* be the case. Most apparent reasons for the superiority of large groups over small groups, and of all groups over individuals, derive from the participation of a greater number of individuals with more diverse viewpoints in the problem-solving effort. Thus, larger groups do better on problems where the quality of the group solution can be

---

117. *See* text at note 70 *supra.*

118. Thomas & Fink, *Effects of Group Size,* 60 PSYCH. BULL. 371, 373 (1963). Studies supporting this conclusion either directly, by indicating that larger groups are better than smaller groups in terms of the quality of their decision-making, or inferentially, by indicating that groups have advantages over individuals in ways that suggest that large groups have advantages over smaller groups, include: Faust, *Group Versus Individual Problem-Solving,* 59 J. AB. & SOC. PSYCH. 68 (1959); Fox, Lorge, Weltz & Herrold, *Comparison of Decisions Written by Large and Small Groups,* 8 AM. PSYCHOLOGIST 351 (1953); Gibb, *The Effects of Group Size and of Threat Reduction upon Creativity in a Problem Solving Situation,* 6 AM. PSYCHOLOGIST 324 (1951); Goldman, Dietz & McGlynn, *Comparison of Individual and Group Performance Related to Heterogeneous-Wrong Responses, Size, and Patterns of Interaction,* 23 PSYCH. REP. 459 (1968); Hall, Mouton & Blake, *Group Problem Solving Effectiveness Under Conditions of Pooling vs. Interaction,* 59 J. SOC. PSYCH. 147 (1963); Hoffman & Maier, *Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups,* in CURRENT STUDIES IN SOCIAL PSYCHOLOGY 458 (I. Steiner & M. Fishbein eds. 1965) [hereinafter CURRENT STUDIES]; Porter, *Information Distribution and Group Size: Some Effects on Group Problem Solving,* 4 INDUS. MANAG. REV. 1 (1963); Steiner, *Models for Inferring Relationships Between Group Size and Potential Group Productivity,* 11 BEHAV. SCI. 273 (1966); Taylor & Faust, *Twenty Questions: Efficiency in Problem Solving as a Function of Size of Group,* in SMALL GROUPS 513 (A. Hare, E. Borgatta & R. Bales eds. 1965); Tuckman & Lorge, *Individual Ability as a Determinant of Group Superiority,* in CURRENT STUDIES, *supra,* at 409; Watson, *Do Groups Think More Efficiently than Individuals?,* 23 J. AB. & SOC. PSYCH. 328 (1928). *But see* Hackman & Vidmar, *Effects of Size and Task Type on Group Performance and Member Reactions,* 33 SOCIOMETRY 37 (1970); Hare, *A Study of Interaction and Consensus in Different Sized Groups,*

measured by the simple sum of the contribution of all members.[110] Larger groups also surpass smaller groups in solving "eureka" type problems: problems in which insightful solutions seen by one group member may be easily explained to others.[120] Obviously the likelihood that one member will perceive the solution increases with group size.

Problems faced by jurors are seldom of the "eureka" type. A kind of "summing" does occur, however, when individual judgments are added and averaged, as is probably the case with many tort damage awards.[121] Small-group research suggests that in many problems involving quantitative estimation, such as judging the number of beans in a jar, the average judgments of statisticized groups[122] tend better to approximate correct answers than average individual judgments.[123] The group's advantage appears especially great when the problem involves unfamiliar material or when opinions as to the correct solution vary widely. Group superiority seems to depend entirely on the fact that averaging larger numbers of estimates reduces error variance; hence larger groups would have an advantage over smaller groups in dealing with such problems.[124]

Larger groups also do better than smaller groups where crucial inputs are needed, because different members may make critical contributions at different stages.[125] Thus, if memory is important, a large group is more likely to contain members who recall crucial

---

17 AM. SOC. REV. 261 (1952); Holloman & Hendrick, *Problem Solving in Different Sized Groups,* 24 PERSONNEL PSYCH. 489 (1971).

119. An example is the task of constructing as many words as possible from a set of letters in a given period of time. *See* Watson, *supra* note 118. But larger groups are typically less efficient in such tasks than smaller groups in that the number of words constructed per person per unit of time is less. *Cf.* Hoppe, *Memorizing by Individuals and Groups: A Test of the Pooling-of-Ability Model,* 65 J. AB. & SOC. PSYCH. 64 (1962).

120. *See* Hoffman, *Group Problem Solving,* 2 ADVANCES IN EXPERIMENTAL SOC. PSYCH. 99, 111 (1965); Lorge, Fox, Davitz & Brenner, *A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920-1957,* 55 PSYCH. BULL. 337 (1958); Kelley & Thibaut, *Group Problem Solving,* in 4 HANDBOOK OF SOCIAL PSYCHOLOGY 64-65 (2d ed. G. Lindzey & E. Aronson 1969).

121. *See* text at note 106 *supra.*

122. A "statisticized" group is an artificially constructed group of individuals who have never actually deliberated together, and whose judgments are averaged by the researcher.

123. *See* Lorge, Fox, Davitz & Brenner, *supra* note 120, at 344-46.

124. If, for example, relatively objective evidence was presented on the extent to which the amputation of an arm would reduce the future earnings of a truckdriver over a 20-year period, one would expect that the judgments of groups of individuals, on the average, would be closer to the amount actually lost than the judgments of individuals deciding alone. Judgments of larger groups, on the average, should be closer than judgments of smaller groups.

125. *See* Faust, *supra* note 118.

facts at each stage of the problem-solving activity. Thus, larger juries are likely to be superior to smaller juries where memory or a good understanding of facts and instructions is crucial to the deliberative process.[126]

A final strength of the larger group, although really another variation of themes already discussed, lies in its greater heterogeneity. Research indicates that heterogeneous groups are more likely to arrive at correct solutions to problems than homogeneous groups.[127] It has already been demonstrated that increasing the size of a randomly selected group tends to increase the heterogeneity of its membership.[128]

A study by Barnlund is suggestive, although it deals with group-individual differences rather than large-group/small-group differences.[129] Barnlund created student groups that were similar in syllogistic reasoning ability but not in other respects. The subjects were then presented with problems in syllogistic reasoning. Although susceptible to standard techniques of logical resolution, the problems were written so that both premises and conclusions contained references to conservative Republicans, communists, atheists, college professors, and other groups likely to arouse the prejudices of college students. Barnlund found that group solutions were substantially better than those of individuals. Indeed, groups whose members were in the lowest fifth of the class on ability rivaled the performance of the most skilled class members working alone. Group superiority was based on both additive and discussion effects. Pre-deliberation decisions constructed for each group on the basis of a majority of individual responses were superior to the average of individual products, and group decisions after discussion were significantly superior to these constructed solutions. Barnlund attributed much of the group superiority to the fact that the different prejudices that the problems elicited in different members were counterbalanced. Hence, individuals deliberating in groups were forced to become more objective. He commented: "The significance of this one factor alone would be hard to overestimate."[130] The existence of such

---

126. *See* Kelley & Thibaut, *supra* note 120, at 69.

