1982

Death-Qualification and the Fireside Induction

Robert M. Berry

Follow this and additional works at: http://lawrepository.ualr.edu/lawreview

Part of the Constitutional Law Commons, and the Criminal Procedure Commons

Recommended Citation

Available at: http://lawrepository.ualr.edu/lawreview/vol5/iss1/1

This Article is brought to you for free and open access by Bowen Law Repository: Scholarship & Archives. It has been accepted for inclusion in University of Arkansas at Little Rock Law Review by an authorized administrator of Bowen Law Repository: Scholarship & Archives. For more information, please contact mmserfass@ualr.edu.
DEATH-QUALIFICATION AND THE "FIRESIDE INDUCTION"

Robert M. Berry*

The phrase "fireside inductions" has been used to describe those common-sense empirical generalizations about human behavior which derive from introspection, anecdotal evidence, and culturally transmitted beliefs. The psychology of the fireside roughly corresponds to what people believe about human behavior and how such behavior is to be described and explained. By and large, the law reflects these fireside inductions, which sometimes do and sometimes do not accord with behavioral science principles. There are occasional instances, however, when the fireside inductions are not represented in the law, and behavioral science research is examined to support, refute, or reconcile particular fireside inductions. Death-qualification is such an instance.

The problems which revolve around death-qualification introduce significant constitutional questions which are currently being considered in Grigsby v. Mabry. The focus of this article, however,

---

* B.A., M.A., and Ph. D., University of Arkansas at Fayetteville, Post-Doctoral study in psychobiology at the University of Chicago, Associate Professor of Psychology and Criminal Justice, University of Arkansas at Little Rock. (For information about unpublished studies cited in this article, contact the author).

2. Grigsby v. Mabry, 483 F. Supp. 1372 (E.D. Ark.), modified, reh'g granted, 637 F.2d 525 (8th Cir. 1980). The primary constitutional issue centers on whether death qualification violates the sixth and fourteenth amendment guarantees to a fair and impartial jury.
is on the psycho-legal issue of death-qualified juries: the lawmaker's use of behavioral science research and the fireside induction. This article first explores the issue of death-qualification and then the fireside induction relative to that issue. These topics are followed by a discussion of how behavioral science research might affect the fireside induction. Relevant research will then be examined. Finally, conclusions and recommendations will be discussed.

THE DEATH-QUALIFICATION ISSUE

Death-qualification is best considered in the context of the following table which illustrates the variety of potential community viewpoints toward administering the death penalty against an individual convicted of a capital offense.

<table>
<thead>
<tr>
<th>Attitude Toward Death Penalty</th>
<th>Automatic death penalty group</th>
<th>Favor death penalty group</th>
<th>Indifferent group</th>
<th>Oppose death penalty group</th>
<th>Automatic life-imprisonment group</th>
</tr>
</thead>
<tbody>
<tr>
<td>will always vote for death penalty against it in every case</td>
<td>favors death penalty but will not vote to impose it in every case</td>
<td>neither favors nor opposes the death penalty</td>
<td>opposes or has doubts about death penalty but will not vote against it in every case</td>
<td>will always vote for life imprisonment instead of death penalty</td>
<td></td>
</tr>
</tbody>
</table>

Until 1968, the practice in many states was to eliminate from capital cases those potential jurors who expressed opposition to the death penalty, i.e., those individuals characterized as belonging to either the “oppose death penalty” or the “automatic life-imprisonment” groups of Table 1. Exclusion was accomplished during the voir dire, when stated opposition to the death penalty could result in removal for cause. Elimination of opponents of the death penalty resulted in a “death-qualified” jury.

3. Hovey v. Superior Court, 28 Cal. 3d 1, 616 P.2d 1301, 168 Cal. Rptr. 128 (1980) (table is an adaptation of materials reported in Hovey).
Oberer traced the origin of the death-qualification practice to a time when conviction of a capital offense was automatically followed by the death penalty. Since a finding of guilt implied death, opposition to the death penalty conceivably could interfere with the determination of guilt. Death-qualified juries were therefore considered essential to the enforcement of the criminal law. The practice of death-qualification was retained even as variations in sentencing developed which functionally separated the trial process into two stages: a guilt phase and a sentencing phase. In the guilt phase, capital trial jurors must decide whether the defendant is guilty of a crime for which he may receive the death penalty. If guilt is determined during this initial phase, the jury must then decide whether to impose the death penalty in the sentencing phase.

In Witherspoon v. Illinois the United States Supreme Court altered the defining attributes of a death-qualified jury. The Court reasoned that it was "self-evident" that juries formed by excluding prospective jurors on the basis of "general objections to the death penalty" could not "speak for the community" and were "uncommonly willing to condemn a man to die." In the Court's view, procedural fairness required that "the decision whether a man deserves to live or die must be made on scales that are not deliberately tipped toward death." The Court ruled that the only prospective jurors who could constitutionally be eliminated because of their opposition to capital punishment were:

[T]hose who made unmistakably clear (1) that they would automatically vote against the imposition of capital punishment without regard to any evidence that might be developed at the trial of the case before them, or (2) that their attitude toward the death penalty would prevent them from making an impartial decision as to the defendant's guilt.

The practical consequence for most states was that capital trial jurors were excluded for cause only when they stated adamant op-

---

5. Id. at 550.
7. Id. at 518.
8. Id. at 522.
9. Id. at 520.
10. Id. at 521.
11. Id. at 521-22 n.20.
12. Id. at 522-23 n.21.
position to the death penalty. That is, a death-qualified jury by Witherspoon standards includes not only those jurors death-qualified by traditional standards, but also those potential jurors who comprise the "oppose death penalty" group of Table 1. In terms of Table 1, a jury qualified under Witherspoon standards excludes only those individuals adamantly opposed to the death penalty. This excluded group can also be designated as "Witherspoon-Excludables" (WEs).

The Court in Witherspoon also considered whether the exclusion of jurors with scruples against the death penalty resulted, at the guilt phase of a trial, in a "tribunal organized to convict" so that a representative, impartial determination of guilt could not be assured. Three unpublished studies were considered as evidence that death-qualified jurors were more likely to convict than were prospective jurors excludable on the basis of their opposition to the death penalty. The Court found these data "too tentative and fragmentary to establish that jurors not opposed to the death pen-

13. For a review of exclusionary practices among the various states, see Annot., 39 A.L.R.3d 550 (1971).
15. The petitioner in Witherspoon did not submit experimental evidence on this issue until the case was before the Supreme Court. The Court was requested to take judicial notice of three unpublished summaries of psychological research. These studies were:
   (a) W. Wilson, Belief in Capital Punishment and Jury Performance (University of Texas, 1964) (unpublished manuscript).
   (c) H. Zeisel, Some Insights into the Operation of Criminal Juries (University of Chicago, 1957) (unpublished manuscript), subsequently published as H. Zeisel, Some Data on Juror Attitudes Toward Capital Punishment (Center for Studies of Criminal Justice, University of Chicago Law School, 1968) [hereinafter cited as Zeisel].

After the petition for certiorari was filed, the Zeisel summary was withdrawn, apparently because Zeisel had not authorized its use in unfinished form. The brief submitted by Witherspoon relied on only the studies by Wilson and by Goldberg. Since these studies were available only in preliminary form, and since they had not been subjected to the scrutiny they would have received had they been offered as evidence to the trier of fact, the Supreme Court noted:

[We can only speculate . . . as to the precise meaning of the terms used in the studies, the accuracy of the techniques employed, and the validity of the generalizations made. Under these circumstances, it is not surprising that the amicus curiae brief filed by the NAACP Legal Defense and Educational Fund stated that with respect to bias in favor of the prosecution on the issue of guilt, the record is almost lacking in the sort of factual information that would assist the Court.

Witherspoon, 391 U.S. at 517-18 n.18.
alty tend to favor the prosecution in the determination of guilt”\footnote{391 U.S. at 517-18.} or that the “exclusion of jurors opposed to capital punishment . . . increases the risk of conviction.”\footnote{Id.}

Thus the Court did not decide the legal question whether a juror's attitude concerning the death penalty is so related to his tendency to convict or acquit that special rules are required for jury selection. The Court encouraged continued research on the question, however.\footnote{Id. at 520 n.18. The Court noted that a future defendant might still attempt to establish that a jury, death-qualified by \textit{Witherspoon} standards, “was less than neutral with respect to guilt.” \textit{Id.}} In terms of Table 1, the unresolved issue is whether the “automatic life-imprisonment” group, excludable by \textit{Witherspoon} standards, includes a subset of “\textit{Witherspoon-Excludable}” subjects who, in spite of their adamant opposition to the death penalty, could nevertheless render a fair and impartial determination of guilt and who should not, therefore, be excluded from participation at the guilt phase of a capital trial.\footnote{The continued exclusion of these subjects from the penalty phase of the trial is not in issue. Continued exclusion from both phases of a trial of those subjects whose opposition to the death penalty would prevent them from being fair and impartial is also not in issue.}

The group excludable under \textit{Witherspoon} standards is comprised of two distinct sub-groups which have in common that they would never vote for the death penalty in the \textit{sentencing} phase. The distinction between the sub-groups lies in their impartiality, or the lack of it, at the \textit{guilt} phase. One sub-group is \textit{not} impartial, knowing that a guilty vote may ultimately mean death. Members of this sub-group are conventionally described as “Nullifiers,” and it is agreed that they are properly excluded from both phases of a capital trial. Members of the other sub-group state that they \textit{can} be impartial on the issue of guilt or innocence. The issue raised by \textit{Witherspoon} is whether the latter sub-group should be included at the guilt phase of a capital trial. Members of the impartial sub-group are sometimes described by the term “Guilt-Phase Includables.”

