Law and Human Behavior, Vol. 8, Nos. 1/2, 1984

The Effects of Death Qualification on Jurors' Predisposition to Convict and on the Quality of Deliberation*

Claudia L. Cowan,† William C. Thompson,‡ and Phoebe C. Ellsworth§

This study provides a straightforward test of the proposition that people who are permitted to serve on juries in capital cases (death-qualified jurors) are more likely to convict a defendant than are people who are excluded from serving on capital juries due to their unwillingness to impose the death penalty (excludable jurors). A sample of 288 subjects classified as death-qualified or excludable under the *Witherspoon* standard watched a 2½-hour videotape of a simulated homicide trial including the judge's instructions, and gave an initial verdict. Death-qualified subjects were significantly more likely than excludable subjects to vote guilty, both on the initial ballot and after an hour's deliberation in 12-person juries. Nine juries were composed entirely of death-qualified subjects (death-qualified juries), while 10 contained from 2 to 4 excludable subjects (mixed juries). On postdeliberation measures, with initial death-penalty attitudes controlled, subjects who had served on the mixed juries were generally more critical of the witnesses, less satisfied with their juries, and better able to remember the evidence than subjects from the death-qualified juries, suggesting that diversity may improve the vigor, thoroughness, and accuracy of the jury's deliberations.

INTRODUCTION

The jury is a uniquely democratic institution. Its membership is not restricted to those with special skills or knowledge. Instead, the wisdom of the jury is collec-

*This research was supported in part by the Veatch Program of the North Shore Unitarian Society and in part by the Abelard Foundation. We are grateful to John Kazubowski, executive officer of the Santa Clara County Superior Court, Gabriel Gorenstein, Hans Zeisel, and Philip G. Zimbardo for their help in various phases of the research. We are especially grateful to Reid Hastie for allowing us to use his videotape and for helping us in numerous other ways. Requests for reprints should be sent to the third author.

†Attorney at Law, 2002 South Circle View Drive, Suite D, Irvine, California 92715.
‡Program in Social Ecology, University of California-Irvine, Irvine, California 92717.
§Department of Psychology, Stanford University, Stanford, California 94305.

0147-7307/84/0600-0053$03.50/0

tive, emerging when a group of ordinary citizens of different backgrounds and viewpoints deliberate together to reach a decision that represents the common sense of the community. A jury is seen as a fair tribunal because it represents the coalescence of a great diversity of community attitudes. The preservation of this diversity is of central importance both to the defendant, who is entitled to the collective judgment of the community, and to the community itself, whose members are entitled to an equal say on matters so fundamental as criminal justice. The selection of a jury *venire* is therefore surrounded by safeguards to assure that the jury will represent all fair and impartial persons in the community.

The United States Supreme Court "repeatedly has held that meaningful community participation cannot be attained with the exclusion of minorities or other identifiable groups from jury service" (*Ballew v. Georgia,* 1978, p. 231–237).[1] The practice of "death-qualifying" capital juries by excluding those who are adamantly opposed to the death penalty is an unusual departure from this commitment to diversity. It is the sole legal rule that permits the categorical exclusion of a whole group of otherwise eligible citizens from jury service on the basis of their beliefs. The purpose of their exclusion is rooted in the traditional reliance on the same jury to determine both guilt and penalty in capital trials. This group, the "*Witherspoon* excludables," have indicated their inability to consider voting for death if the defendant were found guilty of a capital crime and are therefore barred from the determination of the penalty; for reasons of economy and convenience, they are also excluded from participation in the determination of guilt or innocence.

The effect of this exclusion is that there are two distinct types of jury that decide whether people accused of crimes are innocent or guilty. In almost all criminal cases the defendant is tried by a "mixed" jury, representing the full range of community opinion, and in a series of cases extending back over the last century, the Supreme Court has consistently acted to maintain and enlarge this representativeness (cf. *Duren v. Missouri,* 1979).[2] But in capital cases, the defendant is tried by a death-qualified jury, representing a more limited spectrum of community opinion.

Legal scholars and social scientists have long been concerned with the consequences of this anomalous exclusion in capital trials. One effect, the unusually punitive nature of a jury culled of all members who express any opposition to capital punishment, was recognized by the Court in *Witherspoon v. Illinois* (1968)[3]; their conclusion that such a sweeping exclusion resulted in "a tribunal organized to return a verdict of death" (p. 521) led them to limit the excluded group to those who stated that they could never vote to impose the death penalty, regardless of the evidence.[4]

---

[1]*Ballew vs. Georgia,* 435 U.S. 223 (1978).

[2]*Duren vs. Missouri,* 439 U.S. 357 (1979).

[3]*Witherspoon vs. Illinois,* 391 U.S. 510 (1968).

[4]Witherspoon contended that his jury was unconstitutionally biased toward (1) conviction, and (2) death, because all individuals who had conscientious scruples against the death penalty had been excluded. The court held that the jury was biased toward death, but that the evidence was too shaky to persuade them that it was biased toward conviction. It restricted the group who could be

A second effect is the loss of representativeness. In one sense, the departure from representativeness is true by definition: juries from which a significant and identifiable portion of the population of eligible jurors is irrevocably banned cannot possibly represent the whole population. But there are other, less obvious concomitants to the general loss of representativeness. Women and blacks, whose rights to proportional representation on juries have been specifically guaranteed by the Supreme Court (*Taylor v. Louisiana,* 1975[5]; *Norris v. Alabama,* 1935[6]), are disproportionately excluded from capital juries under *Witherspoon,* and so are the poor, and members of certain religions (see Fitzgerald and Ellsworth, this issue). And, because attitudes toward the death penalty are related to other attitudes about the criminal justice system, the meaning of due process protections, and the proper performance of a juror's role, death qualification reduces the chances that certain beliefs and viewpoints will be represented on the jury (Fitzgerald and Ellsworth, this issue). This limitation of the diversity of the opinions and ideas may adversely affect the quality of the jury's deliberation, a point to which we shall return.

This restriction of the range of attitudes is also related to a third possible effect of death qualification: conviction proneness. The suggestion that death-qualified juries are conviction prone was initially articulated by Oberer in 1963:

> My premise is admittedly argumentative: that a jury qualified on the death penalty is one more apt to *convict,* quite apart from the degree of punishment to be assessed. If this premise once be conceded, it follows that a defendant tried before such a jury is denied a fair trial *on the basic issue of guilt or innocence.* (Oberer, 1963, p. 545).

Death Qualification and the Jury's Verdict: Conviction Proneness

In attempting to test the validity of Oberer's premise, social scientists have generally taken two different approaches. The first is to demonstrate that attitudes toward the death penalty are correlated with other attitudes more closely related to the juror's perceptions of the guilt or innocence of criminal defendants. This research has consistently found that those who favor the death penalty are likely to favor the prosecution, relative to those who oppose the death penalty. These studies are reviewed in Fitzgerald and Ellsworth (this issue), whose own work shows that this proprosecution bias is also characteristic of death-qualified jurors, as compared to the group excluded under *Witherspoon.* The second approach is to attempt to demonstrate the conviction proneness of death-qualified jurors more directly, by asking subjects their verdicts in real or simulated cases. In the fol-

---

constitutionally excluded to those who either (1) could not consider voting for capital punishment in any case, or (2) could not make an impartial decision as to a capital defendant's guilt or innocence. These two classes of jurors are barred from both the determination of penalty and the determination of guilt and innocence. The research reported here involves the effects of excluding the *first* group from the guilt-and-innocence decision. The second group (the "nullifiers") were excluded from the sample, as jurors who could not be fair and impartial on guilt in a capital case would be excluded from that decision in any case.

[5]*Taylor vs. Louisiana,* 419 U.S. 522 (1975).

[6]*Norris vs. Alabama,* 294 U.S. 587 (1935).

lowing review of the literature, we shall restrict ourselves to studies embodying this second approach.

Wilson

W. C. Wilson (1964) conducted the first experimental study of the conviction proneness of death-qualified jurors. His subjects, 187 college students, were divided into two groups on the basis of their answer to the question, "Do you have conscientious scruples against the death penalty, or capital punishment for a crime?" Wilson gave each subject written summaries of five capital cases and asked for a verdict of guilt or innocence. Subjects without scruples against the death penalty voted to convict significantly more often than subjects with conscientious scruples.