127. *See* Goldman, Dietz & McGlynn, *supra* note 118; Hoffman, *supra* note 120; Hoffman, *Homogeneity of Member Personality and Its Effect on Group Problem-Solving*, 58 J. AB. & Soc. Psych. 27 (1959); Porter, *supra* note 118; Thomas & Fink, *supra* note 118, at 381. *But cf.* Shaw, *A Note Concerning Homogeneity of Membership and Group Problem Solving*, 60 J. AB. & Soc. Psych. 448 (1960).

128. *See* text at note 75 *supra*.

129. Barnlund, *A Comparative Study of Individual, Majority, and Group Judgment*, 58 J. AB. & Soc. Psych. 55 (1959).

130. *Id.* at 59.

counterbalancing prejudices would, of course, be more likely in a randomly selected group of twelve than in one of six. Barnlund also attributed group success to the fact that membership in the group appeared to increase motivation, to make people more self-critical and thus inhibit mistaken answers, and to provide greater critical resources than those available to individuals working alone.[131] Increased motivation, the first of these factors, is likely to be greater in groups of six than in groups of twelve; with respect to the other two factors, groups of twelve should be marginally superior.

Of course, groups cannot realize the advantages of heterogeneity unless members express different ideas before a final choice is adopted. Lorge and his associates found that the probability that an expressed idea would appear in a group solution was a function of its commonality; ten per cent of the ideas expressed by only one member appeared in the group solution, compared to fifty per cent of the ideas expressed by two or more members.[132] This suggests that the ideas of a minority are less likely to be ignored by a jury of twelve than by a jury of six, although it is possible that ideas must be expressed by proportionately more members in larger groups if they are to receive equal attention.

A leader, by "budgeting" discussion time for members who are alone in their position, can also ensure presentation of minority views. Maier and Solem, asking five- and six-person groups to solve Maier's well-known "horse-trading problem,"[133] found that except when there was an initially large correct majority, the presence of a discussion leader increased the tendency of group members ultimately to accept a correct answer.[134] When the minority was initially correct and there was no group leader, forty-six per cent of the initially incorrect subjects switched to the correct answer. When a discussion leader was appointed, seventy-two per cent of the initially incorrect subjects switched.[135] This suggests that effective leadership may be vital if heterogeneous groups are to realize the full potential of their diversity. Since larger groups are more likely

---

131. *Id.* at 58.

132. Lorge, Davitz, Fox & Herrold, *Evaluation of Instruction in Staff Action & Decision Making,* USAF HUM. RES. INST. TECH. REP. No. 16 (1953).

133. *A* buys a horse for $60. He sells it to *B* for $70. He buys it back from *B* for $80 and resells it for $90. How much if anything has *A* made on the deals? Students usually answer that *A* has made a profit of either $10 or $20. Lawyers, of course, will immediately perceive the correct answer.

134. Maier & Solem, *The Contribution of a Discussion Leader to the Quality of Group Thinking: The Effective Use of Minority Opinions,* 5 HUMAN RELATIONS 277 (1952).

135. *Id.* at 280-81.

to be diverse than smaller groups, skilled leadership may be particularly crucial if twelve-member juries are to perform up to their potential.[136]

Accepting the above evidence on the superiority of decision-making in larger groups, one question remains: Exactly how much better is such decision-making? Gibb, working with individuals and different-sized groups, found that during thirty-minute discussions "creativity," defined in terms of the number of ideas produced, increased as a negatively accelerated function:[137] Each additional member added something to the group's product, but he added somewhat less than his immediate predecessor. Since adding group members generally involves constant or increasing costs, at some point additional members will not contribute enough to the quality of the group decision to justify the additional cost. It is impossible, however, to tell from the write-up of Gibb's study the gains or costs involved in adding group members, and it is generally impossible to draw from the group-decision-making literature any concrete notion of how substantial an advantage a jury of twelve would have over one of six.

G. *Research That Favors Six*

The literature reviewed thus far does suggest that where the decisions of six- and twelve-member groups diverge, the decisions of twelve are likely to be more accurate, more creative, and less bound by prejudice. However, some small-group studies suggest that the larger group is not a better decision-maker, at least not under all conditions. Indeed, some researchers have found groups of five or six to be about the ideal size for group decision-making.[138]

---

136. *See* Hare, *supra* note 118, at 262. Also, leadership skill is probably more important in large groups because participation becomes increasingly stratified in such groups, so the leader comes to dominate more and more of the conversation. *See* A. HARE, HANDBOOK OF SMALL GROUP RESEARCH 231 (1962); Bales, Strodtbeck, Mills & Roseborough, *Channels of Communication in Small Groups*, 16 AM. SOC. REV. 461 (1951); Thomas & Fink, *supra* note 118, at 374. A formal leader may mitigate this effect by taking a relatively neutral role and assuring that all get a chance to participate. *Cf.* Hoffman, *supra* note 120, at 108. Leaders also may be important to break down the closeness of a small group and draw out minority opinion that otherwise would be inhibited. *Cf.* Bovard, *Group Structure and Perception*, 46 J. AB. & SOC. PSYCH. 398 (1951); Thomas & Fink, *supra* note 118, at 375. An effort to provide jurors with some training in techniques of discussion leadership might pay substantial dividends.

137. Gibb, *supra* note 118. Gibb worked with groups of sizes 2, 3, 6, 12, 24, 48, and 96. *See also* Lorge, Fox, Davitz & Brenner, *supra* note 120.

138. *See* Frye, Spruill & Stritch, *Effect of Group Size on Public and Private Coalescence, Efficiency and Change*, 62 J. SOC. PSYCH. 131 (1964); Hackman & Vidmar, *supra* note 118, at 48-49; Holloman & Hendrick, *supra* note 118, at 499-500; Rosenblatt & Rosenblatt, *Six-Member Juries in Criminal Cases. Legal and Psychological Considerations*, 47 ST. JOHN'S L. REV. 615, 633 (1973); Slater, *Contrasting Correlates of*

Like the studies favoring larger groups, these studies are hampered by the subjects used, the problems posed, and the range of sizes investigated.

Evidence in the literature indicates at least four areas in which smaller juries might have an advantage over larger ones: (1) Smaller juries are more likely to have members who are satisfied with their jury service and who accept the results of deliberation; (2) coordination problems that may interfere with effective decision-making will be greater in larger juries; (3) deliberation in smaller juries is more likely to reflect the relatively equal participation of all members; and (4) larger juries will tend to divide into factions that may adversely affect the quality of their decisions. My view is that on balance none of these propositions ultimately justifies the conclusion that juries of six are likely to be better decision-makers than juries of twelve, but they all deserve consideration.