The lawmaker who must decide whether to include the impartial sub-group may inquire whether there is a difference in the willingness to convict of two classes of potential jurors: those who are death-qualified by \textit{Witherspoon} standards and those who are excluded by \textit{Witherspoon} standards. To answer the question, the lawmaker may refer to the fireside induction and to behavioral science research.
THE FIRESIDE INDUCTION

The fireside induction suggests that in those cases in which the evidence is compelling, impartial jurors will vote to convict or acquit on the basis of the evidence, regardless of their attitude toward the death penalty. In equivocal or ambiguous cases in which the evidence is less compelling, however, jurors from the two "favor death penalty" groups at the left of Table 1 are more likely to vote for conviction; jurors from the two "oppose death penalty" groups at the right of Table 1 are more likely to vote for acquittal. Further, the strength of the tendency to convict or acquit is directly related to whether the juror holds one of the more extreme attitudes so that, for example, the automatic death penalty group would more likely vote for conviction in an equivocal case than would those who simply "favored" the death penalty. At the other end of the continuum, WEs would more likely vote for acquittal than would those simply "opposing" the death penalty. Broadly stated, the fireside induction suggests that proponents of the death penalty are conviction-prone and opponents of the death penalty are acquittal-prone.

Stated in terms of the issue raised by Witherspoon, the fireside induction is that WEs are less likely to convict (more likely to acquit) than a group of Witherspoon-qualified jurors.

The objection may be made that this estimate of the fireside induction is biased toward the defense because many defense attorneys not only articulate the belief that proponents of the death penalty are conviction-prone relative to opponents of the death penalty, but also endorse the status of the induction as that which "everyone knows." Indeed, Witherspoon and subsequent cases represent re-

20. The perception of one group as either conviction- or acquittal-prone relative to the other depends upon whether the group is viewed from the perspective of the defense or that of the prosecution. The defense typically chooses to focus on "conviction-prone" to describe jurors included in a capital trial, while prosecutors focus on excluded jurors as "acquittal-prone." The neutral question is whether there is a reliable difference in the frequency of guilty verdicts returned between proponents and opponents of the death penalty.

21. Comment, Grigsby v. Mabry: A New Look at Death-Qualified Juries, 18 AM. CRIM. L. REV. 145, 145-63 (1980). The author cites twenty-two cases heard in federal and state courts which have rejected the claim that the sixth amendment impartiality requirement is violated by death-qualified juries. Id. at 161-62 n.154. The author states:

A chronological analysis of the holdings of some of these cases may provide an insight into the reluctance of these courts to accept the open invitation of Witherspoon. . . . As each new study or piece of evidence came out after Witherspoon, starting with the finished drafts of the Witherspoon evidence. . . . the issue was relitigated. See State v. Ramirez, 116 Ariz. 259, 569 P.2d 201 (1977) (tacit admission that studies persuasive, but not enough since Witherspoon to dispel fragmentary label); Clark v. Fike, 538 F.2d 750 (7th Cir. 1976) (although definite "pattern or
curring attempts by defense attorneys to convince the courts that the induction is valid. However, since prosecutors have vigorously opposed incorporating the alleged induction into law, it may be hypothesized that there exists a countervailing prosecution induction to the effect that proponents of the death penalty are not more likely to convict than those who oppose the death penalty. The author submits that there is no countervailing induction and that the prosecution shares the view of the fireside induction held by the defense. The basis for the argument of a shared view of the induction will be fully developed below, in “The Fireside Revisited.” The conclusion to be anticipated is that the opposition by the prosecution, like the advocacy by the defense, is based on perceived self-interest.

The position of the defense is straightforward: jurors favoring the death penalty are conviction-prone; the inclusion of WEs would increase the chances for acquittal by neutralizing, to some extent, the tendency of death-qualified juries to convict. The position taken by the prosecution is slightly more complex; jurors opposed to the death penalty are seen as acquittal-prone; the inclusion of acquittal-prone WEs would decrease the probability of obtaining a conviction. In addition to the danger of a decreased conviction rate, proposed solutions to the accommodation of WEs in capital cases involve considerations of added cost and inconvenience. The irony of the prosecution’s position is that, because they think the fireside induction is valid and that legal implementation may hinder the prosecution’s case, it is in their interest to oppose incorporation of the induction into the law. As a consequence, the prosecution’s pro-induction view is cloaked in anti-induction rhetoric.

Before the behavioral science research is examined to support, refute, or reconcile the fireside induction, it would be instructive to examine the lawmaker’s use of behavioral science research and the fireside induction.

---

correlation" shown by studies, evidence still too fragmentary). On this one by one basis, however, the courts were reluctant to dispel the “too tentative and fragmentary” label when only one more study had been added since the last time the issue was approached. See Townsend v. Twomey, 452 F.2d 350 (7th Cir. 1971) (too few new studies since the evidence was last declared “too tentative and fragmentary”). Little heed was paid to the fact that many more studies had been added since the evidence was originally found to have been “too tentative and fragmentary.” In this way, twelve years of wholly consistent studies, no less than fourteen in number, had been disregarded as “tentative and fragmentary.”

Id. at 162 n.155.
RESEARCH AND THE FIRESIDE INDUCTION

The judicial process has not been quick to rely on the findings of behavioral science research. It was not until Ballew v. Georgia, in which the Court established a minimum jury size of six in state criminal cases, that "social science research was elevated from the footnotes into the text of a Supreme Court opinion." The primary reason for the lack of reliance on behavioral science research seems to be that very little of such research is available in a form that can assist judicial decisions, i.e., the research was either not designed to produce information relevant to legal issues or was conducted after the judicial decision. For the research to be useful to the lawmaker, it is necessary, as a preliminary, to determine how a study relative to the Witherspoon issue of death-qualification might be designed.

First, it would be necessary to identify a large number of potential jurors and reliably sort those jurors according to their attitudes toward the death penalty. Next, those jurors would participate in a real jury trial, involving a capital case, in which the evidence would be sufficiently equivocal to allow attitudes toward the death penalty to become part of the determination of guilt. At the trial's conclusion, various panels consisting of twelve jurors would be formed. Some panels would be comprised of death-qualified jurors by Witherspoon standards, some panels would be comprised of WE jurors, and some panels would be comprised of various combinations of these groups. The relative tendency of these groups to convict or acquit could then be assessed by comparing the proportion of cases in which each type of jury returned verdicts of not guilty or guilty (on various degrees of the offense).

In reality, such an ideal experiment could never be conducted because compelling ethical and legal restraints protect the trial process. In the absence of any kind of experimental evidence, lawmakers would seem to be forced to rely on their best estimate of the fireside induction. The ideal experiment can be approximated,

24. Id. This article discusses means by which data may be collected and presented in a form suitable for judicial consideration.
25. Katz reports an effort by Strodtebeck to record surreptitiously actual jury deliberations. The resulting controversy has effectively prevented further research of this nature. J. Katz, EXPERIMENTATION WITH HUMAN BEINGS: THE AUTHORITY OF THE INVESTIGATOR, SUBJECT, PROFESSIONS AND STATE IN THE HUMAN EXPERIMENTATION PROCESS (1972).
however, by the use of simulation. Assuming that the results obtained with simulated juries are similar to the results that would have been obtained with real juries, what kind of effect would various experimental outcomes have on a lawmaker's decision?

Meehl has addressed the general nature of the research-fireside induction relationship. He suggests that the most comfortable situation occurs "when there is a sizable and consistent body of research, experimental and nonexperimental (file data and field observation data), yielding about the same results as the fireside inductions." In this case, the congruence between behavioral science research and the fireside induction would provide a consistent basis for decisionmaking.

A slightly more difficult situation, according to Meehl, occurs when "No quantitative or experimental evidence is available or readily collectable before action must be taken. Here we rely upon the fireside inductions, these being all we have." The most difficult situations occur when the research results run counter to the fireside induction. In cases of research-fireside induction conflicts, the research results will probably be regarded as unconvincing counter-fireside induction evidence if the research consists of (1) only a few isolated studies with weak effects, (2) poorly designed studies in which internal validity is questionable, (3) studies conducted under conditions so artificial or remote that generalization to actual situations may not be valid, or (4) a collection of studies yielding inconsistent results. In research-fireside induction conflicts such as these, the lawmaker may be uncomfortable with a decision based on the fireside induction, but it is reasonable to expect that the minimal research support will cause the fireside induction to prevail.