Wilson's "conscientious scruples" question was the appropriate legal standard for excluding jurors from capital cases until the 1968 *Witherspoon* decision. Naturally, it results in the exclusion of a larger (but possibly less distinctive) group of jurors opposed to capital punishment than does the current standard of an "automatic" vote against the death penalty. The underlying question addressed remains the same in both cases: Are attitudes toward capital punishment related to the tendency to vote guilty? But because the standard of exclusion has changed, and because Wilson failed to screen out jurors who would not be fair and impartial on guilt or innocence (the "nullifiers"), Wilson's study, taken alone, is of uncertain relevance for current practice.

Goldberg

A similar method was employed by Goldberg (1970). She presented 16 short written descriptions of assault and homicide cases to 100 black and 100 white Georgia college students. All the cases were possible capital cases at the time of the study. After reading each case, the subjects were asked to record their individual verdicts and to select an appropriate penalty. Goldberg, like Wilson, asked subjects whether or not they had conscientious scruples against the use of the death penalty.

Those subjects with no conscientious scruples against the death penalty voted to convict in 75% of the cases, while those with scruples voted to convict in 69% ($p = .08$). Likewise, the scrupled subjects were less likely to convict the defendants of first degree murder ($p = .06$), and more likely to acquit them on the grounds of insanity ($p = .03$). Similar findings emerged with respect to penalty: scrupled subjects imposed the death penalty or life imprisonment significantly less often than nonscrupled subjects ($p < .01$).

Goldberg concluded that the exclusion of scrupled jurors would create an imbalance

> heavily weighted against the defendant, in the exclusion for lack of impartiality of only those who oppose the death penalty and the correlative assumption that others are impartial. It is valid to interpret the present findings to say that people who lack scruples against the death penalty have a greater tendency to disregard the plea of not guilty on the grounds of insanity and are apt to give a stiffer penalty. (Goldberg, 1970, p. 63)

Like the Wilson study, Goldberg's research suffers from a scruples question that is no longer appropriate in the light of *Witherspoon*, and from a failure to identify and exclude nullifiers; thus the applicability of her conclusions to post-*Witherspoon* conditions is also dependent on the replication of her results with an appropriate definition of the excluded group. Also, both Wilson's and Goldberg's stimulus materials, brief written descriptions of crimes, were unusually lacking in mundane realism (Carlsmith, Aronson, and Ellsworth, 1976).

Zeisel

The same relationship between attitudes toward the death penalty and conviction proneness was reported by Zeisel (1968), using a radically different method of research. Zeisel interviewed actual jurors who had been sitting on felony juries in Chicago and New York. These jurors were asked three questions: (1) What was the first ballot vote of your jury as a whole? (2) What was your own first ballot vote? (3) Do you have conscientious scruples against the death penalty? The sample comprised 464 first-ballot votes of these jurors. Since the jurors had been exposed to so many different types of trial, Zeisel roughly controlled for the weight of the evidence against the defendant by dividing his data into 11 different "constellations" of guilty/not guilty splits on the first ballot, ranging from one guilty vote and 11 not guilty votes to 11 guilty and one not guilty. In 10 of the 11 first-ballot constellations, jurors without scruples against the death penalty voted guilty more often than jurors with scruples. The study is particularly important, since it is the only one that was not a simulation, but measured actual juror verdicts; because its results correspond to those of the simulation studies it strengthens our confidence in their validity.

The Wilson, Goldberg, and Zeisel studies all compared the conviction proneness of groups with or without conscientious scruples against the death penalty. In 1968 the *Witherspoon* decision changed the standard to exclude from capital juries only those who would invariably refuse to impose the death penalty, regardless of the evidence. Results of two post-*Witherspoon* studies that used questions more closely approximating this standard show that the change did not eliminate the relative conviction proneness of death-qualified as compared to excludable jurors.

Jurow

Building on the earlier work, Jurow (1971) conducted a more sophisticated study of attitudes toward the death penalty and conviction proneness among 211 employees of the Sperry Rand corporation. All subjects listened to two tape recordings of simulated murder trials which included short opening statements, examinations of witnesses, closing arguments, and legal instructions. After each tape, subjects gave verdicts of guilty or not guilty. Jurow also obtained individual measures on two different scales of attitudes toward the death penalty. One five-point scale, Jurow's CPAQ(B), included an alternative approximating the appropriate *Witherspoon* categorization: "I could never vote for the death penalty

regardless of the facts and circumstances of the case." (Jurow's question does not identify the nullifiers and it is possible that the format of the question, coming right after a different five-point scale question on death penalty attitudes, may have resulted in some misclassification.[7])

When Jurow divided his subjects into five groups based on their responses to the CPAQ(B), a familiar pattern emerged. Subjects who more strongly favored the death penalty were more likely to convict. The difference was statistically significant for the first tape-recorded case, a robbery–murder. The pattern of results in the second case, a rape-murder, showed a slight trend in the same direction, but the difference was not significant.[8]

Thus Jurow's data provide some confirmation for the contention that, even under the more stringent *Witherspoon* standard, death qualification creates juries that are more likely to convict than juries drawn at random. However, although the wording of Jurow's attitude question approximates a *Witherspoon* criterion more closely than the questions used by his predecessors, this advantage is not reflected in the statistical analysis: Jurow's chi squares are based on all five categories of the CPAQ(B), but the more appropriate comparison is between the *Witherspoon* alternative and the other four. Thus in Case 1, 44.7% of the death-qualified respondents voted to convict, as compared with 33.3% of the excludable respondents. In Case 2, the difference is more extreme, with 60% of the death-qualified respondents and 42.9% of the excludables voting for conviction. While these percentages certainly suggest that death-qualified jurors are more conviction prone, the small size of the excludable group precludes a trustworthy chi square estimate. At the very least, the data indicate that the relationship between death penalty attitudes and conviction proneness does not disappear when the attitudes are defined by a *Witherspoon* criterion.

Harris[9]

The 1971 Harris poll reported in White (1973) and in *Hovey v. Superior Court* (1980)[10] also included an appropriate *Witherspoon* question:

> Suppose you were a juror in a murder trial and it was completely up to the jury to choose whether the penalty would be death or imprisonment, do you think there would be any situations 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?

---

[7]That this misclassification was probably trivial is indicated by the penalties assigned by subjects in Jurow's 14 real and fictitious possible death penalty cases. These cases included some of the most notorious murderers of recent history, such as Lee Harvey Oswald, Sirhan Sirhan, James Earl Ray, and Adolf Eichmann, yet on the average, *Witherspoon* excludable subjects voted for death in less than one (.33) of the 14 cases. It is also notable that those who said that they would *always* vote for death failed to do so in 2.4 of the 14 cases.

[8]Since the order of presentation of the two audiotaped trials was not counterbalanced, one cannot rule out the possibility that the predicted relationship would have occurred in the second case as well, had it been presented first. It is impossible to know what factors may have led to the different levels of significance for Jurow's two trials.

[9]Detailed methodological and statistical information on the whole Harris poll are not available, and our discussion must be qualified accordingly.

[10]*Hovey vs. Superior Court,* 616 Pac 2d 1301 (1980).

Harris surveyed a national probability sample of 2068 adults in an hour-long, in-person interview that included numerous attitudinal questions as well as a measure of conviction proneness. Some of the attitudinal results are summarized in Fitzgerald and Ellsworth (this issue). For the question on conviction-proneness, all the respondents were initially instructed on three basic legal principles: the burden of proof on the prosecution, the standard of reasonable doubt, and the rule that a defendant need not testify. They then read written descriptions of four criminal cases: a typewriter robbery, a manslaughter, an assault on a police officer, and an automobile larceny. After each case, the respondents were told the legal definition of the appropriate crime and asked whether, as jurors, they would find the defendant guilty or not guilty.

In all four cases, those jurors who said they could never vote for the death penalty found the defendant guilty less often than did the death-qualified jurors. This difference was significant for the first three cases ($p < .01$) and marginally significant for the larceny case ($p < .10$).

What is striking about these studies is their consistency; carried out over a fifteen-year span, using very different samples and methods, all have reached the same result.[11] People who oppose the death penalty are less likely to convict a criminal defendant. Taken together with the common finding that jurors' initial voting patterns are a powerful predictor of the final jury verdict (cf., e.g., Hastie, Penrod, and Pennington, in press), these empirical studies are strong evidence that death-qualified juries are different from the juries that try all other criminal cases. They are more prone to convict.

In our attempt to replicate this finding, we tried to improve on the best features of the previous studies, while avoiding their flaws. Thus we used an appropriate *Witherspoon* question and screened out those jurors who could not be fair and impartial, we selected a nonstudent population of eligible jurors, and we used a videotaped trial that was much more realistic than any of the materials used in previous simulations.