The first of these advantages—the greater satisfaction expressed by members of smaller groups—is well documented,[139] but of questionable relevance where the concern is with the quality of decision-making rather than with the feelings of decision-makers. That members are satisfied with their group or with the quality of their decisions does not mean that the group decisions are better in any objective sense. For example, Slater, working with groups of sizes two through seven, has suggested that five is probably the optimal number for solving human relations problems. Members of smaller groups tend to feel that their group is too small for the task, members of larger groups tend to feel that their group is too large, and members of groups of five tend to feel that their groups' size is about right.[140] Slater explains these findings by suggesting that five-person groups are large enough that members feel free to make aggressive efforts toward problem-solving even at the risk of antagonizing others, yet small enough that members feel some regard for the

---

Group Size, in CURRENT STUDIES, *supra* note 118, at 416. The Holloman and Hendrick study, which found groups of six to be better—although not significantly so—than groups of twelve in predicting the outcome of the movie "Twelve Angry Men" must be discounted because groups were constructed to be homogeneous with respect to ability. This artificial construction vitiates one of the principal advantages of a larger group. In Goldman, Dietz & McGlynn, *supra* note 118, at 461, the authors report that the performance of heterogeneous groups improved with size, but not the performance of homogeneous groups.

139. *See, e.g.,* Hackman & Vidmar, *supra* note 118; Hare, *supra* note 118; Thomas & Fink, *supra* note 118, at 378. Of course, these feelings might be important with respect to other values that the legal system might wish to promote. It is predictable that if jury size were reduced from twelve to six the average satisfaction of individuals with their jury experience would go up.

140. Slater, *supra* note 138, at 418.

needs of others. Slater, however, does not attempt to judge the quality of group decisions. Hackman and Vidmar, working with groups of the same size as Slater's on three different problems, replicated Slater's finding.[141] They also independently analyzed the quality of the group product and found that group decisions tended to be better when made by groups both above and below size five. They suggest that middle-sized groups may be too comfortable for effective task performance, and that the stress engendered by the conflict and coordination problems of larger groups may actually induce better task performance. A related explanation that is supported by some research is that conformity increases as group ties become stronger.[142] Thus, as one feels more comfortable in a group or grows to value his group membership more, the likelihood that he will challenge the dominant group perspective decreases.

Coordination problems do, of course, increase with group size. Where resolution of a problem requires a consistent strategy or coordinated movements, as with certain motor problems or designing a crossword puzzle, group performance is often poorer than that of the best individual member.[143] Smaller groups probably would have an advantage over larger groups in such situations. Jury-type problems, however, are unlikely to place such a high premium on coordination. Fox and his associates, for example, report that when 400 air force officers were given thirty minutes to solve a complex human relations problem, groups of twelve and thirteen produced solutions superior to those reached by groups of six through eight.[144] These officers probably had an advantage over the ad hoc groups typically studied by social-psychologists and over juries in that pre-existing rank differences could ease coordination problems. However, juries have an offsetting advantage in that they need not render verdicts within a particular length of time, and the selection of a foreman aids in organization. Coordination difficulties might force juries of twelve to take longer to reach verdicts than juries of six, but there is no reason to expect that the solutions finally reached would be inferior.

The increased inequality of participation in larger groups is a third factor that implies that smaller juries might be superior to

---

141. Hackman & Vidmar, *supra* note 118, at 48-49.

142. *See* Bovard, *supra* note 136; Deutsch & Gerard, *supra* note 96; Dittes & Kelley, *Effects of Different Conditions of Acceptance upon Conformity to Group Norms*, 53 J. AB. & Soc. PSYCH. 100 (1956); Kidd, *supra* note 93. *Cf.* Bales, *The Equilibrium Problem in Small Groups*, in SMALL GROUPS, *supra* note 118, at 444.

143. *See* Kelley & Thibaut, *supra* note 120, at 69-70, 74-75.

144. Fox, Lorge, Weltz & Herrold, *supra* note 118.

692   *Michigan Law Review*   [Vol. 73:643

larger ones. The normative model of the jury process suggests that
each juror should contribute to the deliberations, and there is some
evidence in the small-group literature that highly skewed partici-
pation will adversely affect decision-making. Hare, for example,
reports that when boy scouts were asked to rank the relative use-
fulness of various pieces of camping equipment, the most influential
boy in groups of five was the one with the most camping experi-
ence; the most influential boy in groups of twelve was the one who
talked the most.[145] In a review of the literature, Hoffman cites addi-
tional evidence that the individual who talks the most will have the
greatest impact on group problem-solving.[146] He reports an experi-
ment by Riecken that indicates that groups rarely adopt "elegant"
solutions when their proponents are the least talkative members of
the groups, but almost always adopt them when their proponents
are the most talkative members.[147] Hoffman suggests that the most
self-confident members may prevent a group from using all of the
information available because they so dominate discussion that they
can focus attention on matters they think important.[148] Hoffman
also feels that the disproportionate influence of the most self-
confident group member will increase with the size of the group,
because potential dissenters will grow increasingly reluctant to
voice their opinions for fear of being thought deviant.[149] This last
observation, at least, is somewhat questionable; Thomas and Fink
among others suggest that smaller groups tend to inhibit disagree-
ment because of their greater solidarity.[150]

The findings of Hare and Hoffman, if applicable in the jury
setting, would imply that the decision-making advantages that the
twelve-member jury enjoys by virtue of the greater heterogeneity
of its membership would be more or less canceled during the
deliberation process. If members do not contribute to the group
decision, the fact that they have unique ideas and values means

---

145. Hare, *supra* note 118.

146. Hoffman, *supra* note 120, at 105-07.

147. *Id.* at 105, *citing* Reicken, *The Effect of Talkativeness on Ability To Influence
Group Solutions of Problems*, 21 SOCIOMETRY 309 (1958). There is also evidence that
high participators exercise disproportionate influence in a jury setting. Strodtbeck and
his associates report that individuals who participated more expressed greater satis-
faction with the deliberation and shifted their position less often in the process of
arriving at a verdict. Other jurors also disproportionately reported high participators
as being helpful in reaching a verdict. Strodtbeck, James & Hawkins, *supra* note 70,
at 3-11.

148. Hoffman, *supra* note 120, at 107.

149. *Id.*

150. Thomas & Fink, *supra* note 118, at 375. *See also* studies cited note 142 *supra.*

little. Although the arguments that favor this conclusion are reasonable, close evaluation of available evidence suggests that such cancellation effects do not occur to any great extent. Even though low participators in a group of six can be expected to participate more than low participators in a group of twelve,[151] the advantages of the increased heterogeneity of the larger group should not be substantially diminished.

With respect to the danger that a single member will dominate the decision-making, the careful research of Bales and his associates indicates that in groups of six working with a human relations discussion problem, the most active member will be responsible for about forty-three per cent of the interaction initiated and the next most active member for about nineteen per cent.[152] In twelve-member juries the evidence implies that the degree to which one member dominates conversation will be substantially less. Strodtbeck and his associates found that jury foremen, the most active jury participants, initiated only about a quarter of the interaction, and tended to take a neutral rather than a partisan position on a damage issue.[153] Bevan and his associates found that even high status jury foremen who had been instructed to dominate discussion had only limited success in raising the jury damage award to a target set by the experimenter.[154] Thus, increasing jury size from six to twelve does not necessarily cause the jury to become increasingly dominated by one or two persons; indeed, it seems likely that such an increase would reduce the absolute participation of the leading participator, although the extent to which the leader's participation proportionately exceeds that of other jurors might be unaffected or increase.