Less difficult situations are encountered in those cases in which the research evidence still conflicts with the fireside induction, but the evidence is more convincing, *e.g.*, the studies show stronger effects; there are no problems of internal or external validity; and/or the studies produce consistent results. In such conflict situations the strong research evidence should decrease the lawmaker's discomfort, and the research results could be expected to guide the

---

26. This assumption can be questioned. For a general discussion of the simulation problem see Bermant, McGuire, McKinley, & Solo, *The Logic of Simulation in Jury Research*, 1 CRIM. JUST. & BEHAV. 224 (1974).
27. Meehl, supra note 1.
28. Id. at 27.
29. Id.
lawmaker's decision. The preceding discussion suggests a continuum of potential dilemmas which may be introduced by research which conflicts with the fireside induction.

BEHAVIORAL SCIENCE RESEARCH

The major empirical question is whether a jury formed under Witherspoon standards is more likely to convict than a similar jury which also includes WEs. The question can be addressed by comparing the guilt judgments rendered by Witherspoon-qualified jurors with the guilt judgments made by WEs in cases involving the same defendant.

Prior to a review of the most relevant studies, a set of three pre-Witherspoon studies will be considered. Since all three studies were conducted prior to establishment of the Witherspoon standard, researchers asked whether a jury formed by excluding all subjects having scruples about the death penalty was more likely to convict than a jury comprised of subjects not having such scruples. The comparative difference in conviction behavior between scrupled and non-scrupled subjects, broadly defined, does not directly address the Witherspoon issue since the scrupled group includes, from Table 1, (1) the Witherspoon-qualified “oppose death penalty” subjects, and (2) the two sub-groups which comprise the “automatic life-imprisonment” group. These early studies are therefore of marginal relevance to the empirical question derived from Witherspoon. They are strongly relevant to the fireside induction, however.

A. Pre-Witherspoon Studies

The Wilson Study. Wilson administered a brief written description of five criminal cases to 187 Texas college students and 61 college students in New York. At the same time, Wilson determined whether the students did or did not have scruples against capital punishment. The students were then asked to render a decision of guilty or not guilty for each case. Since one of the cases involved two defendants, the total number of guilty verdicts for any given subject could range from zero to six. Table 2 illustrates Wilson's results.

The study revealed a systematic tendency for non-scrupled sub-

---

30. Supra note 15.
31. Id.
### TABLE 2
Percentage of Guilty Verdicts for Scrupled and Non-scrupled Subjects

<table>
<thead>
<tr>
<th>Number of Guilty Verdicts</th>
<th>Subjects' Death-Penalty Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scruples Against Death Penalty</td>
</tr>
<tr>
<td>5 - 6</td>
<td>17.0%</td>
</tr>
<tr>
<td>3 - 4</td>
<td>61.3%</td>
</tr>
<tr>
<td>0 - 2</td>
<td>21.7%</td>
</tr>
</tbody>
</table>

Subjects to reach a guilty conclusion more often than scrupled subjects. Almost 22% of the scrupled subjects voted guilty on two or fewer occasions, compared with less than 9% of the non-scrupled subjects. Only 17% of the scrupled subjects voted guilty in five or in all six cases, while almost 30% of the non-scrupled subjects did so. These differences would have occurred by chance less than 2% of the time. The Wilson study was thus the first study to demonstrate that subjects without scruples against the death penalty were more likely to convict than subjects with scruples. The Wilson results support the fireside induction.

Wilson also found that non-scrupled subjects were more confident in their decisions concerning guilt and that they assessed more severe penalties. Further, the students expressed their agreement or disagreement with fifteen statements measuring bias toward the prosecution and bias against the insanity defense. The results strongly suggested that non-scrupled subjects were biased in favor of the prosecution and against the insanity defense.

One study is seldom definitive in resolving an empirical question in behavioral science research, and the Wilson study is vulnerable to several criticisms. In the first place, even though simulation procedures have proven their scientific utility, there may be skepticism about extrapolating from a situation involving brief written descriptions of criminal cases to an actual capital case. Further, as one wag has suggested, psychological science seems largely based on data collected from white rats and college sophomores. While Wilson’s students would seem to be more like actual jurors than

---

32. Unfortunately, the original source of this suggestion is unknown.
would rats, the extent to which students are representative of real jurors is not known, even though many of Wilson’s subjects may have been members of the actual jury population. A third problem to be noted is that Wilson’s data seem most relevant to juror, not jury, behavior. The student jurors simply indicated their guilt judgments with no opportunity for group deliberation. This procedure most closely resembles a first ballot in actual jury procedure. Finally, as previously noted, the broad comparison between scrupled and non-scrupled subjects is not the appropriate comparison for the issue raised by Witherspoon.

Wilson’s results are nonetheless important as an empirical demonstration of the link between attitude toward the death penalty and (simulated) conviction behavior and as support for the fireside induction. 

The Goldberg Study.\textsuperscript{33} Goldberg independently conducted a study very similar to Wilson’s and reported similar results. Goldberg asked 200 undergraduate students in Georgia whether they had scruples against the death penalty. Sixty-one percent (122) indicated they had such scruples. The remaining 39\% (78) presumably answered that they had no such scruples. Each student-juror judged the guilt or innocence of defendants in sixteen written descriptions of criminal cases involving the death penalty. Each defendant could be found guilty of first degree murder or guilty of a lesser crime, or could be acquitted on the grounds of insanity or for lack of evidence. If judged guilty, a sentence was assigned which ranged in severity from probation to the death penalty. Goldberg reported four primary comparisons between scrupled and non-scrupled subjects, which appear in Table 3.

All of the differences reported in Table 3 are in the same direction as Wilson’s results and, as a consequence, provide further support for the developed view of the fireside induction. It should be noted, however, that the conventional decision rule adopted by Goldberg\textsuperscript{34} led to the conclusion that the observed difference in guilty verdicts (and the difference between first degree murder verdicts) between scrupled and non-scrupled subjects could reasonably

\textsuperscript{33} Supra note 15. 

\textsuperscript{34} To determine whether the differences between the means of two groups are significant, the threshold question to be resolved is whether the means of two groups differ only by chance. In tests of other groups of scrupled and non-scrupled subjects, or in tests of Wilson’s and Goldberg’s subjects a second time, the scores on the second test might not be the same as those on the first—they might even be reversed. This occurrence would generate skepti-
be attributed to chance rather than to attitude toward the death penalty.35

TABLE 3
Percentage of Different Verdicts for Scrupled and Non-scrupled Subjects

<table>
<thead>
<tr>
<th>Juror's Verdict</th>
<th>Subjects' Death-Penalty Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scruples Against Death Penalty</td>
</tr>
<tr>
<td>Guilty: 1st Degree Murder</td>
<td>44%</td>
</tr>
<tr>
<td>Guilty: Lesser Offense</td>
<td>25%</td>
</tr>
<tr>
<td>Not Guilty: Insanity</td>
<td>14%</td>
</tr>
<tr>
<td>Not Guilty: Lack of Evidence</td>
<td>17%</td>
</tr>
</tbody>
</table>

Goldberg also reported important racial differences relative to death penalty attitudes. Black respondents were more likely to oppose the death penalty (76%) than white respondents (47%). No sig-

cism about the significance of the differences, and the conclusion might be drawn that whatever apparent differences existed were due merely to chance.

Statistical procedures are used which examine the hypothesis that there are no genuine differences between the groups being tested (called the null hypothesis). The statistic yields a numerical statement of the probabilities of any observed differences between groups being the result of chance instead of the factors presumed to be causing them—in this case, attitudes toward the death penalty.

Before the conclusion can be drawn that an observed difference is a genuine one and is not due to chance, a probability value of 5% or .05 must be reached. The probability value of .05 has, by convention, been adopted by researchers in the behavioral sciences as the highest appropriate risk for concluding that the mean differences are genuine and are not attributable to chance. For example, if the probability is higher than 5% that the results would have occurred by chance, the researcher will conclude that chance can reasonably account for the observed difference between groups. The researcher's decision might, of course, be wrong, because he or she will erroneously conclude that real differences are due to chance 5% of the time. (This failure to reject a null hypothesis which should be rejected is called a Type I error. The erroneous conclusion that there is a genuine difference when the differences are due to chance is called a Type II error). The 5% convention is, therefore, simply an agreed-upon rule for decision based upon the relative consequences of making a Type I, as opposed to Type II, error. In circumstances in which there is particular concern about a Type II error, a probability level of .08, .10, or even .20 might be preferred. A probability level higher than .05 would seem warranted when considering the studies reviewed here since judicial decisions based on their results may literally mean the difference between life and death.

significant differences in attitude were observed between male and female respondents.

The same criticisms that applied to Wilson's study, and perhaps other criticisms, apply to Goldberg's study. Again, the extent to which the college students were representative of actual jurors is unknown; the stimuli involved written case descriptions; there were no deliberations of the kind engaged in by juries; and the comparison was between broadly defined scrupled and non-scrupled subjects.