Death Qualification and the Process of Jury Deliberation: The Effects of Homogeneity

The steady accumulation of research indicating that death-qualified jurors are more likely to vote for conviction represents an obvious threat to the fairness

[11]Osser and Bernstein (1968) compared the conviction/acquittal rate for first-degree murder trials to that for robbery and burglary trials over a 6-month period in 1967 and found no difference, leading them to conclude that death-qualification does not produce conviction proneness. They provide no data on the proportion of death-qualified juries or jurors in each group, but simply assume that jurors in all the murder cases were death-qualified and that all juries in the robbery and burglary cases included some excludables. Even if this dubious assumption were correct, it would mean that in this study, unlike any of the other studies reported here, type of jury would be perfectly confounded with type of case. Burglary and robbery cases may require stronger evidence to get to trial, defense attorneys may defend people accused of murder more energetically, and there are numerous other differences that might mask the conviction proneness of death-qualified juries. To assume that the only difference between murder trials and robbery and burglary trials is the composition of the jury is obviously absurd.

of the tribunal, and is especially serious since it is "an imbalance to the detriment of . . . the defense" (*Ballew v. Georgia,* 1978, p. 236). Momentous though this consequence is, however, it may not be the only unfortunate effect of death qualification. The Supreme Court has repeatedly stated that proper jury functioning involves effective deliberation. Hastie, Penrod, and Pennington (in press), in an analysis of recent Supreme Court decisions on the function of the jury, note the Court's emphasis on the jury's ability to "remember each of the important pieces of evidence or argument" and to "make critical contributions necessary for the solution of a given problem" (*Ballew v. Georgia,* 1978, p. 233); the "counterbalancing of various biases" (*Ballew,* p. 234) in the course of "earnest and robust argument" (*Apodaca v. Oregon,* 1972, Justice Douglas, dissenting, p. 389)[12]; and the maintenance of the reasonable doubt standard by allowing those who favor acquittal "through full deliberation to temper the opposing faction's degree of certainty of guilt" (*Apodaca,* Justice Douglas, dissenting, p. 390). Death qualification may impair these essential jury functions.

Death penalty attitudes do not exist in a vacuum, but are associated with a whole cluster of attitudes, experiences, and knowledge (cf. Vidmar and Ellsworth, 1974; Fitzgerald and Ellsworth, this issue; Tyler and Weber, 1982). And the process of jury decision making is much more than a simple aggregation of guilty and not guilty votes; it is an exchange of these attitudes, experiences, and knowledge as they pertain to the case at hand. When a death-qualified jury retires to the deliberation room, it has lost more than a particular number of Not Guilty votes. It has lost a set of perspectives on the world and the evidence.

How might this restriction of the diversity of viewpoints affect the quality of the jury's deliberation? With the loss of a particular set of perspectives the jury may lose the only juror who remembers a crucial fact or point of law. Certain arguments, scenarios, or interpretations of the defendant's state of mind may never be raised during deliberation. For example, in their study of jurors' responses to a simulated homicide trial, Hastie, Penrod, and Pennington (in press) found that jurors from lower socio-economic backgrounds did not find the fact that the defendant was carrying a knife exceptional, and were less likely than middle-class jurors to infer premeditation from this fact. Finally, the restriction may hush the voice of a dissenting minority, whose criticisms of the majority viewpoint would have fostered the careful scrutiny of all relevant issues and the maintenance of the standard of proof beyond a reasonable doubt.

In making these predictions, we need not assume that those jurors excluded by death qualification are smarter, more convincing, or somehow better than the rest of the jury; they need only be different. In losing a different viewpoint the jury loses some of its capacity for controversy and self-criticism. Unchallenged opinions need not be carefully considered, and thus the exclusion of a group that is likely to hold different opinions may impair the accuracy of the jury's perception of the evidence, the thoroughness of their deliberation, and the stringency of their standard of proof.

The study reported here was designed to examine the effects of death qual-

[12]*Apodaca et al. vs. Oregon,* 406 U.S. 408 (1972).

ification on both aspects of jury functioning: initial verdict preference and the quality of the deliberation process. In this article, we shall present data on jurors' verdict preferences before and after deliberation, and on postdeliberation measures related to the quality of the deliberation. Analysis of the videotapes of the actual deliberation process will be reserved for a future publication.

METHOD

Overview

A group of adults qualified for jury service watched a videotape of a simulated homicide trial, which represented all major aspects of an actual criminal trial. After hearing the evidence, arguments, and instructions, the jurors gave an initial verdict. Jurors were then divided into 12-person juries and allowed to deliberate for one hour. Half the juries were composed exclusively of death-qualified jurors, subjects who had previously indicated that they were willing to consider voting to impose the death penalty (death-qualified juries). The remaining juries comprised a majority of death-qualified jurors, but also included two to four jurors who would be excluded under current law because of their views on capital punishment (mixed juries).[13] After deliberation on either a mixed or death-qualified jury, each subject completed a postdeliberation questionnaire tapping different aspects of the jury's functioning and rendered a postdeliberation verdict.

It is important to note that there are two very different types of questions that can be addressed by this design. First, at the level of individual subjects, we can ask how the *Witherspoon*-excludable jurors differ from the death-qualified jurors. In line with the prior research, we predicted that death-qualified jurors would be more likely to favor a guilty verdict before they entered the jury room than would excludable jurors. We also wanted to find out whether exposure to other points of view during the course of deliberation would erase any differences in conviction-proneness, or whether these differences would persist even after deliberation. We predicted that death-qualified jurors would perceive the trial evidence in a manner more favorable to the prosecution.

Second, we can ask how the restriction of diversity on a death-qualified jury affects the process of that jury's deliberation. This is a *jury* level analysis which compares the responses of subjects who had deliberated on a death-qualified jury to those of subjects who have been exposed to the full range of community views on a mixed jury. If our speculations on the effects of diversity are correct, we would expect the mixed juries to show higher levels of controversy, a more critical

[13]Of the 288 subjects who saw the videotape and recorded initial ballot preferences, it was only possible to assign 240 to juries, ten 12-person juries in each condition, since sometimes subjects would fail to keep their appointments, so we would be left with extra jurors after creating as many juries as we could. In addition, one death-qualified jury had to be dropped from the analysis, because one of its members was an amateur actor who had recently starred in a production of *Twelve Angry Men*, and who dominated deliberation using arguments and reasoning drawn from that play.

view of the evidence, and perhaps a better memory for the facts of the case and the judge's instructions.

Subjects

Two hundred and eighty-eight adults eligible for jury service in Santa Clara or San Mateo County, California participated in the study. Thirty-seven of them were recruited from the venire lists of Santa Clara County Superior Court after completion of their terms as jurors. Of the remaining 251 subjects, 218 were people who had responded to a classified advertisement in local newspapers asking for volunteers for a study of "how juries make decisions." The final 33 people were referred by friends who had heard of the study or who were subjects themselves. Each subject was paid $10 for participation.

Definition of Experimental Groups

This study was designed to compare the attitudes and behavior of death-qualified and excludable venirepersons who might serve as jurors in a capital trial. Juror eligibility and division into experimental groups was established in the initial telephone contact with potential subjects. Venire list subjects were contacted by the experimenter and invited to participate in the study. The remaining subjects were interviewed when they called in response to the newspaper advertisement.

In the telephone interview, two questions derived from the *Witherspoon* decision were used to identify the subject groups. The experimenters explained that judges in homicide trials ask questions about the jurors' attitudes, and subjects should answer just as they would if they were being questioned prior to a real trial. The experimenters then asked the first *Witherspoon* question:

> Now assume that you have been called as a possible juror in a first degree murder trial. The prosecutor is asking for the death sentence. Since this is a case where the death penalty may be imposed, the judge will ask you certain questions about your attitude toward the death penalty before deciding whether you should be chosen to serve on the jury.
>
> There are two parts to any trial where the death penalty may be imposed. In the first part, the jury decides whether the person on trial is guilty or not guilty. If the person is found guilty, there is a second part—a separate trial—in which the jury decides whether he or she should get the death penalty or life in prison.
>
> The judge will ask you this question:
>
> Is your attitude toward the death penalty such that as a juror you would never be willing to impose it in any case, no matter what the evidence was, or would you consider voting to impose it in at least some cases?
>
> a) I would be unwilling to vote to impose it in any case.
>
> b) I would consider voting to impose it in some cases.