Low participators in a larger group, on the average, do participate less by percentage than low participators in a smaller group, but this difference is in part artifactual. Assuming equal participation by each member, the percentage participation of members of a larger group will always be less than the percentage participation of members of smaller groups, simply because there are more members to share the time.[155] Where there is no fixed time limit on discussion, as in the jury setting, low percentage participation does not

---

151. *See* A. HARE, *supra* note 136, at 232.

152. *See* Bales, *supra* note 142, at 458.

153. Strodtbeck, James & Hawkins, *supra* note 70, at 716.

154. Bevan, Albert, Loiseaux, Mayfield & Wright, *Jury Behavior as a Function of the Prestige of the Foreman and the Nature of His Leadership*, 7 J. PUB. L. 419 (1958).

155. Diamond, *supra* note 12, at 526-27, makes this same point in her reanalysis of the data presented in Note, 6 U. MICH. J. L. REF. 712, *supra* note 4.

necessarily mean that an individual has been unable to make all of the points that he thinks important. Indeed, even the comment, "I don't understand" may be helpful if it forces other members to clarify their arguments.[156]

A second reason why average participation rates of low participators might be less on twelve-member juries than on juries of six is that larger jury size increases the probability that the argument or insight one was going to offer will be offered first by another. Thus, an individual who finds high participation personally uncongenial may be able to withdraw substantially from a twelve-member discussion without depriving the group of ideas or information.[157]

Of course, it is also possible that those who withdraw from discussion on a twelve-member panel do have original arguments and ideas but are intimidated from presenting them by the size of the group they face. Hawkins, however, found that when a jury was divided into factions, the minority faction initiated substantially more than its proportionate share of the conversation.[158] Majority factions apparently place considerable pressure on minorities to defend their positions. In this setting, Hawkins found that minority jurors generally increased their participation, and the increase was likely to be quite substantial if the individual ultimately found himself a minority of one.[159] This suggests that, even on a larger jury, recalcitrant members will participate when no other member makes their arguments.[160]

A final factor that mitigates any advantage smaller juries might

---

156. Barnlund states that such professions of limited understanding helped his groups in their task of syllogistic reasoning. Barnlund, *supra* note 129, at 58.

157. Social status is noticeably associated with participation in the twelve-member jury. Hawkins found that, after excluding the foreman, college-educated individuals were responsible for about 9.8 per cent of the total participation, while those with only grammar school educations were responsible on the average for about 5 per cent of the participation. Controlling for occupation as well, the range was from college-educated proprietors, who averaged 13.5 per cent, to grammar-school laborers with an average participation level of 3.7 per cent. C. Hawkins, *supra* note 32, at 37, Table 9. Thus, those on the jury more used to resolving problems verbally may do most of the talking. If high-status groups are more capable than members from low-status backgrounds in dealing with jury-type problems, these status effects may improve the final product of the deliberations.

158. C. Hawkins, *supra* note 32, at 123-37. For evidence of the same phenomenon, see Diamond, *supra* note 12, at 525.

159. Of course, in this setting members in the majority direct their arguments to the one dissenter and call on him to respond. Individuals in factions of two or more may refrain from expressing persuasive arguments, allowing their allies to express less persuasive ones. *See* C. Hawkins, *supra* note 32, at 136-37.

160. Where jurors are alone in the minority, the increased pressure to participate makes holding out quite difficult. The lone dissenters who succeeded in hanging Hawkins' juries apparently did so by withdrawing or making a mockery of the deliberations. *Id.* at 135-37.

have in encouraging more widespread participation is that on an absolute basis, larger size means that more viewpoints are likely to be represented, even if more views are also suppressed. In her law review note, Kessler reports that participation was unanimous in six of eight experimental six-member juries, compared to only one of eight twelve-member juries.[161] However, no twelve-member jury had fewer than seven participants, and six of the eight had nine or more participants.[162] Moreover, these results unquestionably underrepresent the breadth of participation on twelve-member panels. Either Kessler's student jurors were generally unmotivated or her experimental case was too easy or both, for only five of her panels, including just one twelve-member jury, took longer than twenty minutes to reach a decision.[163] Simon, by contrast, reports that cases in which all twelve jurors did not contribute at least one comment were unusual.[164] Hawkins' research suggests that the organization of the deliberation process may promote full participation. Jurors frequently seek to determine where their jury stands by the use of open polls, a tactic that not only forces jurors to reveal their views but may give each juror a structured opportunity to explain his position.[165]

The final and most intriguing argument for the superiority of decision-making by smaller juries is based on Hawkins' doctoral dissertation, part of the Chicago Jury Project. Hawkins monitored forty-six twelve-member juries deliberating mock tort actions. Jurors were citizens who had actually been called to duty in the federal district court. Hawkins was able to distinguish between two quite different modes of proceeding, one of which he called "deliberating in unity" and the other "deliberating in factions."[166] When jurors deliberate in unity, they seek to resolve issues without the use of advocates. Jury members may present arguments for either party, and the decision, which emerges from the bulk of facts and evaluations, cannot be traced to any particular person or set of persons.[167] When jurors deliberate in factions, collective opinion, instead of

---

161. Note, 6 U. MICH. J. L. REF. 712, *supra* note 4, at 731.

162. *Id.*

163. *Id.* at 725.

164. R. SIMON, *supra* note 32, at 244 n.9. Simon was working with real jurors on a mock case. She studied 98 juries, but her comment may be based only on 39 cases in which deliberations were transcribed. In any case, her experiment was more sophisticated in its design than Kessler's, presented more difficult issues to the jury, and had the advantage of real jurors. All these factors should make her results more reliable.

165. *See* C. Hawkins, *supra* note 32, at 74-82.

166. *Id.* at 106-10.

167. *Id.* at 107.

being a product of discussion, is a product of preexisting personal opinions. Members argue in favor of preferred outcomes and unite with other members who share their goals, without regard to the others' reasons. Discussion is organized into an exchange of speeches by advocates. Those in a faction are more concerned with the effectiveness with which the other side is being attacked than with the arguments put forth to attack them. While deliberation in unity can involve all jurors, deliberation in factions involves what are really just two "individuals" or subgroups.[168]

Hawkins recognizes that these descriptions are "ideal types" and that actual jury deliberations will often have some of the characteristics of one and some of the other. However, the description of the ideal types indicates Hawkins' apparent opinion that juries deliberating in unity are likely to produce "better" results than juries deliberating in factions. At one point he states: "A group which tries to reach a decision without dividing into factions does not pay any attention to how the people are lined up but only to how the weight of rational argument is lined up . . . ."[169] At another point he concludes that when deliberations are in unity, "we would expect that the content and logic of the individual arguments will have primary importance."[170]

If Hawkins is correct, smaller juries should have substantial qualitative advantages over larger juries, because considerable evidence suggests that larger groups are more likely than smaller groups to become factious.[171] But there is little evidence to support Hawkins' interesting conjectures. Hawkins himself consciously ignores the characteristics of deliberation in unity in the most detailed portion of his analysis, and he acknowledges that the formal characteristics of group decision-making that he does examine would probably yield little understanding of groups that hew close to the ideal of deliberation in unity.[172]

The major evidence supporting Hawkins' conclusion that deliberation in unity is the more rational process is his finding that the length of discussion when groups are unaligned tends to be unrelated to the number of individuals in the minority,[173] but that once

---

168. *Id.* at 108.