The significance of the last point is demonstrated by what Goldberg regarded as a curious feature of her data. Fifty-two of her scrupled subjects imposed one or more death sentences, a finding which Goldberg regarded as inconsistent. These "inconsistent" subjects can simply be regarded as samples from either or both of the "oppose death penalty" or WE groups of Table 1. At the other end of the death penalty continuum, five non-scrupled subjects were "inconsistent" in imposing no death sentences. These "inconsistent" subjects apparently belonged to either the "Indifferent" or "Favor Death Penalty" groups.

Although it was available to the Witherspoon Court, the Goldberg study is too seriously flawed to have contributed to the Witherspoon decision other than as a "tentative and fragmentary" study. The primary flaws derive from Goldberg's treatment of her results. It would have been informative to have reported (as Wilson did) the number of subjects in each group who convicted no defendants, those who convicted only one defendant, and so on through the maximum possible number of convictions (sixteen). The analysis of data organized in this manner is relatively straightforward. Goldberg did not present her results in this manner and it is not at all clear how she analyzed the results she did report. There is a distinct possibility that she used an inappropriate analysis and/or made a serious computational error.36

In view of these important reservations, the prudent course is to minimize the significance of the Goldberg results which directly compare the conviction performance of scrupled and non-scrupled subjects. Acknowledging the reservations, the data still permit the statement that Goldberg's results are consistent with the fireside in-

36. Goldberg might have been expected to count each subject sixteen times, one time for each case. This would generate $122 \times 16 = 1952$ votes. A chi-square analysis of such total votes yields a significant result. However, Goldberg suggested that she utilized a chi-square analysis and obtained $p=.08$, a result that was not significant. A chi-square analysis is not appropriate to analyze the total voting frequencies because the votes are not independent, as chi square requires. The actual statistical procedure used by Goldberg is not clear.
duction since the differences reported in Table 3 are in the same direction as those of Wilson.

The common criticisms directed toward the Wilson and Goldberg studies are fairly obvious ones growing out of the context in which the research was conducted. Because of the ethical and legal restraints which prohibit the study of actual jurors in actual trials, it would seem that behavioral science research is limited to offering only the most realistic simulation it can design. Surprisingly, however, there is one study of actual scrupled and non-scrupled jurors which examined jurors' courtroom behavior in actual trials. This unique study by Hans Zeisel was initially available to the Witherspoon Court but was withdrawn prior to the Witherspoon hearing because it had not been completed or analyzed.

The Zeisel Study. Zeisel interviewed jurors who had participated in actual criminal cases and asked them three key questions on the last day of their term of service:

1. Whether they had scruples against the death penalty.
2. How they voted on the first ballot after the jury began its deliberations.
3. How the entire jury voted on the first ballot.

The survey results generated a "grab bag of Guilty and Not Guilty votes on many different cases, where, most of the time, we did not even have all twelve jurors from a particular case. . . . How were we to compare a Guilty vote in a case where the jury voted unanimously Guilty on the first ballot, with a Guilty vote that was given on the first ballot in opposition to eleven jurors who voted for acquittal?"

Zeisel reasoned that the total number of guilty votes cast on the first ballot indicates the general thrust of the evidence. Since unanimous first-ballot votes did not reveal potential differences between scrupled and non-scrupled subjects, Zeisel analyzed the 464 split first-ballot votes. These 464 votes were grouped into "constellations," according to whether they were given in a first ballot in which there was only one vote for guilty, two votes for guilty, and so forth, through eleven first-ballot votes for guilty. Figure 1 presents, for each voting constellation, a comparison between the proportion of jurors with and without scruples against the death penalty who voted guilty on the first ballot.

37. Supra note 15.
Several observations are appropriate relative to Figure 1. First, Zeisel's concept of the "strength of the evidence" seems to be supported. The proportion of subjects voting guilty, both scrupled and non-scrupled, is functionally related to the number of first-ballot guilty votes. Second, there is an interesting difference between the proportion of scrupled and non-scrupled subjects voting guilty as a function of the strength of the evidence. Non-scrupled subjects reached the point of "equal-likelihood" of voting either guilty or not guilty (p=.50) when the strength of the evidence yielded only four first-ballot guilty votes, i.e., when the evidence was relatively weak. In contrast, scrupled subjects did not exceed the point of "equal-likelihood" until there were eight first-ballot votes for guilty, i.e., when the evidence was relatively strong. Finally, non-scrupled subjects voted guilty more often than scrupled subjects ten times out of the eleven possible comparisons. Had there been only chance differences between scrupled and non-scrupled subjects, then non-scrup-
pled subjects could be expected to convict more often than acquit in ten or eleven of the comparisons fewer than six times in a thousand. The chance odds of 172 to 1 against this rare event strongly suggest that the differences observed by Zeisel represent real differences in conviction behavior between the two groups and not simply chance differences.\(^{39}\)

Because it used actual jurors and actual trials, the Zeisel study is an important demonstration that the differences between groups in conviction behavior extend to actual situations and are not limited to simulations. An important reservation, of course, is that Zeisel's scrupled and non-scrupled groups, like Goldberg's and Wilson's, were only broadly defined. This is hardly surprising since all three studies were conducted prior to the Witherspoon decision and could not have been expected to form their comparison groups on the basis of Witherspoon criteria.

B. Post-Witherspoon Studies

The appropriate comparison groups for a study of the Witherspoon issue would involve a contrast between a group of subjects death-qualified by Witherspoon standards and a group of subjects who (1) indicated they would vote against the death penalty in every capital case, regardless of the evidence, or (2) could not be fair in judging a capital defendant's guilt or innocence. Three studies are now available which have attempted to reflect the appropriate comparison groups.

**The Jurow Study.**\(^{40}\) Jurow's subjects consisted of 211 industrial workers roughly comparable to many actual juror pools. New York judges responsible for court administration refused Jurow permission to use prospective jurors. Nevertheless, one-third of the 211 subjects had previously participated in jury service. The use of a nonstudent subject pool diminishes the critical issue of the representativeness of the subject group associated with the earlier Wilson and Goldberg studies.

---

39. Figure 1 is adapted from Zeisel's Table 9. The percentages reported by Zeisel for each voting constellation were first converted to raw frequencies and the "unknown" votes were eliminated. Then, for each constellation, the number of scrupled subjects who actually voted guilty was expressed as a percentage of the total number voting either guilty or not guilty. The same procedure was followed to determine the percentage of non-scrupled subjects who voted guilty.

Each subject was administered the two-part Capital Punishment Attitude Questionnaire (CPAQ), developed by Jurow to assess general attitudes toward capital punishment and, in part B, to assess how the subject would consider the death penalty if serving on a jury. The CPAQ-B is directly relevant to the *Witherspoon* question because subjects responding to statement one on the scale were identified as belonging to that *Witherspoon*-excludable category of subjects who indicated they would always vote against the death penalty, regardless of the evidence. No questions were asked about whether the subjects could or could not be fair in judging a capital defendant's guilt or innocence. The *Witherspoon*-relevant CPAQ-B is reproduced below as Table 4 and also includes the distribution of subjects endorsing each alternative.

**TABLE 4**

Capital Punishment Attitude Questionnaire (CPAQ), Part B, and percentage distribution of subjects endorsing each alternative.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Choice</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>1</td>
<td>I could not vote for the death penalty regardless of the facts and circumstances of the case.</td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
<td>There are some kinds of cases in which I know I could not vote for the death penalty even if the law allowed me to, but others in which I would be willing to consider voting for it.</td>
</tr>
<tr>
<td>63%</td>
<td>3</td>
<td>I would consider all of the penalties provided by the law and the facts and circumstances of the particular case.</td>
</tr>
<tr>
<td>5%</td>
<td>4</td>
<td>I would usually vote for the death penalty in a case where the law allows me to.</td>
</tr>
<tr>
<td>2%</td>
<td>5</td>
<td>I would always vote for the death penalty in a case where the law allows me to.</td>
</tr>
</tbody>
</table>

After completing the CPAQ and several other attitude scales,\(^{41}\)

---

41. The other scales administered by Jurow were the Thurstone Capital Punishment Attitude Scale, the Conservatism-Liberalism Scale, the California F Scale, the Legal Attitudes Questionnaire, the Average Penalty Scale, and the Death Penalty Scale.
Jurow’s subjects listened to two audiotapes of simulated criminal trials, including abbreviated opening statements, examination of witnesses, closing arguments, and the judge’s instructions to the jury. The first thirty-three minute tape involved the murder of a liquor store proprietor during a holdup, and the second fifty minute tape involved a narcotics addict charged with robbing, raping, and killing a girl in her apartment.