All persons who gave response (a), stating unambiguously that they could not vote for the death penalty in any case, were designated the *Witherspoon*-excludable group. Those who gave response (b) became the death-qualified group. The second *Witherspoon* question followed:

> Now suppose that you were a juror in the *first* part of the trial, just to decide whether the accused person is guilty or not guilty of the crime. The judge instructs you

> that in reaching your verdict you are only allowed to consider the evidence presented in court, and must follow the law as he will state it to you. If the accused is found guilty, there will be a separate trial to decide whether or not he or she should get the death penalty.
>
> Which of the following expresses what you would do if you were a juror for the first part of the trial?
>
> a) I would follow the judge's instructions and decide the question of guilt or innocence in a fair and impartial manner based on the evidence and the law.
> or
> b) I would 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.

Those persons choosing alternative (b), indicating that they could not be impartial in determining the guilt or innocence of a defendant, would be disqualified from service on any jury for reasons distinct from their ability to consider imposing the death penalty. We used this question to screen out this incontestably ineligible group from participation in the study.

Two further groups of potential subjects were excluded: (1) those ineligible for jury service in California because they were under eighteen years old or had neither a California Driver's License nor voter registration; (2) lawyers and reporters.

After scheduling difficulties in composing twelve-person juries and imposition of these selection criteria, 258 death-qualified and 30 excludable subjects were included in the final sample.

The Trial Videotape

After viewing simulated trial materials prepared by several other social scientists interested in jury behavior and considering creating materials of our own, we decided to use the videotape prepared by Dr. Reid Hastie for use in his research on the jury deliberation process (Hastie, Penrod, and Pennington, in press). We found this tape to be representative of the procedures, setting, style, and issues that commonly occur in actual homicide trials. The case was complex enough to afford several plausible interpretations and verdict preferences. It resembled most real murder trials in that the fact that the defendant had killed the victim was not in controversy; rather, the evidence centered on the precise sequence of events preceding the killing, and on the defendant's state of mind at the time. Finally, the tape was far more vivid and realistic than any others we have encountered. We felt that it was highly unlikely that we could construct a better tape with our resources.

Hastie's videotape is a reenactment of an actual homicide case based on a complete transcript of the original trial. Each actor portraying a defense or prosecution witness was provided with "a summary of the case highlighting his or her testimony" (Hastie, Penrod, and Pennington, in press). The judge and the lawyers, portrayed by an actual judge and two experienced criminal attorneys, were given "unabridged copies of the judge's instructions, selections of relevant testimony, and the attorney's opening and closing arguments as they were originally presented" (Hastie, Penrod, and Pennington, in press). The attorneys were asked to develop their cases as they would for a real trial, the witnesses were

asked to review their materials to get their version of the events firmly in mind, and then, for the actual taping, all actors put aside their materials and engaged in "spontaneous improvisations closely following the original case." (Hastie, Penrod, and Pennington, in press).

We modified the tape in two ways for the present research. First, we shortened it slightly by deleting one defense witness whose testimony added little. Second, we replaced the segment of the original tape containing the judge's instructions, which had been based on Massachusetts law, with a new sequence in which the applicable California law was given. The new instructions were derived from *California Jury Instructions, Criminal* (CALJIC, 1970) and were assembled in consultation with trial attorneys and law professors. Professor William Keogh of the Stanford Law School portrayed the presiding judge in the new sequence. The version of the tape we used lasted two and one half hours, including a half hour of judge's instructions. Pretesting indicated that the tape was regarded as convincing and realistic.

In the trial videotape, the defendant Frank Johnson is charged with first degree murder for the stabbing of Alan Caldwell outside a neighborhood bar. The prosecution brings evidence that the defendant and victim had argued in the bar earlier that day, and that Caldwell had threatened the defendant with a straight razor. Johnson left, but returned with a friend that evening, carrying a fishing knife in his pocket. Caldwell later came into the bar, and he and the defendant went outside and began to argue loudly. Two witnesses testify that they saw Johnson stab down into Caldwell's body. The victim's razor was subsequently found folded in his left rear pocket. For the defense, Johnson testified that he had returned to the bar that evening on the invitation of his friend, and had entered only after ascertaining that Caldwell was not there. Caldwell had come in and asked him to step outside, he supposed for the purpose of patching up their quarrel. Once outside, Caldwell had hit him and come at him with a razor. He had pulled out a fishing knife that he often carried in his pocket and Caldwell had run onto the knife. Johnson's friend corroborated much of his testimony. In cross-examination the defense attorney casts doubt on the ability of the prosecution eyewitnesses to see the scuffle, and shows that medical evidence cannot establish whether the defendant stabbed down into the victim or the victim ran onto the knife.

Four verdicts are possible in this case, dependant upon the jury's findings of the facts. The defendant may be guilty of first degree murder, of second degree murder, of voluntary manslaughter, or he may be not guilty by reason of self-defense or accidental homicide.

Procedure

The study was conducted on weekend afternoons at Stanford University. Each subject group consisted of 12 to 36 subjects.

Upon their arrival, all subjects were shown to an auditorium, asked to fill out an informed consent form and a preliminary questionnaire, and given a brief overview of the study. The questionnaire focused on background (demographic)

characteristics, general attitudes toward the death penalty and toward criminal defendants, and general attitudes with respect to crime control and due process (The Legal Attitudes Questionnaire: Boehm, 1968). The experimenter then introduced the trial videotape:

> Now we would like to show you a videotaped reenactment of an actual criminal trial. This trial took place in Boston, and nothing has been changed, except, of course, for the names. You will be asked to reach a verdict based on the facts of the case and the law the judge explains to you, just as you would if you were an actual juror. We would like you to pay close attention to the testimony and the judge's instructions and to try to take the case as seriously as you would if you were actually serving on a jury. While you are listening to the case, you should not communicate with anyone else, and you should not take any notes, because we want your experience to be as much like that of a real juror as possible.

During the 2½ hours while the subjects were watching the videotape, the experimenters assigned the participants to twelve-person juries. The assignments were designed to produce two types of jury: death-qualified juries, composed totally of death-qualified jurors, and "mixed" juries, composed of death-qualified jurors and two to four *Witherspoon* excludables. If five or more excludables were present, they were assigned to two juries; otherwise all excludables were assigned to the same jury. Death-qualified jurors were then assigned to the remaining spaces on the juries.

All jury assignments were random subject to the following constraints: (1) juries could not contain members who were acquainted, (2) juries could not contain more than one student member, and (3) numbers of male jury members and of persons recruited from the jury list were roughly equalized across juries.

As soon as the videotape was over, the experimenter reminded the subjects that they should not converse, distributed a verdict questionnaire, and continued as follows:

> Now you've seen the whole case and heard the judge explain the law. If you were sitting on a jury, and had to vote right now, what would your vote be? Take a few moments to consider, and mark your answer on the sheet that was handed out to you. This should be your own personal individual decision, so please keep your feelings to yourself and don't look to other people to see what they are thinking. Choose one of the four verdicts on the sheet: first degree murder, second degree murder, manslaughter, or not guilty.

Subjects' responses to this question were the basic measure of predisposition to convict.

After collecting the initial verdict questionnaires, the experimenter read off the jury assignments and directed the members of each jury to a separate room for deliberation. The rooms were seminar rooms, each containing a table with ample space for 5 people along each side and one at each end. A videotape camera was placed behind the person sitting at one end of the table, and two ceiling microphones were used to record the audio track. The camera and microphones recorded the deliberations for later analysis, and also allowed the experimenters to view the deliberations on a monitor outside the deliberation room, in order to detect problems that might jeopardize the validity of the study.

Once the subjects were settled in the jury room, the experimenter told them that their next task was to discuss the case and try to reach a verdict. S/he told them that their immediate postvideotape verdict was confidential, and that they need not feel committed to it. S/he informed them that most juries begin by taking a straw vote, and that in any case they should choose a foreman before beginning their deliberation. S/he continued as follows:

> As you discuss the case, it is important to put yourselves into the role of jurors. Imagine that you are a real jury and that your verdict will actually determine the fate of the defendant you saw on the tape. We want you to make your decision only on the basis of what you saw on the tape. Although the characters in the trial you saw were actors, we want you to treat them as if they were real. In short, we want you to make the decision you would make if you were a real jury and if you had seen in court exactly what you saw on the tape.