169. *Id.* at 115.

170. *Id.* at 155. Presumably this statement does not hold for deliberation in factions, where such formal characteristics as the relative size of the majority and minority are seen to be of primary importance.

171. *See, e.g.,* Hare, *supra* note 118; Holloman & Hendrick, *supra* note 118; Thomas & Fink, *supra* note 118.

172. C. Hawkins, *supra* note 32, at 154-55.

173. *Id.* at 113-15.

jurors have voted and revealed the existence of coalitions, a rela-
tionship emerges between the length of subsequent arguments and
the number in the minority.[174] Hawkins apparently interprets the
fact that this relationship exists only when coalitions are acknowl-
edged as evidence that the jurors are attending to alignments and
irrelevant attempts at persuasion rather than to the rational analysis
of fact.[175] Implicit in this interpretation is the notion that there is
only a constant number of rational arguments to be made and
facts to be analyzed; hence one would expect discussion confined to
the rational and factual to take the same amount of time regardless
of the way in which opinion is divided.[176] Stating the position in
this way reveals its weaknesses: Even if there are only so many
rational arguments to be made or facts to be analyzed, it doesn't
follow that all arguments will always be made or that some may not
be made more than once.

Furthermore, Hawkins' findings are just as consistent with the
hypothesis that deliberation in unity inhibits the expression of
minority opinions. It has already been suggested that the closeness
of small groups, a closeness that may lead to more unified delibera-
tions (or vice versa), might tend to inhibit dissent.[177] Indeed, the
results of one study appear contrary to Hawkins' conjectures. In an
examination of the contents of interaction in various groups,
Hamblin and Miller found that evaluative statements were more
common in smaller groups. Their conclusion is that the presence
of factions, more likely in larger groups, may cause members to com-
municate more factual material and less opinion.[178] Perhaps
Hawkins' results may also be explained by the fact that the periods
of deliberation in unity that he examined tended to come dispropor-
tionately at the start of discussions, where the minority's role might
differ from what it is when a verdict is close at hand.

The only other evidence in Hawkins' study bearing on the
possible superiority of deliberation in unity is anecdotal observa-
tions that jurors aligned in factions may direct harassing tactics or
unreasonably extreme arguments to the opposite faction.[179] The
presence of irrational arguments, however, does not mean that

---

174. *Id.*

175. *Id.* at 115.

176. Hawkins is not very clear on his analysis here. This is my interpretation of
his chain of reasoning and may be unfair.

177. *See* text at note 142 *supra. But see* Hoffman, *supra* note 120, at 107.

178. Hamblin & Miller, *Variations in Interaction Profiles and Group Size*, 2 Soc. Q.
105 (1961).

179. *E.g.,* C. Hawkins, *supra* note 32, at 131-32.

rational arguments have been suppressed, and may indicate that factional deliberation tends to ensure that all available arguments, of whatever type, are put before the group.

In short, while Hawkins' thesis is intriguing, there is virtually no hard evidence that suggests that deliberation in unity or the deliberations of smaller juries should yield more rational results than deliberation in factions or the deliberations of larger juries.[180]

H. *On Balance*

The arguments in this section, both those that are statistically derived and those that are based more broadly on the social-psychological literature, are crude. The statistical analysis requires too many assumptions that are obviously false in at least some instances, and the social-psychological analysis is both based on imperfect research and, for the most part, too far removed from the jury situation to allow confident generalization. Nevertheless, the evidence is generally consistent and overwhelming. *Current knowledge justifies the general conclusion that where the verdicts of six- and twelve-member juries diverge, the verdicts of twelve are likely to be of somewhat higher quality than the verdicts of six, and are likely to be superior with respect to other important values.*

Had the proponents of the six-member jury in *Williams* and *Colgrove* been assigned the burden of showing that the change in size was unlikely to decrease the average quality of jury justice, they would not have prevailed given our present state of knowledge. Indeed, a fair evaluation of the relevant evidence suggests the opposite conclusion. Available evidence, however, is insufficient to permit much more than a general statement about the likely *direction* of size effects. In particular, the evidence is insufficient to assess the magnitude of size effects. Yet this may be the crux of the problem. The increased cost of larger juries means that the magnitude of size

---

180. *But see* Rosenblatt & Rosenblatt, *supra* note 138. The Rosenblatts are apparently persuaded by Hawkins and their reading of the small-group literature, for they conclude, largely on the basis of Hawkins' hypothesis, that "the six member jury affords a superior and more permanent consensus of opinion." Rosenblatt & Rosenblatt, *supra*, at 633. If they mean that the members of smaller juries will be more satisfied with their decisions after reaching them and less likely to change them in some post-deliberation setting, I agree with them. If they mean that two six-member juries hearing the same case are more likely to agree than two twelve-member juries hearing that case, or that the verdict of the six-member jury has resulted from a more rational consideration of the evidence—or is better by some other standard—then I must disagree. It may be easy to maintain an appearance of complete rationality where minority opinions are suppressed or not present. But so long as the minority position is not itself irrational, a good decision should give a minority the weight its arguments deserve.

effects is likely to be crucial in determining the desirability of smaller juries.

I have suggested that in only a fraction of all jury trials would the verdicts of twelve differ from those of six.[181] Review of the literature indicates that Zeisel and Diamond's ideal experiment probably would reveal that where verdicts do differ, those of twelve would, on the average, be superior. But in many cases it would be impossible to say that the verdict of one jury is better, and in some cases the verdict of the smaller jury would appear preferable. It is impossible to weigh dollar and time savings of the smaller jury against these speculative outcomes. The costs of six-member verdicts might be so low as to be clearly outweighed by the savings they engender, or they may be so substantial as to warrant immediate reversal of the decisions in *Williams* and *Colgrove*. If, however, the sixth and seventh amendments do not allow one to weigh dollars against any benefits—however small—of larger juries, then it appears that *Williams* and *Colgrove* should be reversed irrespective of the magnitude of the size effect. My best guess is that these decisions should be reversed on either test: The savings of the smaller jury are quite likely outweighed by the costs, some of which we may never measure.

Still, based on current knowledge it is impossible to be scientifically certain that future research will not disclose substantial advantages to juries of six. For reasons such as those suggested by Hawkins and others, one may find that juries of six do better with respect to certain values than juries of twelve. In addition, research capable of revealing discernible differences may show none. My impression is that the final judgment on six versus twelve will turn on the values that individuals subjectively place on the presence of minority views in the jury room. This judgment may and perhaps should depend on the roles that minorities play in jury deliberations and the ways in which they play them, something we currently know little about.

III. Research Design and Other Matters

A. *Measuring Verdict Differences*

Thus, I join Zeisel and Diamond in their call for further research, but I favor different strategies than the carefully controlled "real world" experimental approach they emphasize. If I am correct in my argument in Part I, most cases will so clearly favor one party

---

181. *See* section IB *supra.*

that studies of actual trials are unlikely to reveal substantial dif-
ferences in the percentage of plaintiff or prosecution verdicts ren-
dered by different-sized juries. One statistic, however, can be
expected to respond sufficiently to jury-size differences to provide
evidence as to whether important size effects do or do not exist: the
percentage of hung juries in jurisdictions that mandate different-
sized juries.