After listening to the tapes, subjects considered their verdict and voted guilty or not guilty. The relationship between judgments of guilt and the endorsed CPAQ-B item are reproduced in Table 5. Table 5 also reports, for each group, whether the direction of the majority vote was to convict or acquit.

| Case I | | Case II | |
| CPAQ-B Statement | Number voting | CPAQ-B Statement | Number voting | |
| Endorsed | Guilty | Innocent | Majority Vote | |
| 1 | 7 | 14 | 67% Acquit | 9 | 12 | 57% Acquit |
| 2 | 12 | 30 | 71% Acquit | 25 | 17 | 60% Convict |
| 3 | 59 | 73 | 55% Acquit | 76 | 56 | 57% Convict |
| 4 | 10 | 1 | 91% Convict | 9 | 2 | 82% Convict |
| 5 | 4 | 1 | 80% Convict | 4 | 1 | 80% Convict |

The two cases were not equally likely to generate guilty judgments. The average judgment of the first case was not guilty, a judgment made by 56.4% of the subjects. The average judgment of the second case was guilty, a judgment made in 58.3% of the cases. Both cases thus seem to be relatively equivocal in evoking judgments of guilt or innocence, one criterion developed earlier to characterize an “ideal” study. Had they been ideal, both of Jurow’s cases would have produced a three-two or two-three majority-vote split among
the five groups. The four-one split for conviction in Case II suggests that the "strength of the evidence" in Case II may have been too strong to permit detection of potential group differences.

For Case I, the vote to convict or acquit was systematically related to the subject's attitude toward the death penalty. Subjects endorsing CPAQ-B statements one, two, and three were more likely to acquit than convict; subjects endorsing statements four and five favoring the death penalty were more likely to convict.

These same data can be examined in a slightly different way. If the average value of the CPAQ-B endorsed by subjects who convicted is compared to the average value endorsed by subjects who acquitted, the result is that subjects who convicted in Case I scored significantly higher in the direction of favoring the death penalty on the CPAQ-B than subjects who acquitted.

For Case II, the results were in the same direction as in Case I, as evidenced by the changing percentage of particular CPAQ-B groups who voted guilty. An overall analysis of these differences was not significant, which suggests only a chance relationship between death-penalty attitude and conviction behavior for Case II. However, the average CPAQ-B item endorsed by those convicting in Case II was significantly higher than it was for those acquitting. The discrepancy between the conclusions from these two measures may derive from a combination of (1) the relatively low frequencies in groups one, four, and five, (2) the relative insensitivity of Case II as a measuring device because of the strength of the evidence, and (3) the shift toward guilty judgments from Case I to Case II for CPAQ-B groups two and three.

With respect to the third possibility above, Jurow actually examined the shift in voting patterns between Case I and Case II. He

42. Supra Table 4.
43. Id.
44. The obtained value of $t$ was 3.36. If the statistical test inquires whether the groups were different, it is a non-directional test and, in this case, $t$ would be associated with a critical value of 1.97. On the other hand, if the statistical question specifies in advance that one particular group should score higher than the other (as all these studies do), then a directional test may be appropriate. In this case, the critical value of $t$ for a directional test is 1.65. Since the obtained $t$ of 3.36 exceeded both critical values, the directional and non-directional tests lead to the same conclusion.

45. The obtained value of $t$ was 1.86 which is less than the critical value of 1.97 associated with a non-directional test (supra note 44). This comparison led Jurow to conclude that there was no significant difference in the average CPAQ-B item endorsed by those convicting or acquitting in Case II. The author's conclusion that Jurow's results demonstrate a significant difference is based on the observation that the obtained $t$ of 1.86 exceeds the critical value of 1.65 associated with a directional test.
DEATH-QUALIFIED JURIES

noted that "groups one, four, and five are relatively constant in their voting behavior, the former [sic] constantly voting to acquit and the latter two to convict. Whatever flexibility in voting behavior is exhibited by the subjects comes largely from groups two and three."46

The methodological criticisms applicable to the Goldberg and Wilson studies are diminished in the Jurow study. While it is true that Jurow studied juror behavior, and not jury behavior, he used Witherspoon-relevant groups, used more realistic audiotaped simulation materials, and his subjects approximated the characteristics of real jurors. Nevertheless, Jurow obtained results consistent with those of prior studies. In short, Jurow's results support the conclusion that the results reported in the pre-Witherspoon studies occurred in spite of, and not because of, methodological limitations.

The Harris Study.47 White48 reported a Harris poll in which 2068 adults, a stratified sample of the United States, were asked whether, in a murder trial, "there would be any situation in which you might vote for the death penalty, or do you think you could never vote for the death penalty, regardless of the circumstances?"49 Responses to this question identified the same comparison groups as those investigated by Jurow.

Each subject-juror was instructed about such legal principles as reasonable doubt and the prosecutor's burden of proof, after which they were given written descriptions of four criminal cases. The type of case and the percentage of each group voting guilty appears in Table 6. In the first three cases, the results were regarded as statistically significant in that each of the observed differences would have arisen by chance fewer than one time in a thousand. In the fourth case, the observed difference was in the expected direction but, because there was a one in ten possibility that such a difference might be due to chance, the results were not regarded as statistically significant. In three of the four cases, therefore, the results were highly reliable in demonstrating that subjects who could never vote for the death penalty voted to convict less often than death-qualified jurors. Thus, the fireside induction is again supported, this time with a national sample of subjects in which the comparison groups were distinguished on the basis of relevant Witherspoon criteria.

---

46. Jurow, supra note 40, at 583-84.
49. Hovey v. Superior Court, 28 Cal.3d 1, 37, 616 P.2d 1301, 1322, 168 Cal. Rptr. 128, 149 (1980) (quoting the precise language used in the Harris Study).
However, like the previous studies, the results of jury deliberations were not investigated.

### TABLE 6

Percentage of Guilty Verdicts in Four Cases by *Witherspoon*-Excludable and *Witherspoon*-Qualified Subjects

<table>
<thead>
<tr>
<th>Case</th>
<th>Classification by Death-Penalty Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Witherspoon</em>-Excludable</td>
</tr>
<tr>
<td>Typewriter Robbery</td>
<td>57%</td>
</tr>
<tr>
<td>Manslaughter</td>
<td>67%</td>
</tr>
<tr>
<td>Assault of Officer</td>
<td>32%</td>
</tr>
<tr>
<td>Automobile Larceny</td>
<td>69%</td>
</tr>
</tbody>
</table>

The methodological improvements associated with the Jurow results and with the Harris results make the research results considerably less "tentative and fragmentary" than was the evidence at the time of the *Witherspoon* hearing. Even so, a major remaining reservation may be expressed relative to the composition of the group defined as *Witherspoon*-Excludable in both the Jurow study and in the Harris study. In both cases the identified WEs were relevant to the first exclusionary criterion because the WE subjects indicated they would vote against the death penalty in every capital case, regardless of the evidence. The current issue, however, turns on the second criterion: the exclusion of subjects adamantly opposed to the death penalty who could not be fair in judging a capital defendant's guilt or innocence.

A legitimate reservation, therefore, is that the WE groups used in the Jurow and Harris studies may have included Nullifiers and are thus too inclusive to indicate whether a subset of WEs, described earlier as Guilt Phase Includables, differ in conviction behavior from those who are death-qualified by *Witherspoon* standards.

The kind of study needed is one involving a realistic simulation in which the conviction performance of *Witherspoon*-qualified jury-eligible subjects is compared with WEs, excluding Nullifiers. Ideally, the study would examine jury performance in addition to juror
behavior. Precisely such a study was recently completed by Phoebe Ellsworth and her associates.

*The Ellsworth, Thompson, and Cowan Study.* Ellsworth eliminated the Nullifiers from a sample of jury-eligible adults in California by asking two basic questions. The first question asked whether potential subject-jurors would or would not consider voting to impose the death penalty in at least some capital cases. A second question asked whether, in the guilt phase of a trial, the subjects would "(a) . . . decide the question of guilt or innocence in a fair and impartial manner based on the evidence and the law, or (b) . . . not be fair and impartial in deciding the question of guilt or innocence, knowing that if the person was convicted he or she might get the death penalty." Subjects choosing alternative (b) were identified as Nullifiers and eliminated from the experiment. Subjects who expressed adamant opposition to the death penalty but who also selected alternative (a) were identified as (Guilt Phase Includable) WEs. The final sample consisted of 258 death-qualified subjects and 30 WEs, a total of 288 (of whom 45% had experienced prior jury service).

**TABLE 7**

Percentage of Witherspoon-Excludable and Witherspoon-Qualified Subjects Voting for Various Verdicts

<table>
<thead>
<tr>
<th>Classification by Death-Penalty Attitude</th>
<th>Witherspoon-Excludable</th>
<th>Witherspoon-Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Degree Murder</td>
<td>3.3%</td>
<td>7.8%</td>
</tr>
<tr>
<td>2nd Degree Murder</td>
<td>23.3%</td>
<td>21.3%</td>
</tr>
<tr>
<td>Manslaughter</td>
<td>26.7%</td>
<td>48.9%</td>
</tr>
<tr>
<td>Acquittal</td>
<td>46.7%</td>
<td>22.1%</td>
</tr>
</tbody>
</table>

The subjects were then shown, in groups ranging in size from 12 to 36, a two-and-one-half hour scripted, videotaped reenactment of an actual criminal trial, regarded in pretesting as convincing and

---


51. Id.
realistic. After viewing the videotape, each subject voted for a verdict of either not guilty, or guilty of first degree murder, second degree murder, or voluntary manslaughter. The results appear in Table 7. The pattern of results appearing in Table 7 could have arisen by chance fewer than one time in a hundred. The results, therefore, are significant and demonstrate that subjects who were death-qualified by Witherspoon standards were more likely to convict than subjects who would never vote for the death penalty, but who could be fair and impartial at the guilt phase of a trial.