The experimenter closed by informing the subjects that they had one hour in which to deliberate, and that they should try to reach a decision in that time, although s/he realized that one hour might not be long enough to reach a consensus. The purpose of this instruction was simply to assure that the subjects worked on their deliberation seriously and tried to reconcile their differences of opinion; in fact we did not expect them to reach a verdict, nor did we ask that a final vote be taken.

Subjects were then left to discuss the case. Although they appeared to be slightly self-conscious in the presence of the recording equipment for the first minute or two, as soon as the deliberations revealed disagreements among the members (almost immediately in all juries), they became highly involved in the discussion and seemed to forget about the camera. After an hour the experimenter returned, stopped the deliberation, and handed out the postexperimental questionnaires.

Postdeliberation Questionnaire

The six-part postdeliberation questionnaire measured individual jurors' reactions to the case and deliberations. The questionnaire was designed both to compare the responses of death-qualified and excludable jurors, and to compare the responses of subjects who had served on death-qualified or mixed juries. The measures were as follows, and always appeared in the following order:

1. A second verdict measure, designed to assess opinion change;
2. A 17-item questionnaire designed to assess subjects' feelings about the performance of their jury—its heterogeneity, fairness, thoroughness, etc.;
3. A 12-item questionnaire designed to assess subjects' perceptions of the believability and helpfulness of each of the six witnesses;
4. A 6-item questionnaire designed to assess subjects' evaluations of the prosecutor and the defense attorney;
5. An 18-item true–false test for recall of the judge's instructions, balanced for acquiescence and errors favoring defense or prosecution;
6. A 14-item multiple choice test of memory for testimony. Each item included one correct response choice, one erroneously slanted toward the

defense, one erroneously slanted toward the prosecution, and one "neutral" error.

Sections 2, 3, and 4 required responses on a 7-point Likert-type scales. Most subjects completed this battery of questionnaires within half an hour.

RESULTS

Sample Characteristics

General Characteristics

The subject sample was fairly representative of the suburban upper-middle-class community surrounding Stanford University, except that males and minorities were under-represented. The sample was 93% white, comprising 35.4% males and 64.6% females. The average age of subjects was 43, and 64.4% of the sample was currently employed outside the home. Married persons were 46% of the sample; 25% were single, 19% divorced, 4% separated, and 6% widowed. The median educational level was slightly less than a baccalaureate degree. Registered Democrats were 45.3% of the sample, Republicans 33.5% and unregistered 6.8%, with the remainder divided among Independent voters and small parties. In the category of religious affiliation, 35% listed themselves as Protestant, 17% as Catholic, 9% as Jewish, 25% listed no affiliation, and 14% listed other religions. Finally, 45% of the sample had previously done jury duty, while 36% had actually served on juries.

Death-Qualified vs. Excludable Subjects

Witherspoon-excludable jurors did not differ significantly from death-qualified jurors in any of these characteristics with two notable exceptions. Excludables were more likely to be female ($\chi^2_{(1)} = 5.7$, $p < .02$) and Catholic ($\chi^2_{(4)} = 9.3$, $p < .06$). Neither did excludable and death-qualified jurors differ on a crude measure of their standard of reasonable doubt. When asked, "What degree of probability that the defendant committed the crime would justify a verdict of guilty?" the responses of excludable ($\bar{x} = 86.4\%$) and death-qualified jurors ($\bar{x} = 86.2\%$) were nearly identical ($t = .03$). There were slight differences in the jurors' *a priori* estimates of felony defendants' guilt. Of the people brought to trial on felony charges, excludables estimated that 75.3% are "actually guilty," while death-qualified jurors estimated 80% ($t = 1.74$, $p = .09$). Naturally, excludable and death-qualified jurors differed radically in their scaled attitudes toward capital punishment ($\bar{x}_{EX} = 1.10$, $\bar{x}_{DQ} = 3.12$ on a 5-point scale; $t = 10.28$, $p < .001$).

Conviction Proneness

Death Qualification and Conviction Proneness

The distribution of initial verdict preferences is presented in Table 1. Our results provide strong support for the hypothesis that death-qualified jurors are

Table 1. Verdict Choices of Death-Qualified and Excludable Jurors

	Predeliberation ballot	
	Death-qualified	Excludable
First degree murder	7.8% (20)[a]	3.3% (1)
Second degree murder	21.3% (55)	23.3% (7)
Manslaughter	48.9% (126)	26.7% (8)
Not guilty	22.1% (57)	46.7% (14)
	100% (258)	100% (30)

$\chi^2_{(3)} = 10.2, p < .02$
$\chi^2_{(1)}$ for guilty/not guilty split $= 7.46, p < .01$

	Postdeliberation ballot[b]	
	Death-qualified	Excludable
First degree murder	1% (2)	3.4% (1)
Second degree murder	17.3% (34)	13.8% (4)
Manslaughter	68% (134)	48.3% (14)
Not guilty	13.7% (27)	34.5% (10)
	100% (197)	100% (29)

$\chi^2_{(3)} = 10.3, p < .02$
$\chi^2_{(1)}$ for guilty/not guilty split $= 7.79, p < .01$

[a]Numbers in parentheses are cell frequencies.
[b]Two of the jurors who deliberated are not included in the postdeliberation ballot data because one wrote "Can't Decide" on the ballot and the other left it blank.

more likely to convict than are jurors excludable under the *Witherspoon* criteria. The most direct test of this hypothesis is a comparison of Guilty and Not Guilty verdicts among the two groups of jurors. Among the death-qualified jurors 22.1% voted Not Guilty while 77.9% found the defendant guilty of some level of homicide. Among the excludable jurors, 46.7% voted Not Guilty, and 53.3% voted Guilty of some offense. This difference is highly significant ($\chi^2_{(1)} = 7.46, p < .01$) and indicates that the departure from representativeness created by the process of restricting juries in capital cases to death-qualified jurors creates a bias against defendants in death penalty trials.

It might be argued that initial verdict preferences are a questionable measure, since they do not reflect the impact of exposure to other points of view. Previous research on the high correlation between first-ballot verdicts and final jury decision is strong evidence for the validity of initial verdict preferences as a measure of conviction proneness (cf. Hastie, Penrod, and Pennington, in press). Nonetheless, in order to find out whether the differences in initial verdict preferences were robust enough to withstand exposure to different opinions, we took a second measure of verdict preference following the hour-long deliberation. These results are also presented in Table 1. Among the death-qualified jurors, 13.7% still voted Not Guilty after an hour of deliberation, while 34.5% of the excludable jurors

voted Not Guilty ($\chi^2_{(1)} = 7.79, p < .01$). Thus it is clear that the excludable jurors do not abandon their position when faced with majority opposition, but maintain it long enough to contribute their arguments to the deliberation and the ultimate decision.

Other Variables and Conviction Proneness

In general, personal characteristics of the jurors showed little correlation with first ballot vote. Harsher verdicts (1 = not guilty, 4 = first degree murder) were not associated with: age ($r = .07$), sex ($r = .02$), level of education ($r = -.04$), previous experience on a jury panel ($r = -.04$), or the jurors' views of the level of subjective certainty that would justify a verdict of guilty ($r = .05$).

Those subjects who were not employed ($r = .09, p < .08$) and those who had actually served on a jury before ($r = -.10, p < .06$) were slightly more likely to have strict first ballot verdicts, as were those subjects who estimated that more felony defendants who come to trial are "actually guilty" ($r = .11, p < .05$).

The variables most successful in predicting first ballot votes were the two measures of legal attitudes. The best predictor of voting severity was a favorable attitude toward the death penalty ($r = .19, p = .001$). The "Legal Authoritarianism" score from Boehm's (1968) Legal Attitudes Questionnaire (LAQ) was also associated with more severe first ballot votes ($r = .16, p < .02$). These two attitudes were highly intercorrelated ($r = .46, p < .001$).

Differences Between Death-Qualified and Excludable Jurors on Postdeliberation Measures

Perceptions of Witness Credibility

In general, as might be expected from their verdicts, death-qualified jurors were more impressed with prosecution witnesses than were excludable jurors. Two measures of credibility were taken for each witness: how believable his testimony was and how helpful it was in understanding what happened on the night of the killing. The results are presented in Table 2.

In contrast, there were no significant differences in death-qualified and excludable jurors' ratings of defense witnesses. Both groups found the defense eyewitness and the defendant relatively noncredible. Death-qualified jurors' ratings of the believability of the defendants' companion and the defendant were 2.9 and 3.9, respectively, while excludables' ratings were 3.2 and 4.1.