The preceding section indicates why one would expect twelve-
member juries to hang more frequently than six-member juries. The
validity of this expectation could be tested by comparing hung-jury
rates in states that try crimes to twelve with those in states that try
similar crimes to panels of six. The experimental approach advo-
cated by Zeisel and Diamond would not be desirable because the
low incidence of hung juries would require an exceedingly large
number of experimental trials before one could expect differences
in the hung-jury rate to achieve statistical significance. Moreover,
experimental controls are unnecessary because the array of criminal
cases tried in a state that mandates six-member juries is likely to
be so large and varied that these trials as a group are unlikely
to differ greatly from the array of criminal cases tried to twelve-
member juries in a neighboring state; hence the expected rate
of hung juries in the two states, absent size effects, would be the
same. This conclusion entails an assumption that all other things
are equal that could be checked both by examining the kinds of
cases tried to juries in the two states and by comparing hung-jury
rates in the six-member jury state with these rates in a number of
neighboring states trying criminal cases to twelve. If additional states
reduce jury size in all cases, the design may be further improved by
comparing hung-jury rates in these states with the rates in twelve-
member states and with their own rates before the switch to six.[182]
If six-member jury states have lower hung-jury rates than twelve-
member neighbors and lower rates than they themselves had before
the change in size, one could reasonably conclude that the difference
was due to size effects and not to some confounding factor.[183]

---

182. On the design of such an experiment generally, see Lempert, *supra* note 12,
at 130-32. One could also use other means to increase the certainty that any variation
in hung-jury rates was due to jury size and not to some other factor. For example, if
one worried that attorneys make different decisions on whether to demand jury trial
when juries are composed of six rather than twelve, one might interview attorneys
in different jurisdictions to learn what factors they weigh when considering the pos-
sibility of a jury trial. In addition, one could check the rate at which attorneys de-
mand jury trials before both sizes of juries to evaluate the extent of any differences.

183. I have already cited data collected by Zeisel that indicates that the hung-jury
rate in a nationwide sample is more than twice as great as the hung-jury rate in one

In theory, a second statistic calculable on the basis of trial results that would reflect differences attributable to jury size is the variance of the amounts awarded in damage actions.[184] One would expect the variance to be larger in jurisdictions where damage actions are tried to six than in jurisdictions where they are tried to twelve. Again, experimental controls might not be essential. Those damage actions brought to trial in a six-member jury state could be expected as a group to be generally similar to damage actions tried to juries in neighboring twelve-member states. Here, however, more caution would have to be exercised in making comparisons because there are a larger number of plausible explanations for any differences that might be discovered. Data from states that allow attorneys to elect among different-sized juries would be unacceptable, and data from states that provide different-sized juries for different kinds of cases would present peculiar problems.[185] Other confounding factors include potential differences in the settlement policies of insurance companies, differences in the quality of the bar of each state, and specific state laws or other conditions (*e.g.*, allowance of punitive damages) that might affect the size of awards.

These potentially confounding factors should not imply that study of the association between jury size and damages awards might not be useful. One fortunate feature of the expected association between jury size and awards that provides an important check on many of the factors that could confound a comparison of variance statistics is the statistical expectation that the average of awards rendered by different-sized juries in similar types of cases would be the same even if the variance in awards is greater for the smaller jury. A finding that the average amounts awarded in two states were not greatly different would suggest that a comparison of the variance in awards is an appropriate indicator of size effects. Visual comparison of the range of cases that make up the averages would be a further check.

Florida county, *see* note 99 *supra* and accompanying text, but clearly more controlled comparisons of a similar nature are needed before one can be certain that these results are attributable to jury-size differences.

184. This statistic may be based on all damage awards or on a subset of damage awards, such as automobile personal injury cases.

185. Such states, however, would be particularly suitable for bi-state comparisons if cases in the comparison state could be sufficiently broken down by type. In comparing a state in which civil juries were always composed of six with another state in which juries of six were used only in automobile negligence cases one would expect no difference in the variance of awards rendered in automobile negligence cases, but a greater variance in the awards rendered in actions that were tried to twelve in the second state.

An alternative, and perhaps superior, strategy for exploring the association between jury size and variance in damage awards is experimentation with mock trials. Experimentation allows complete control over the facts and rules of law involved. Because all juries, except those that hang, will reach a decision on damages, a moderate number of experimental trials could produce either statistically significant differences in the variance of awards[186] or reasonable certainty that such differences in variance do not exist to any important extent. In addition, working with mock juries allows the experimenter, by monitoring deliberations, to test the validity of the theory that it is averaging processes that cause juries of twelve to show less variation than juries of six.[187]

Mock juries also provide the preferred strategy for examining other probable differences between juries of twelve and six. This "laboratory strategy" allows selection of cases that are sufficiently similar to reveal systematic but unanticipated size effects. In addition, members of visible minorities, such as blacks, could be placed on juries to determine the effects of their presence. Even if the presence of visible minorities had the same influence on juries of twelve as on those of six, the bare finding of influence would indicate likely size effects because the presence of one or more minority group members is more likely on larger juries.[188]

The most important advantage of mock-jury experimentation is that it allows the researcher to monitor the deliberation process. In a close case it may be impossible to know or to prove that one outcome is better than another, yet it may be possible to conclude that a particular decision-making process is preferable. Thus, if research reveals that six-member juries tend to reach their verdicts by deliberations in which all jurors make rational contributions while twelve-member juries are dominated by one or two members or draw lots in frustration, one might well conclude that the quality of justice rendered by six is likely to be superior to that rendered by twelve, even though nothing in observed verdict differences makes this conclusion obvious.

If mock juries are used, the researcher must be careful to avoid certain problems. The most obvious pitfall inheres in the choice of jurors. While college sophomores may be good subjects in experiments designed to explore basic social-psychological processes, a

---

186. This would depend, of course, on the extent to which individual awards varied both within and between groups.

187. *See* text at notes 121-24 *supra.*

188. *See* Table One *supra.*

group composed of college students generally will be unsuited for
jury-size research. There is simply too much evidence that juror
interaction is strongly influenced by factors such as age and occupa-
tional status.[189] The homogeneity of college students on these and
other dimensions not only indicates that their deliberations are
unlikely to represent actual jury deliberations, but also diminishes
any advantage that juries of twelve enjoy over juries of six. The
model for the selection of mock jurors should be the "laboratory"
studies generated by the Chicago Jury Project.[190] Project researchers
received permission from federal district courts in three cities to
use as mock jurors individuals who were part of the court's venire
but who were not otherwise empaneled on the day of the experi-
ment. Absent such judicial cooperation, experimenters should try
to get jurors from the larger, nonstudent community.[191]

Although access to real jurors is ideal, jury-size research probably
should vary from the Chicago model in certain respects. Jurors in the
Chicago project were neither examined on voir dire nor subjected to
challenge. Yet voir dire questioning with a right to challenge adds an
important element of realism to jury-size experimentation. Effective
use of voir dire might minimize some of the differences between
juries of six and twelve.[192] The Chicago researchers' use of pre-
deliberation votes also should be avoided. While such information
is important to measure the effects of the deliberation process on
individuals' attitudes, the work of Gerard and his associates indi-
cates that prior written commitment will affect the extent to which
individuals are influenced by groups.[193] Since the goal of the pro-
posed research is investigation of influence processes within the jury
as a whole, researchers should not risk the potential interference
with the processes of attitude change that prior commitment might
engender.[194]

---

189. *See* R. SIMON, *supra* note 32, at 98-119; C. Hawkins, *supra* note 32, at 22-25,
30-41; Strodtbeck, James & Hawkins, *supra* note 70.