After recording their individual decisions on the issue of guilt, most of the subjects were divided into twelve-person juries and allowed to deliberate. Some juries were composed exclusively of Witherspoon-qualified jurors while some were a mixture of Witherspoon-qualified jurors and (up to four) WEs. Although instructed to try to reach a unanimous verdict, no jury had reached a verdict after one hour of deliberation. Deliberations were then stopped and personal verdicts were again recorded. The results appear in Table 8.

TABLE 8

Percentage of Witherspoon-Excludable and Witherspoon-Qualified Subjects Voting for Various Post-deliberation Verdicts

<table>
<thead>
<tr>
<th>Classification by Death-Penalty Attitude</th>
<th>Witherspoon-Excludable</th>
<th>Witherspoon-Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Degree Murder</td>
<td>3.4%</td>
<td>1.0%</td>
</tr>
<tr>
<td>2nd Degree Murder</td>
<td>13.8%</td>
<td>17.3%</td>
</tr>
<tr>
<td>Manslaughter</td>
<td>48.3%</td>
<td>68.0%</td>
</tr>
<tr>
<td>Acquittal</td>
<td>34.5%</td>
<td>13.7%</td>
</tr>
</tbody>
</table>

Of those subjects who deliberated, there was an unexplained loss of sixty-one subjects from the Witherspoon-qualified group and one subject from the WE group. No information is available on the original votes of these “lost” subjects, and no information is available on the overall pattern of voting shifts. The results in Table 8 are, therefore, best viewed independently of the data appearing in Table 7. An analysis of the data appearing in Table 8 demonstrates that the observed pattern of voting differences between the two
groups would have occurred by chance less than 5% of the time and so indicates a systematic voting difference. The same pattern of results observed as juror behavior before deliberation was also observed after jury deliberation. It would thus appear that the similar patterns of predeliberation results observed in the earlier studies were differences that would also characterize jury, as well as juror, behavior.

The six reviewed studies, along with the fireside induction, form a collective "package" with consistent, convergent results. There is a conspicuous absence of studies reporting contrary results or even conditions limiting the effect of death-penalty attitudes on conviction behavior. A contemporary lawmaker, relying on a judicious mixture of research and the fireside induction, might find that the package provided convincing evidence that there are real differences between WEs and Witherspoon-qualified jurors in their willingness to convict.

C. The ADP Issue

In Hovey v. Superior Court California lawmakers found the package convincing relative to the Witherspoon issue, but not relevant to death-qualification as practiced in California. California provides for the exclusion of not only those who, under Witherspoon, would never vote for the death penalty, but also for the exclusion of those jurors who would always vote for the death penalty. This group corresponds to the Automatic Death Penalty (ADP) group at the extreme left of Table 1. The conclusion in Hovey was that studies relevant to California procedure would compare a combined group of California excludable subjects (WEs and ADPs) with California death-qualified subjects. The current package therefore consisted of studies which provided convincing evidence of an irrelevant issue.

In Grigsby v. Mabry the State of Arkansas also contended that

52. The "package" actually includes numerous additional studies demonstrating other attitudes correlated with death-penalty attitude, survey results demonstrating that blacks and females are more likely to oppose the death penalty and thus are more likely to be excluded than whites and males, and a study demonstrating that the death qualification process itself increases the probability of conviction.

53. 28 Cal. 3d 1, 616 P.2d 1301, 168 Cal. Rptr. 128 (1980).

54. Id. at 68, 616 P.2d at 1346, 168 Cal. Rptr. at 173.

the reviewed studies proved the "wrong" hypothesis. The state alleged that Arkansas, like California, eliminates both ADPs as well as WEs from capital cases. This exclusion of ADPs casts doubt on the generalizability of any study which fails to identify ADPs and classify them with the other excludables.

In *Hovey*, the defense contended that the size of the ADP group was disproportionately small relative to the size of the WE group. If such were the case, then the subtraction of a few ADP jurors from the large class of *Witherspoon*-qualified jurors would be of little consequence to the results of the reviewed studies. The *Hovey* court, however, had no reliable basis from which the relative size of the ADP group could be estimated; tentative indications were that it "may be as small as 1 percent (or less) of the adult population or as large as 28 percent."56 In the absence of information concerning the size and voting performance of the ADP group, the *Hovey* court concluded that there was not enough evidence to support a constitutional decision.

After *Hovey*, it was not clear whether death-qualification by excluding both WEs and ADPs resulted in a disadvantage to the defendant. Whether the exclusion does disadvantage the defendant obviously depends on the relative distribution of WEs and ADPs. If the number of WEs is disproportionately large relative to ADPs, then a jury formed by excluding a large number of WEs and only a few ADPs would operate to the disadvantage of the defendant. It would result in a jury prone to convict, like those described in the six reviewed studies. The State of Arkansas introduced a study in *Grigsby*, however, which suggested that the size of the ADP group was not trivial, but substantial.

*The Shure Study.*57 Shure conducted a telephone survey of an area around West Los Angeles. The jury-eligible status of the respondents is unknown but a small percentage (3%) of the 369 respondents were reported to have experienced prior jury service. On the basis of responses to *Witherspoon* questions developed and used by Ellsworth, Shure initially classified 310 of his respondents as death-qualified and 59 as WEs. On the basis of further questions, the respondents were reclassified. A central question was whether the respondent would vote for the death penalty in a specific case, a question apparently adapted from Jurow's Case I but much more

56. 28 Cal. 3d at 65, 616 P.2d at 1344, 168 Cal. Rptr. at 171.
heinous. The case involved a drug addict who allegedly broke into an apartment where a young college girl and her boyfriend were studying. The boyfriend was shot to death and the girl was raped and shot in the head five times. She survived but was left paralyzed, mute, and partially blind. Based primarily on their response to this case, respondents were reclassified. The reclassified sample was comprised of 84 WEs (22.76%), 123 ADPs (33.33%), and 162 (43.9%) death-qualified by California standards.

If Shure's data are representative, the relative size of the ADP group is substantial and the State's (Arkansas') argument should be seriously considered. Shure's procedure, however, may have been seriously flawed. For example, ADPs were identified according to whether they would vote for the death penalty in a specific, atrocious case rather than whether they would always vote for the death penalty. Further, the results of other studies suggest that Shure's results are not at all representative and that the ADP population is relatively small.

The Jurow Study. The CPAQ-B administered by Jurow and the distribution of responses which appears in Table 4 should be compared with Shure's results. Statement five of the CPAQ-B is an ADP question which was endorsed by only 2% of Jurow's sample, a result in marked contrast to the 33% reported by Shure.

The Hamilton Study. Hamilton administered the CPAQ to 167 university students in New York. The distribution of responses to the CPAQ-B were that 35% of the respondents were classified as WE, 64% were death-qualified by California standards and only 1% were ADPs. The Hamilton results, like those of Jurow, indicate that the ADP population is extremely small both in an absolute sense and relative to the WE population.

The Berry and Butler Study. Berry and Butler administered the CPAQ to a sample of 91 subjects in Arkansas, consisting of employees of Teletype Corporation, the U.S. Corps of Engineers, and adult students at the University of Arkansas. More than one-half of

58. Supra note 40.
this sample was eligible for jury service. The distribution of responses to the CPAQ-B was such that 10% were classified as WE by endorsing statement one, 88% were death-qualified by California standards, and only 1% were classified as ADP. Again, the results indicate that the size of the ADP group is negligible.

The 1981 Harris Study. A national telephone poll was conducted by the Harris organization in January 1981. The sample consisted of 1499 adults from 100 clustered telephone exchanges, stratified by geographic region and metropolitan/non-metropolitan residence. Respondents were first asked a question about their general views on capital punishment. The 256 who reported that they strongly favored the death penalty were then asked whether they could be fair and impartial in determining guilt or innocence in a capital case. The 213 who said they could be fair were then asked the final ADP-defining question of whether they could always vote to impose the death penalty. Thirteen respondents indicated they would always vote for death, hence thirteen subjects from the national sample of 1499 were classified as belonging to the ADP group. After standard weighting procedures were applied, the data revealed that ADP classification defines less than 1% of the adult American population.

The Cowan and Thompson Study. Cowan and Thompson conducted telephone interviews with ninety-nine subjects randomly selected from the jury rolls in Santa Clara County, California. In response to a question concerning their general attitude toward the death penalty, nine subjects reported they strongly favored it. None of the nine, however, indicated that they would always vote for the death penalty in a capital case. Cowan and Thompson were thus unable to locate any ADPs for their proposed experiment on ADP behavior, which led them to abandon their experiment as “impracticable” given the low frequency of ADPs.