Perceptions of Attorneys

Death-qualified and excludable jurors did not differ in their perceptions of the likability, competence, or believability of the defense attorney. Both groups found the prosecution attorney equally competent, but excludable jurors found him less believable ($\bar{x}_{EX} = 4.17, \bar{x}_{DQ} = 4.77; t = 2.01; p < .05$) and less likable ($\bar{x}_{EX} = 2.97; \bar{x}_{DQ} = 3.64; t = 2.37, p < .02$) than did death-qualified jurors.

Table 2. Death-Qualified and Excludable Jurors' Perceptions of the Believability and Helpfulness of Prosecution Witnesses

		Believability ratings[a]			
		Mean	*df*	*t*-value	Probability
Police eyewitness:	Death-qualified	5.54	223	4.38	.001
	Excludable	4.24			
Arresting officer:	Death-qualified	5.59	223	2.11	.04
	Excludable	5.00			
Forensic pathologist:	Death-qualified	6.55	225	4.51	.001
	Excludable	5.62			
Prosecution eyewitness:	Death-qualified	5.18	226	1.97	.05
	Excludable	4.65			
		Helpfulness ratings[a]			
		Mean	*df*	*t*-value	Probability
Police eyewitness:	Death-qualified	5.06	222	3.69	.001
	Excludable	3.86			
Arresting officer:	Death-qualified	3.94	223	1.14	NS
	Excludable	3.52			
Forensic pathologist:	Death-qualified	5.51	221	3.50	.001
	Excludable	4.31			
Prosecution eyewitness:	Death-qualified	4.28	221	1.80	.08
	Excludable	3.72			

[a]Ratings were made on a scale from 1 to 7, where 1 = not at all believable/helpful, and 7 = very believable/helpful.

Memory for Judge's Instructions

In general, jurors' memory for the trial judge's instructions was not good. Overall, jurors answered only 11.77 of 18 true–false questions correctly; random guessing would result in 9 right answers. Excludable jurors ($\bar{x}$ = 12.0) did not score significantly higher than death-qualified jurors ($\bar{x}$ = 11.73).

Though the questions were very difficult, it was clear from the videotapes of jury deliberation that many jurors failed to understand even such basic matters as the distinctions between possible verdicts, and that deliberation often failed to eliminate these misconceptions. One jury concluded that second degree murder was a premeditated killing without extreme hatred (their definition of malice aforethought). A second jury defined second degree murder as a killing in the heat of passion when acting with hatred of the victim. Two juries concluded that the category correctly defined as not guilty by reason of self-defense was actually "involuntary manslaughter," an incorrect categorization never mentioned in the judge's instructions. This generally poor understanding of legal instructions has also been found by other students of jury performance (cf. Hastie, Penrod, and Pennington, in press).

Memory for Evidence

In contrast, jurors' memory for the evidence presented in the trial was fairly good. Jurors were able to identify an average of 8.8 correct answers out of 14 questions with four alternatives (chance level would be 3.25 correct answers). *Witherspoon* excludables ($\bar{x}$ = 9.17) did not remember evidence significantly better than death-qualified jurors ($\bar{x}$ = 8.83). There were also no significant differences by juror attitudes in the rate of errors favoring defense or prosecution.

Assessments of Jury Deliberation

All ten questions on jurors' perceptions of the quality of their deliberations show that *Witherspoon* excludables were less satisfied with their jury's actions, though the difference reached a significance on only a few questions. Excludables were marginally less likely than death-qualified jurors to believe the deliberation "brought out facts" ($t = 1.67$, $p < .10$). They were less likely to believe the jury was "fair" ($t = 3.68$, $p < .001$), and thought the jury's consideration of the case was not as "serious" ($t = 1.95$, $p < .06$). This finding is in line with Hastie, Penrod, and Pennington's (in press) finding that jurors who are in a minority faction are less satisfied than jurors in the majority.

Death-Qualified vs. Mixed Juries

We have argued that the preservation of diversity of opinion on a jury may affect the deliberation process in and of itself, independently of the number of Not Guilty votes. A thorough examination of this question would involve a detailed analysis of the videotapes of the jury deliberations. For the time being, however, we can compare the postdeliberation questionnaire responses of those who served on the homogeneous death-qualified juries with those who served on juries that included some excludable jurors. (No juries reached a verdict in the allotted hour.) We can look at how the experience of serving on one or the other type of jury affects the jurors' views of the witnesses, satisfaction, memory for the evidence, and so on. Because we were interested in the effects of diversity *per se*, over and above the effects of initial verdict preference, all comparisons were performed using analysis of covariance controlling for predeliberation verdicts. In general the results of this analysis were identical to the results of the standard *t*-tests, and we will report only those that differ.[14]

[14]Since excludable jurors served only on mixed juries, person and treatment effects on the postdeliberation measures are confounded. The analysis of covariance controls for the most relevant difference between the two types of jurors—verdict preference, but does not completely solve the problem. Actually, there are two problems: (1) The diversity provided by juries may have no influence on the deliberations or the jurors' experiences; the average responses of subjects on mixed juries may differ from those of subjects on homogeneous juries solely because the divergent responses of the Excludables are averaged in. (2) The Excludable jurors may respond differently from the death-qualified jurors because their experience was exclusively on mixed juries. In order to find out whether the experience of serving on a homogeneous jury differed from the experience of serving on a mixed jury (question 1), and thus to assess the effects of diversity *per se*, we compared the

Perceptions of Witness Credibility and Attorneys

Excludable jurors, it will be recalled, were more critical of the prosecution and, if anything, slightly less critical of the defense witnesses than were death-qualified jurors. When ratings of jurors on death-qualified and mixed juries were compared, the same pattern emerged. Mixed juries rated the police eyewitness as less believable ($\bar{x}_{\text{mixed}} = 5.1$; $\bar{x}_{\text{DQ}} = 5.7$; $t = 2.84$, $p = .005$) and less helpful ($t = 1.96$, $p < .06$) than did death qualified juries. The forensic pathologist was also rated less believable ($\bar{x}_{\text{mixed}} = 6.2$; $\bar{x}_{\text{DQ}} = 6.7$; $t = 3.05$, $p < .005$) and helpful ($t = 3.36$, $p = .001$) by mixed juries. Mean ratings by mixed or death-qualified juries for the two remaining police witnesses were in the same order, but did not reach significance either by t-tests or covariance analysis. Mixed and death-qualified mean ratings of believability for the arresting officer were 5.4 and 5.6, respectively, and for the prosecution eyewitness they were 5.1 and 5.2.

Although there were no significant differences between death-qualified and mixed juries' ratings of defense witnesses when initial vote preference was controlled by covariance analysis, the pattern of results is suggestive. Although excludable jurors as individuals had shown a slight tendency to rate defense witnesses and *more* credible than did death-qualified jurors, people who had served on mixed juries tended to rate them as less credible than did people who had served on the homogeneous death-qualified juries. Mean ratings of both defense witnesses were lower on mixed juries than on death-qualified juries, though the difference reached significance only for helpfulness ratings of the defendant's testimony ($\bar{x}_{\text{DQ}} = 4.35$, $\bar{x}_{\text{mixed}} = 3.91$; $t = 2.00$, $p < .05$). Both measures of the defendant's credibility also showed significant differences when only death-qualified jurors from the two jury types were compared (believable: $t = 1.98$, $p < .05$; helpful: $t = 2.08$, $p < .04$). Again, these differences are merely suggestive, because they disappear when initial vote preference is controlled for, but they are interesting because they indicate that the mixed juries may provoke more controversy and criticism, so that jurors become more skeptical of *all* witnesses for both sides, compared to jurors on death-qualified juries.

Likewise, mixed juries showed a slight tendency to rate both attorneys lower on all measures than did death-qualified juries, but these comparisons were also nonsignificant.

---

responses of the death-qualified jurors who served on homogeneous juries with those of the Death-qualified jurors who served on mixed juries. The results of these comparisons are identical to those reported in the text for the general comparisons of death-qualified vs. mixed juries, indicating that the experience of diversity does have significant effects on jury members' perception and memory.

In order to find out whether the excludable jurors' responses on the postdeliberation measures differed from those of the death-qualified jurors solely because of the excludables' experience on mixed juries (question 2), we performed a separate analysis, considering only those subjects who served on mixed juries, and comparing the excludable subjects with the death-qualified subjects. The results of this analysis are identical to those of the general comparison between death-qualified and excludable subjects, indicating that the excludables differ from the death-qualified subjects over and above the effects of their experience on mixed juries. Thus, in sum, there are independent effects of person and treatment.