190. *See, e.g.*, R. SIMON, *supra* note 32, at 98-119; C. Hawkins, *supra* note 32, at
22-25, 30-41; Strodtbeck, James & Hawkins, *supra* note 70.

191. *Cf.* Bevan, Albert, Loiseaux, Mayfield & Wright, *supra* note 154.

192. Indeed, it would be very interesting to work with juries that have been sub-
ject to voir dire challenges and those that have not. Comparisons could give some in-
dication of the extent to which elimination of certain jurors and the indoctrination
that is often possible on voir dire affect potential jury-size differences. One danger to
be avoided or experimentally tested is that voir dire will be taken more seriously or
be more thoughtfully conducted in the experimental setting than it would be by the
typical practicing attorney in an ordinary jury trial.

193. Gerard, *Deviation, Conformity, and Commitment,* in CURRENT STUDIES, *supra*
note 118, at 263; Gerard, Wilhelmy & Conolley, *supra* note 93.

194. Refraining from obtaining pre-deliberation votes would prevent researchers

The selection and presentation of experimental cases also poses important problems. The facts of the cases used for a study of size effects must support at least two possible verdicts; otherwise there is little reason to expect that different-sized juries will decide differently. Civil cases have a particular advantage in that they generate information on two very different decision-making processes: the decision on liability and the decision as to damages.[195]

When cases are close, differences in jury verdicts may stem from random rather than systematic factors. One would then expect verdicts of juries of the same size to differ from each other as often as they differ from those of juries of a larger or smaller size. If, however, verdict differences are systematically associated with jury size, it will be necessary to explain why these differences exist. Here monitoring and analysis of jury deliberations is essential. One might note the potential problem that the monitoring process could make the jury self-conscious and affect its deliberations, but the available evidence suggests that the problem exists more in theory than in practice.[196] Jury monitoring also allows one to investigate the influence of specific factors, such as the race of criminal defendants, on the deliberations of different-sized and differently composed juries. Cases constructed to include such factors will probably show higher rates of verdict disagreement than other close cases. This finding would be important, because size differences in such cases may have particular relevance to the realization of other values of the legal system.

An additional advantage of using mock rather than real juries is that presentation of cases on audio or videotape ensures that the members of each experimental jury see essentially the same case. However, the use of such devices decreases the realism of the mock trial. This, plus the fact that mock jurors know that they are not actually determining an individual's fate, are probably the two most important weaknesses of this experimental approach. Both suggest that experimental juries will be less motivated to reach subjectively correct decisions and less willing to extend their deliberations than they would be if the cases were real. Nevertheless, the experience of the Chicago researchers working with recorded trials is heartening

from following the suggestion of Zeisel & Diamond, *supra* note 10, at 292, that initial vote distributions be equalized by assigning members to juries based on initial ballots. As Zeisel and Diamond recognize, the problem they are concerned with diminishes as sample size grows. The problem also will be less when close cases are used, since almost all juries can be expected to have some dissenters.

195. *See generally* C. Hawkins, *supra* note 32.

196. *See, e.g.,* R. SIMON, *supra* note 32, at 38.

in this regard. Simon reports that most of her jurors took their roles quite seriously, and that the average experimental jury's deliberations lasted as long as the average deliberation for a trial that takes two or three days.[197] Observation of the length and tone of the jury deliberations, and post-verdict interviews, can provide good checks on the extent to which the necessary realism has been achieved.

The mock juries in the Chicago project also hung more frequently than do trial juries generally.[198] This may be entirely due to the difficulty of the Chicago project cases,[199] but it may also be due to the fact that experimental juries do not feel as much compulsion as real juries to reach a conclusion. Perhaps a social-scientific version of the "Allen Charge" is called for.[200]

B. *So What?*

Suppose experiments such as those suggested above are conducted, and the results indicate that verdicts of twelve are superior to those of six. One might still ask, "So what—haven't *Williams* and *Colgrove* decided the matter?" While there is no reason to be confident that the Supreme Court would do an about-face on the jury-size question if presented with these hypothesized results, the question of how the Court should react to such research is an interesting one.

In his classic commentary on the empirical evidence presented to the Supreme Court by the plaintiff in *Brown v. Board of Education*,[201] Professor Cahn concluded that the Court was misguided even in acknowledging the material with so much as a footnote.[202] In justifying this position, he commented: "I would not have the constitutional rights of Negroes—or of other Americans—rest on any such flimsy foundation as some of the scientific demonstrations in these records."[203] Cahn was, of course, correct. Not only was the Court's scientific foundation flimsy, as Cahn proceeded to demon-

---

197. *Id.*

198. *See* C. Hawkins, *supra* note 32, at 43-44.

199. *Id.* at 44.

200. *See* Allen v. United States, 164 U.S. 492 (1896). *Cf.* Note, *supra* note 90, at 100-06. My purpose here is not to design a jury-size experiment, but to sketch some of the factors that should be considered before such experiments proceed. Hopefully the discussion also suggests more generally the potential power of laboratory experiments to elucidate questions of practical interest to lawyers.

201. 347 U.S. 483 (1954).

202. Cahn, *Jurisprudence*, 30 N.Y.U. L. REV. 150 (1955).

203. *Id.* at 157-58.

strate,[204] but the values involved made even convincing proof of the evils of segregation irrelevant. Blacks are degraded by segregation regardless of the extent to which they are hurt or, indeed, helped by such a system, and the Constitution in 1954 could have been, should have been, and was interpreted to prohibit states from legally separating the races.[205]

But the jury-size cases are very different from the segregation cases, and the influence of the results of empirical research should be quite different. The Court was not deciding any great value question when it decided in *Williams* and *Colgrove* that neither the sixth nor the seventh amendment required juries of twelve. Rather, relying on its premise that a reduction in jury size would have little effect on jury verdicts, the Court concluded that a reduction in jury size was not constitutionally prohibited because *no* value *was* substantially affected. If changes in jury size do not affect the values that inspired the jury trial guarantees,[206] it becomes almost irrelevant that the framers equated the concept of a "jury" with a body of twelve.[207]

Given the Court's functional response to the apparent intent of the framers, if a substantial body of well-conducted research should demonstrate that jury size does affect verdicts in ways that are detrimental to important values, the Court should reconsider its conclusions in *Williams* and *Colgrove*. There is little left of these opinions once the premises on which they rest are proved false.[208]

---

204. *Id.* at 159-65.