The Arkansas Archival Study. A more direct approach to the ADP question is to examine actual capital cases and count the number of potential jurors who have been excluded on the basis of

identification with either the WE or ADP groups. Such a direct study was conducted in Arkansas for all capital cases since 1973, which includes the years in which the present Arkansas death-penalty statute has been in effect. Researchers identified a total of forty-six capital murder cases in which the \textit{voir dire} was transcribed. Of the 2142 jurors involved in these cases, 284 (13.25\%) were excluded as WEs while only 11 (0.51\%) were excluded as ADPs.

It is possible, of course, that defense attorneys simply failed to capitalize on their privilege to ask questions during the \textit{voir dire} which might lead to the elimination of ADPs. Accordingly, the Arkansas Archival Study also examined the distribution of excluded WEs and ADPs in only those cases in which ADP questions were, in fact, asked. The total number of jurors involved in \textit{voir dires} meeting this more stringent criterion was 1557 with the eleven excluded ADPs constituting 0.70\% of this total. In terms of actual Arkansas practice, therefore, ADPs comprise less than one percent of the excluded juror population, a result which accords with all other attempts to estimate the size of the ADP group, with the single exception of the Shure study.

The discrepancy between the estimate of the ADP population provided by the Shure results and the estimates provided by the remaining studies is presently unexplained. One possibility that reasonably could have contributed to the discrepancy has already been outlined: Shure’s use of a specific case to identify ADPs.

It is also possible that one or more of the samples were not representative. Since several studies suggested a small ADP population, including the well-conducted national survey by Harris, the probability is that if any study used a nonrepresentative sample, it was the Shure study. This would seem especially probable given that the West Los Angeles area sampled by Shure included such nonrepresentative communities as Bel Air, Beverly Hills, and Westwood.

At least, the size of the ADP group appears smaller than the size of the WE group. A conservative measure of the difference in group size may be extrapolated from the empirical data compiled by the Arkansas Archival Study in which ADP questions were actually asked. The size of the WE percentage (10.7) was fifteen times as large as the size of the ADP percentage (0.7). With a jury of twelve, the probability of selecting at least one WE is 0.74 and the
probability of selecting at least one ADP is .028.64 In Arkansas, therefore, between 1973 and 1981, the conservative chance of encountering a WE in a jury of twelve was twenty-seven times more likely than the chance of encountering an ADP. It therefore seems that the apparent neutrality afforded a defendant by the disqualification of potential jurors on the basis of either their adamant support of, or adamant opposition to, the death penalty, is only an appearance of neutrality. In fact, the defendant is disadvantaged.

THE FIRESIDE REVISITED

When the fireside induction was originally discussed, vigorous opposition from prosecutors suggested the possibility of a countervailing induction. In this context, there is one observation concerning the ADP issue which should not go unnoticed. The entire thrust of the state’s ADP argument is based on the assumption that ADPs are conviction-prone and WEs are acquittal-prone. This means that the state’s ADP argument affirms the fireside induction of a systematic relationship between death-penalty attitude and willingness to convict. This development is not altogether surprising since, on other occasions, the prosecution’s acknowledgment of the induction is even more explicit. For example, in People v. Ray,65 the prosecution argued that the inclusion of jurors who opposed the death penalty would lead to an acquittal-prone jury, i.e., the fireside induction was used to support the prosecution’s argument for the continued exclusion of WEs. A similar argument was presented in Spinkellink v. Wainwright.66 Another justification for attributing a belief in the fireside induction to the prosecution is the not uncommon prosecutorial strategy of seeking a more favorable “death-qualified” jury by eliminating WEs and, following conviction at the guilt-phase, of waiving the death penalty. For example, this pattern of death-qualification followed by post-conviction waiver of the death penalty occurred in the Grigsby case and raises the question whether the waiver was motivated by a post-conviction impulse of compassion or, as seems more likely, reflected a strategy designed to increase the probability of conviction at the guilt phase.

It thus appears that the state is aligned with the defense in en-

64. The probability value of .74 was obtained from \(1 - (100 - 10.7)^{12}\). The value of .028 was obtained from \(1 - (100 - .7)^{12}\).


66. 578 F.2d 582, 596 (5th Cir. 1978).
dorsing the fireside induction! An alternative, but simultaneous, prosecution strategy used both in *Hovey* and in *Grigsby* is that of denying the fireside induction by challenging the validity of the research studies. In effect, this denial asserts the countervailing induction that death penalty attitude is *not* related to a juror's willingness to convict.

The challenge involves a recitation of the methodological problems associated with each study. The primary criticisms are those outlined earlier,\(^6\)\(^7\), e.g., the use of college students' responses to written case materials. Legitimate criticisms can be directed toward any one study, particularly the pre-*Witherspoon* studies of Wilson and Goldberg. Criticisms of the "package" are somewhat strained since the studies consistently report the same effects using (1) different cases in different locales; (2) college students (Wilson; Goldberg), approximations to real jurors (Jurow), jury-eligible subjects (Ellsworth), or real jurors (Zeisel); (3) written cases (Wilson; Goldberg), audiotapes (Jurow), videotapes (Ellsworth), or actual jurors (Zeisel); (4) subjects with broadly defined scruples (Wilson; Goldberg; Zeisel) or precisely defined *Witherspoon* criteria (Jurow; Harris; Ellsworth); and (5) non-deliberating jurors (all studies) or jury deliberations (Zeisel; Ellsworth).

The state's challenge to the research has been directed primar-
rily at the external validity (the generalizability to actual jury situations), and not at the internal validity (that the differences in conviction behavior derive from differences in death-penalty attitudes) of the results. The question of external validity can never be answered to the critic's satisfaction, of course, since the ideal experiment cannot be performed, and the critics decline to offer any satisfactory, alternative experimental design. In the absence of exacting, impossible standards of scientific proof, most lawmakers will recognize that the Zeisel study and the three post-Witherspoon experiments obviously constitute more than a simple replication of the earlier results—they significantly extend external validity. In the absence of any serious challenge to internal validity, the studies provide convincing support for the fireside induction.

It is nonetheless interesting to explore the implications of the countervailing induction to determine the consequences to a defendant if there are no differences in the willingness to convict of Witherspoon-qualified and WE jurors. Consider the decision-making process confronting a jury as illustrated in Table 9. The columns describe the actual, but unknown, facts concerning the guilt or innocence of the defendant. The left column describes an innocent defendant or a defendant whose guilt is reasonably doubted. The right column describes a defendant guilty beyond a reasonable doubt. The question for the jury is whether the defendant is guilty and should be convicted, or whether the defendant should be acquitted as innocent. The decision to convict or acquit is described by the rows in Table 9.

**TABLE 9**
Decision Matrix: Varieties of Errors and Correct Decisions

<table>
<thead>
<tr>
<th>Jury's Verdict</th>
<th>Actual Condition of Defendant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Innocent</td>
</tr>
<tr>
<td>Convict</td>
<td>Harsh Error</td>
</tr>
<tr>
<td>Acquit</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

A correct decision can be reached in two possible ways. A correct decision includes, as possibilities, either the conviction of a guilty defendant or the acquittal of an innocent defendant. Simi-

68. Supra note 15.
larly, there are two possible ways to reach an erroneous decision: the conviction of an innocent defendant (a harsh error) or the acquittal of a defendant who is guilty (a lenient error). The purpose of a jury trial is realization of the goal of the criminal justice system, *i.e.*, reaching a correct decision and avoiding both errors. The risk of either type of error depends on the criterion the juror uses to determine what constitutes reasonable doubt. The dilemma is that any change in the juror's criterion which minimizes harsh error will lead to an increase in lenient error and vice-versa. In adopting a criterion for decision, the costs associated with each type of error must be considered. The costs associated with the two types of error are not equivalent—a harsh error is much more serious.

A comparison of the effects on the different error rates, given juries which include WEs under the assumption of either the fireside or the countervailing induction, should be instructive. First assume the fireside induction. If WEs are more likely to acquit in equivocal cases, then the effect of including them in capital cases will be that (1) fewer guilty defendants will correctly be convicted, *i.e.*, an increase in lenient error can be anticipated, and (2) more innocent defendants will correctly be acquitted, *i.e.*, a decrease in harsh error can be anticipated. Under the fireside induction, the inclusion of currently excluded WEs would decrease harsh errors. Under the presumption of innocence, it is precisely toward minimizing harsh error that the principles of the Constitution are directed.