Memory for Evidence and Judge's Instructions

It will be recalled that excludable subjects did not differ significantly from death-qualified subjects in their memory for the evidence. The presence of both types of jurors on the same jury, however, significantly improved the members' memory. Subjects on mixed juries remembered more evidentiary facts correctly ($\bar{x} = 9.18$) than did subjects on death-qualified juries ($\bar{x} = 8.54$, $t = 2.09$, $p < .04$). When covariation analysis partialed out the effects of prior voting, the effect was essentially equivalent ($F_{cov} = 4.95$, $p < .03$). The increased recall of mixed juries did not appear to result from the correction of any particular type of evidence error. Mixed and Death-qualified juries did not differ significantly in their ratios of prodefense, proprosecution, or neutral errors though, of course, members of mixed juries made fewer errors in all categories.

Unfortunately, jury diversity did not significantly improve the jurors' dismal performance on the test of memory for judge's instructions. Members of mixed juries scored an average of 11.89 out of the 18 true–false items, while members of death-qualified juries scored 11.63.

Assessments of Jury Deliberation

Excludable jurors were more critical of their juries than were death-qualified jurors. The same pattern of self-criticism was observed for mixed juries as compared with death-qualified juries. Mixed juries were less likely to believe the deliberations brought out facts ($t = 1.78$, $p < .08$) and were fair ($t = 1.99$, $p < .05$). Mixed juries found the case more difficult than did death-qualified juries ($t = 2.35$, $p = .02$). Additionally, subjects on mixed juries felt that the jury as a whole had changed less ($t = 3.58$, $p < .001$). In fact, there were no significant differences in voting changes between the two jury types.

DISCUSSION

In our research we have addressed two general questions about the effects of excluding all those who are adamantly opposed to the death penalty from the determination of guilt or innocence in capital cases. The first is the question of conviction-proneness: Are death-qualified jurors more likely than excluded jurors to find a defendant guilty? This question has been raised repeatedly by the courts and studied intensively by social scientists; our study is the latest in a long line of research. The second question involved other negative effects of the reduction in the diversity of viewpoints on the jury that results from death qualification. Since only one previous study included jurors who had had an opportunity to deliberate (Zeisel, 1968) and none has included postdeliberation questions about anything but the verdict, our study is merely a beginning to our understanding of the other effects of death qualification on jury performance.

Conviction Proneness

Let us begin with the bias toward conviction, where the results are conclusive. In 1968 the *Witherspoon* court decided that the data before them (prelimi-

nary versions of the Wilson, Goldberg, and Zeisel studies) were insufficient to persuade them that death qualification "results in an unrepresentative jury on the issue of guilt or substantially increases the risk of conviction" (*Witherspoon v. Illinois,* 1968, pp. 517–518), but that the question was an empirical question which might be resolved by more compelling data in a future case. Since then, the Zeisel and Goldberg studies have been published, and their results have been replicated by Jurow, Harris, and ourselves, to mention only those studies that deal directly with the correlation between willingness to impose the death penalty and the likelihood of voting to convict in real or simulated cases. There are of course numerous other studies, some reviewed and some presented in this issue, which have employed other methods but which converge on the same conclusion: death-qualified juries are more likely to convict. The validity and robustness of this conclusion are enhanced by the variety of empirical methods, samples, and stimulus contexts involved in the research. Samples have included black and white Southern college students, northern industrial workers, eligible jurors in California, actual jurors in New York and Illinois, and a large national random probability sample. All have found the same result. Stimulus contexts have ranged from brief paper-and-pencil descriptions of cases, to half-hour audiotapes, to our own 2½ hour videotape followed by deliberation, to actual criminal cases followed by real jury deliberations. All have found the same result. Empirical methods have ranged from retrospective interviews with jurors about their experiences, to national public opinion surveys, to a series of increasingly elaborate laboratory simulations. All have found the same result. As the controlled simulations become more realistic, the differences between death-qualified and excludable jurors becomes more pronounced.

The one study that failed to find a difference, that of Osser and Bernstein (1968) used type of case as a surrogate variable for death qualification, a confound which renders its results both irrelevant and uninterpretable. No such systematic error exists in any of the studies that have found evidence of conviction proneness, in all of which type of case was either held constant or allowed to vary at random. All in all, we can think of few topics of social scientific research where the evidence is so consistent and compelling.

The results of the present study add to the body of previous research in several ways. First, our videotaped trial was far more involving than any of the other stimulus materials used, with the probable exception of Zeisel's real trials. Members of our simulated juries debated heatedly, and many were loath to stop when the hour was over, although they had already spent close to four hours in the study and it was getting close to dinner time. Those who had participated in real juries felt that the experience matched their actual jury experience. Second, we determined that the differences persisted after an hour's deliberation. Third, our definition of the *Witherspoon* excludable group was more careful than any previous classifications, in that we restricted our sample to those jurors who would be eligible to serve *but* for their attitudes toward the death penalty, i.e., to those jurors who could be fair and impartial on the question of guilt. Fourth, we determined that the differences between the two groups were in fact due to their excludable vs. qualified status, and not to any differences in background

variables or other biases in the sample. Because our results confirmed those of the earlier, less sophisticated studies, we gain confidence that their results were not a function of particular weaknesses in the individual studies or of poor definition of the excludable group, but instead reflected true differences between death-qualified and excludable jurors.

Other Effects of Loss of Diversity

Our study also allows us to present some initial data on some other detrimental effects of the loss of diversity on the death-qualified jury. It must be remembered that these effects are not just side effects of conviction-proneness: they exist *after* we have partialed out the effects of first ballot verdicts, and thus represent additional, independent effects of the homogenization of viewpoints on the jury.

In considering the possible effects of loss of diversity, we suggested that death-qualified juries might be less stringent in their application of the reasonable doubt standard, less critical in their evaluation of the testimony, less thorough and energetic in their arguments, and less able to remember all the important evidence.

The only information collected in this study that might be relevant to the reasonable doubt standard is the subjects' responses to the question "What degree of probability that the defendant committed the crime would justify a verdict of guilty?" Probability judgments of this kind are hard for people to make, and there was high variability in responses. Neither juror attitude nor type of jury experience was significantly related to these judgments. Thompson, Cowan, Ellsworth, and Harrington (this issue), using a more sensitive and less confusing measure, found significant differences between death-qualified and excludable jurors, but the effects of homogeneous vs. mixed juries remains to be ascertained.

Our data do indicate that participation on the more diverse juries led jurors to be more critical in their evaluations of the witnesses. When we look at *individual* jurors, the direction of differences is what we would expect: Death-qualified jurors trust the prosecution witnesses more than excludable jurors do, and the defense witnesses less. The pattern is not exactly the same when we compare death-qualified and mixed juries. The mixed juries are more critical of *all* the witnesses, defense and prosecution, than the homogeneous juries. For some witnesses the difference reaches significance, while for others it does not, but the general direction suggests that a juror's unquestioning acceptance of *any* witness' testimony is unlikely to survive on a mixed jury: There will be someone who will point out inconsistencies or sources of error in the testimony. The credibility of each witness becomes a matter of debate, and the testimony may be examined more carefully.

Although we do not yet have data on the vigor and thoroughness of the jury's arguments, we do know that members of the death-qualified juries were more satisfied with their experience. Members of mixed juries felt that the case was more difficult and were less sure that they were being fair. They believed that their jury had changed less, even though voting changes were equal in the two

types of jury, and they were less likely to believe that the deliberations brought out important facts, even though they remembered more facts after the deliberations than did members of the death-qualified juries. These results suggest several possibilities. The mixed juries may have been more contentious, so that more people felt worried by the lack of harmony or frustrated by exposure to strong counterarguments. (Remember that the measures were taken *before* consensus was reached.) The level of controversy may have made them feel that they had a longer way to go before they could reach consensus. Or perhaps the lack of agreement on a large number of points raised their standards of adequate jury performance, making them realize just how painstaking their scrutiny of the facts and issues must be.

The data on memory for evidence is in some ways the most fundamental of all. If the jury fails to get the facts straight, it matters little how strenuously they argue or how critically they evaluate the witnesses. Our data show that the members of mixed juries remember the evidence better than the members of death-qualified juries. The facts we asked our subjects about were important ones: Could the doctor determine whether the stab wound came from a downward thrust or from someone running against the knife? Could the police officer see whether the victim was holding a razor? Mistakes could be consequential in determining the verdict. We expect that the superiority of the mixed juries is also a function of the likelihood that errors of fact are more likely to be corrected when there is a wider range of viewpoints and a higher level of controversy.