205. This is not to say that the state of the scientific study of race relations should have had no effect on the *Brown* decision. Had there been a body of reputable scientific literature that demonstrated that all races benefited substantially from segregation it is quite possible that segregation would not have been seen as inherently degrading for one race; had science "demonstrated" that black people were necessarily inferior to white people, the degradation entailed by segregation might not have been perceived as unfair. A somewhat different view of what was "known" about black people may have led to the decision supporting segregation in Plessy v. Ferguson, 163 U.S. 537 (1896). The implication of Kalven's question, "[F]or social science learning to have an impact on the living law, will it first have to become *popular* learning and enter law via the normal political process?" is an important insight into the relationship between empirical science and fundamental values, at least if the judicial system is considered to be part of the normal political process. Kalven, *The Quest for the Middle Range: Empirical Inquiry and Legal Policy,* in LAW IN A CHANGING AMERICA 56 (G. Hazard ed. 1968).

206. The majority in *Williams* professed an inability to discern precisely what the framers meant when they used the word "jury," but they acknowledged that the "usual expectation" in 1789 was that juries were to be made up of twelve. 399 U.S. 78, 98 (1970). Justice Harlan, concurring, emphasized the majority's disingenuousness in expressing even this degree of uncertainty. 399 U.S. at 122-29.

207. *See* note 55 *supra.*

208. Justice Harlan's concurring opinion, based on his rejection of the "incorporation theory," would not necessarily be undercut by empirical findings. However, re-

At the very least, the results of such research should force the Court to face the difficult question whether adverse results associated with juries of six are justified by the savings that smaller juries engender.[209]

---

search might show such an extreme increase in irrationality or prejudice when jury size is reduced that one taking Harlan's perspective would conclude that a jury of twelve is essential to due process in any case tried by a jury. My hunch is that the differences between six and twelve will not be extreme enough to sway one who accepts Harlan's position.

209. Good research on the exact magnitude of these savings would be particularly helpful. Lacking such research, the Court, if proved wrong on its initial premise, might rest an affirmance of *Williams* or *Colgrove* on the premise that the savings from a reduction in jury size are substantially greater than they actually are.

Three points relating to questions of empirical research and jury size have not been discussed in this paper, but are worth brief mention before concluding. First is the fact that a reduction in jury size implicates values unrelated to differences in verdicts. These values include such factors as the involvement of ordinary citizens in the ongoing processes of democratic government and the appearance of fairness to individuals on trial. Cutting jury size in half will substantially decrease the number of citizens who serve on juries or who have close acquaintances with such experience. A decrease in jury size also will decrease the percentage of minority group represen-tatives who sit on juries. *See* sections IIB, C *supra*. In a society as pluralistic as America is today, there may be a positive value in minimizing the number of situa-tions in which minority group members are judged by groups lacking minority rep-resentation, whether or not jury judgments would be affected by the presence of minority group members. These values have been largely ignored in this article, in part because the Court has largely ignored them in the jury-size cases and in part because I have attempted to focus this disscussion of empirical evidence on a single legal issue.

The second point worth mentioning before concluding is that, while the question of the unanimous verdict has not been discussed in this paper, the general approach of the analysis would apply to this problem as well. Statistical analysis and the social-psychological literature suggest that the effects of allowing nonunanimous verdicts are likely to be much more drastic than the effect of reducing jury size. Zeisel has called the nonunanimous verdict "reduction [in size] with a vengeance" and has demon-strated the statistical sense in which this is true. Zeisel, *supra* note 8, at 722. The de-liberation process particularly may suffer from a situation in which a majority does not have to contend rationally with the arguments of a minority sufficiently small to be outvoted. One can only speculate on what the effects of such a situation are likely to be on jurors in the overridden minority or on individuals who learn that they have been convicted over dissent. The Supreme Court has approved the use of non-unanimous verdicts in state courts, at least so long as nine votes out of twelve are required to convict. *See* Apodaca v. Oregon, 406 U.S. 404 (1972); Johnson v. Louisiana, 406 U.S. 356 (1972). It has not yet, however, faced a case in which only a majority of six was required to convict. Research on six-member juries with four, five, or six votes required to convict could be of great importance should such a case ever reach the Court.

A final point implicit in this article and applying generally to empirical research on legal issues is that the common strategy of gathering data on interesting law-re-lated problems "to see what is out there" is at best likely to be inefficient and at worst likely to produce misleading results. A researcher should have as good an idea as is possible of what he is likely to discover before he begins to collect data. At a mini-mum, existing bodies of relevant theory and data should be examined. With this as background, and with information gained from pre-tests, exploratory investigations, or general knowledge, the researcher can check his experimental design against the information he is likely to discover. Such a review can provide a forewarning of likely analytical problems, suggest interesting questions that can be asked of the data, and

indicate a need for collecting information on matters that would otherwise appear only tangentially relevant to matters being investigated. Where such a review allows the generation of reasonable hypotheses, the efficiency of the entire investigation is enhanced because one-tailed significance tests become appropriate, thus increasing the power of the tests. *See* H. BLALOCK, *supra* note 47, at 127-28, 188-93. Such a review may also suggest that, given the available data, the research is not worth doing because the likely results will have only the most equivocal relationship to the questions of legal interest.

I should also note that, as this article goes to press, the results of a number of recent studies on the jury-size question are beginning to be reported. Volume 2 of the Social Action & the Law Newsletter (Center for Responsive Psychology, Brooklyn College March, 1975) reports three studies: one by Valenti and Downing that found that when a case was prosecution-oriented six-member juries were more likely to convict and less likely to hang than twelve member juries, one by Snortum, Klein and Sherman that found that one assertive confederate arguing for conviction had more influence on juries of six than on juries of twelve, and one by Buckhout and Weg that found support for "the belief that 6-person juries tend to convict at a higher level of guilt (in a homicide) . . . ." *Id.,* at at 9. None of these studies is currently available in print, so I am unable to evaluate their methodological adequacy.

In another study, Beiser and Varrin, looking at civil cases in four New England federal district courts before and after the adoption of six-member juries, report that six-member jury trials took an average of 13.2 hours as compared to an average of 16.4 hours for twelve-member jury trials, Beiser & Varrin, *Six-member Juries in the Federal Courts,* 58 JUDICATURE 425, 430 (1975); the median trial-length, however, was 12.1 hours for both. *Id.* The probability of settlement, number of witnesses, time per witness and length of voir dire showed little or no relationship to jury size. *Id.,* at 428-30. Deliberations lasted somewhat longer in the twelve-member jury trials. *Id.,* at 430. The study also found that twelve-member juries more often found for plaintiffs and granted larger damage awards than did six-member juries, even after controlling for cause of action. *Id.,* at 431. However, the number of trials on which the analysis of verdicts is based is so small that little confidence should be placed in the verdict-related findings.

Finally, I understand that Drs. Barton and Padawer-Singer of the Bureau of Applied Social Research of Columbia University are engaged in a well-financed large-scale research project, investigating the differences between six- and twelve-member juries. Hopefully, this ongoing research will be able to discern whatever differences exist between juries of twelve and those of six, thus resolving some of the questions this article had to leave open.