Under the countervailing induction, if there are no differences between WEs and *Witherspoon*-qualified jurors in their willingness to convict, then the inclusion of WEs will have absolutely no effect on either correct decision, or harsh error or lenient error. If there were really no differences, then the procedure of asking about jurors' attitudes toward the death penalty during the *voir dire* makes no sense, because it would be of no significance in the outcome of the guilt phase of a trial. Under the countervailing induction, the state would have no interest in excluding jurors on the basis of death-qualification except for the inconvenience and/or cost associated with the inclusion of WEs at the guilt phase of a trial. One potential solution would be to require bifurcated guilt-sentencing proceedings using independent juries. The increased inconvenience and cost resulting from such a solution might be used to argue that the state's interest in saving time and/or money outweighs the defendant's interest in an impartial trial. In the current death-qualification case of *Grigsby v. Mabry*, however, this argument was not
introduced. Indeed, no evidence was introduced regarding the value of death qualification to the state. It therefore seems unlikely that the state genuinely believes that attitude toward the death penalty and tendency to convict are independent. This revisit to the fireside suggests that the state has adopted the inconsistent posture of affirming the fireside induction with the ADP issue and simultaneously denying the induction by challenging the validity of the research studies. The goal, continued exclusion of WEs from jury trials, is the thread of consistency which unites the state's seemingly inconsistent positions.

CONCLUDING COMMENTS

It has been argued that the effect of the exclusion of jurors under the *Witherspoon* standard ranges from no effect under the countervailing induction to a clear disadvantage to the defendant under the fireside induction. It has also been argued that the weight of the evidence indicates that potential jurors excludable under *Witherspoon* differ in their tendency to convict from those who could participate as jurors under *Witherspoon*. Most observers would probably agree that additional research would not be likely to result in serious modification of the basic experimental findings. There would be some interest in discovering limiting conditions to the observed differences, but the major contemporary interest has shifted to research designed to understand potential theoretical mechanisms responsible for the observed effects of death penalty attitudes.

One possible mechanism was suggested by Jurow. Jurow theorized that the distinction between conviction-prone and acquittal-prone jurors corresponded to Packer's distinction between a "due process model" and a "crime-control model" of the criminal process.69 In the "due process model,"

[T]he defendant is considered guilty if, and only if, the state meets proper procedural requirements from the moment it commences the criminal process. Thus, under certain circumstances, a defendant cannot be found guilty no matter how clear it may be that he committed the crime if, for example, illegally obtained evidence is used at his trial, or if he confessed without first being advised of his right to remain silent.70

The "crime-control model," in contrast, "focuses on apprehending

---

and convicting the maximum number of criminals. The emphasis is on 'factual guilt' in the sense that due process procedural requirements are viewed as 'niceties' or 'obstacles' that often hamper law enforcement personnel in arresting and convicting the 'obviously guilty.'”71

In addition to the studies reviewed here, several other studies, including numerous attitude surveys, were introduced in Grigsby v. Mabry.72 Two of these are briefly described because they represent the manner in which WEs and Witherspoon-qualified subjects establish criteria for making decisions relative to Table 9.

*The Ellsworth and Fitzgerald Study.*73 Ellsworth and Fitzgerald surveyed 123 WEs (actually guilt-phase includable) and 594 Witherspoon-qualified subjects, randomly selected from jury-eligible subjects in Alameda County, California. Table 10 reports each statement that significantly distinguished the two groups. Perhaps the statement most obviously related to the decision-making criterion was that it is better for society to let some guilty people go free than to risk convicting an innocent person. Over 62% of guilt-phase inculables agreed while a majority (56.1%) of the Witherspoon-qualified subjects disagreed. The probability of this difference resulting by chance is less than one in a thousand. The pattern of response to the other statements show a similar prosecution, crime-control bias.

*The Ellsworth Study.*74 In another study by Ellsworth, fourteen guilt-phase inculables and fifteen Witherspoon-qualified subjects were asked to imagine that they had served on a jury in a homicide case, reached a unanimous verdict, and later discovered the verdict was wrong. Each subject was asked to rate on a scale from 1 to 100 the amount of regret they would feel for either a harsh or a lenient error. The results indicated that Witherspoon-qualified subjects expressed equal regret for both types of error, while the guilt-phase inculables expressed significantly more regret for harsh than for lenient errors.

---

71. Id.

72. 483 F. Supp. 1372 (E.D. Ark.), modified, reh'g granted, 637 F.2d 525 (8th Cir. 1980).

73. Ellsworth & Fitzgerald, Due Process vs. Crime Control: The Impact of Death Qualification on Jury Attitudes (unpublished manuscript, 1979) (to be published in LAW & HUMAN BEHAV. (Summer, 1982)).

74. Hovey v. Superior Court, 28 Cal. 3d at 40, 616 P.2d at 1325, 168 Cal. Rptr. 152 (unpublished study, discussed in Hovey as the Ellsworth Post-Deliberation Follow-up Study).


<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage agreeing with statement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Guilt-Phase Includable</td>
</tr>
<tr>
<td>It is better for society to let some guilty people go free than to risk convicting an innocent person.</td>
<td>62.5</td>
</tr>
<tr>
<td>A person on trial who doesn't take the witness stand and deny the crime is probably guilty.</td>
<td>23.5</td>
</tr>
<tr>
<td>Even the worst criminal should be considered for mercy.</td>
<td>77.8</td>
</tr>
<tr>
<td>District attorneys have to be watched carefully, since they will use any means they can to get convictions.</td>
<td>53.1</td>
</tr>
<tr>
<td>All laws should be strictly enforced, no matter what the results.</td>
<td>46.3</td>
</tr>
<tr>
<td>A person would not be brought to trial unless he or she were guilty of a crime.</td>
<td>26.5</td>
</tr>
<tr>
<td>If the police obtain evidence illegally it should not be permitted in court, even if it would help convict a guilty person.</td>
<td>63.9</td>
</tr>
<tr>
<td>The plea of insanity is a loophole allowing too many guilty people to go free.</td>
<td>59.2</td>
</tr>
<tr>
<td>Harsher treatment of criminals is not the solution to the crime problem.</td>
<td>80.0</td>
</tr>
<tr>
<td>Defense attorneys have to be watched carefully, since they will use any means to get their clients off.</td>
<td>64.7</td>
</tr>
</tbody>
</table>
Results such as those described in these two studies suggest that the source of the differences in willingness to convict between WEs and *Witherspoon*-qualified subjects may derive from their respective decision-making criteria and their respective estimates of the costs of various types of error.

When the other survey evidence is considered, the clear thrust of the evidence has been to establish that persons who favor the death penalty are "uncommonly" predisposed to find for the prosecution and against the defendant.

All of the studies and all of the surveys display a common element: they demonstrate that there is a significant difference between those potential jurors excludable under *Witherspoon* and those jurors who could be "death-qualified" under *Witherspoon*. That is, all of the studies uniformly support the fireside induction. Under these conditions, Meehl noted that:

> The legislator's, judge's, or administrator's situation is most comfortable when there is a sizable and consistent body of research, experimental and nonexperimental (file data and field observation data), yielding about the same results as the fireside inductions. While one may be scientifically skeptical even in this harmonious situation, in the pragmatic context of decision making, rule writing, or policy adopting, such rigorous skepticism can hardly lead to pragmatic vacillation. Some sort of action is required, and all we have goes in the same direction.\(^7\)

The appropriate action would seem to require the conclusion that death-qualified juries are not only "uncommonly," but also unconstitutionally, prone to convict. A constitutional decision will, of course, have to meet the constitutional standard, which requires that the WE group constitute a distinctive, identifiable group.\(^6\) It is obvious, however, that the WE group is distinctive and identifiable, since members of this group are currently excluded on the basis of their distinctive and identifiable attitudes toward the death penalty. The convergent conclusions from the fireside induction and from social science research is that the exclusion is not irrelevant, but basic to fair and impartial judgments of guilt or innocence.

The appearance of a neutral jury is achieved by eliminating both the ADP group and the WE group. In reality, jury neutrality is not achieved, however, and the inadequacy is more than a matter of

---

75. Meehl, *supra* note 1, at 27.

the unequal distribution of ADPs and WEs. Haney\textsuperscript{77} has discussed the range of possible solutions to the multiple problems introduced by death qualification and has concluded that even modified forms of death qualification would violate the constitutional requirements of a fair and impartial trial. Haney's thoughtful analysis suggests that the most reasonable solution is a bifurcated trial with death-penalty attitude forming a basis for exclusion only at the penalty phase. In many jurisdictions, bifurcated trials already occur in civil cases with one jury assessing liability and another deciding damages. This practice is especially likely to be observed as the importance of the case increases, although no such case even begins to approach the life or death seriousness of capital cases involving death qualification.

Bifurcation may be in the state's, as well as in the accused's, interest. Death-penalty trials will not always occur in capital cases; for example, in Grigsby's case, the death penalty was waived. In such cases, the expensive death-qualifying \textit{voir dire} which now precedes the guilt phase of capital trials can be eliminated, saving substantial amounts of time and money.

An editor of this journal suggested to the author that a psychologist could change a light bulb only if the light bulb \textit{wanted} to change. The fireside induction and the social science research indicate that the practice of death qualification, like the light bulb, is in need of change. The real question is whether the legal system, like the psychologist, can afford to wait.

\textsuperscript{77} Haney, \textit{Juries and the Death Penalty: Readdressing the Witherspoon Question}, 26 \textit{Crime \& Delinq.} 512 (1980).