It is disappointing that the memory for legal instructions did not improve with diversity in the same way. It is our impression, from watching the taped deliberation, that jurors were considerably more willing to speak up and express different views of the facts than they were to disagree with an erroneous statement of the verdict category or other points of law, and that more jurors participated in the reconstruction of the events and the evaluation of the witnesses than in the legal definitions. Often one or two dominant jurors would "lay down the law"—incorrectly, and the few attempts at correction were hesitant and ineffective. It is our hunch that a confidant juror will often win out over an accurate juror in getting his or her version of the law accepted.

Our findings suggest that the courts' strong emphasis on the diversity of the jury is justified. Although our questions about the effects of the homogenization caused by death qualification were exploratory, and our findings should be viewed with caution until they are replicated, our data are sufficient to at least raise serious questions about the sacrifice in the quality of deliberation that occurs as the diversity of the jury is restricted.

Implications for the Law

The law is often slow to accept evidence based on empirical studies by social scientists. Questions are rarely posed in empirical terms; decisions are rarely influenced directly by empirical data. The *Witherspoon* case was exceptional for stating that whether the death qualification of juries substantially increases the risk of conviction is an empirical question. The Supreme Court was unconvinced

by the empirical data that were before it in 1968, but it suggested that empirical data could in principle provide a convincing answer to the question of conviction proneness. The court's rejection of the data in *Witherspoon* was justified, we feel. As social scientists, we would not like to see the law changed by preliminary summaries and second-hand reports of empirical studies.

We also feel that the more general caution about the direct application of social science research is justified, although it is often frustrating to us as social scientists. Judges and legal scholars have demanded full disclosure of the methods and analyses used in reaching conclusions, review of all the research on an issue whether it supports or disconfirms a position, acceptance of the findings by other social scientists or some degree of consensus within the social science community (e.g., Bazelon, 1981), and exposure to the "traditional testing mechanisms of the adversary process" (Opinion of Powell, J. in *Ballew v. Georgia,* 1978, p 246). These requirements are in some ways much more stringent than the standards that govern the more traditional methods of evaluating the issues in appellate court cases, but then the use of social science is relatively new, and appellate court judges, less familiar with social science methods, may feel they need more explicit safeguards.

We feel that this caution is justified if it is a sincere expression of the need for high standards to be met before research can be influential, if it is a concern expressed in good faith. If, on the other hand, the standards are intended to be *impossibly* high, designed to keep empirical social science research out of the courts altogether, then "caution" is simply a disingenuous mask for closed-mindedness. If the nature of the issue is changed from one case to the next, so that data collected in the first are no longer relevant to the second; if new criteria for excellence are added just as the old ones are met; if the issue is suddenly redefined as non-empirical just when the research is completed, then the frustration felt by the social scientists will legitimately turn to suspicion, and the law may be deprived of a valuable source of information.

The only published appellate court opinion that has considered all of the data on conviction proneness (and on the other questions discussed in this issue) is *Hovey v. Superior Court* (1980). This opinion, we feel, represents a sincere, careful, and highly competent consideration of empirical data by an appellate court, the best we have seen. The California Supreme Court concluded that the data were sufficient to demonstrate the conviction proneness of juries under *Witherspoon,* but that it was impossible to extrapolate from petitioner's showing about a " '*Witherspoon*-qualified jury' to his intended conclusion as to a 'California death-qualified jury' " (p. 1345), which also excludes those who would *automatically vote to impose the death penalty* in all cases where it is legally available. If substantial numbers of these "automatic death penalty excludables" are also barred from jury service, and if they are extremely conviction-prone, then their removal might reduce the imbalance toward conviction that we found among our death-qualified jurors.

When this issue surfaced, after our study had been completed, we went back to our subjects to see whether any of them should be excluded from the death-qualified group on this ground. On the death penalty scale question, 23 subjects

had indicated that they "strongly favored" the death penalty. Since the next most extreme opinion on the scale was to "favor the death penalty except in some cases," only those 23 subjects could possibly be "automatic death penalty excludables." We engaged a person who had no previous connection with the study to contact each of these 23 people and ask them whether they would consider voting for life imprisonment in some capital cases or whether they would automatically vote for the death penalty. The format of the question was parallel to the *Witherspoon* exclusion question.[15] Of the 22 subjects we managed to recontact, none indicated that he or she would automatically vote for the death penalty in every capital case. This leaves the remote possibility that the one subject we failed to locate was misclassified in our study, but his or her reclassification would not materially alter the results.

From the point of view of the California Supreme Court, however, the validity of this particular study is not the issue. They are raising an additional empirical question that the data did not address: How would the results on conviction proneness *in general* be affected if the role of the automatic death penalty jurors is considered? Although our initial disgruntled reaction was to see this as a new question sprung upon us without warning, it is a valid empirical question in the California context. A new round of research was necessary to answer this question, and the answer is that inclusion of this group of jurors would not materially alter any of the results on conviction-proneness or prosecution proneness (see Kadane, this issue, for a summary). In general, the response of the California Supreme Court has led to a strengthening of the data on conviction proneness. As social scientists, we were pleased with their careful analysis of the empirical issues, despite the outcome.

We are, of course, eagerly awaiting the further use of these data both in the California Supreme Court and in other courts. We have tried very hard to provide the sort of research that the courts have requested and so far we are encouraged by the response. Most of the data on conviction proneness, including our own, were collected in direct response to an empirical question posed by the United States Supreme Court. The designs, methods, and results of all of the research on this issue have been fully described in court, and the witnesses subjected to cross examination according to "the traditional testing mechanisms of the adversary process" (Powell, J., concurring, *Ballew v. Georgia,* 1978). Thus, as social scientists, we cannot help regarding the cases on death qualification that come before the appellate courts as test cases, not only of the Constitutional issues, but of the use of social scientific research in judicial decision making.

---

[15] The wording of the question was as follows:

> If you served as a juror in a first degree murder trial, and if the defendant was convicted of a crime for which the law permitted the jury to sentence him either to death or to life in prison, would you automatically vote for the death penalty in every case, no matter what the evidence was, or would you be willing to consider voting for life imprisonment or death depending on the evidence that you heard in the case?
>
> (1) I would automatically vote for the death penalty in every case.
>
> (2) I would be willing to consider voting for life imprisonment or death depending on the evidence that I heard in the case.

REFERENCES

Bazelon, D. L. Veils, values, and social responsibility. Address to the 89th annual convention of the American Psychological Association, Los Angeles, California, 1981.

Boehm, V. R. Mr. Prejudice, Miss Sympathy, and the authoritarian personality: An application of psychological measuring techniques to the problem of jury bias. *Wisconsin Law Review,* 1968, 734.

Carlsmith, J. M., Ellsworth, P. C., & Aronson, E. *Methods of Research in Social Psychology.* Reading, Massachusetts: Addison-Wesley, 1976.

Fitzgerald, R., and Ellsworth, P. C. Due process vs. crime control: Death qualification and jury attitudes. *Law and Human Behavior,* 1984, **8,** 31–52.

Goldberg, F. Toward expansion of Witherspoon: Capital scruples, jury bias, and the use of psychological presumptions in the law. *Harvard Civil Rights Law Journal,* 1970, **5,** 53.

Louis Harris & Associates, Inc. Study No. 2016 (Unpublished), 1971.

Hastie, R., Penrod, S. D., & Pennington, N. *Inside the jury.* Cambridge, Massachusetts: Harvard University Press (in press).

Jurow, G. New data on the effect of a death-qualified jury on the guilt determination process. *Harvard Law Review,* 1971, **84,** 567–611.

Oberer, W. E. Does disqualification of jurors for scruples against capital punishment constitute denial of fair trial on the issue of guilt? *University of Texas Law Review,* 1961, **39,** 545.

Osser, A. I., & Berstein, B. The death-oriented jury shall live. *University of San Fernando Valley Law Review,* 1968, **1,** 253–258.

Thompson, W. C., Cowan, C. L., Ellsworth, P. C., & Harrington, J. C. Death penalty attitudes and conviction proneness: The translation of attitudes into verdicts. *Law and Human Behavior,* 1984, 8, 95–114.

Tyler, T., & Weber, R. Support for the death penalty: Instrumental response to crime or symbolic attitude? *Law and Society Review,* 1982, **17,** 21–46.

Vidmar, N., & Ellsworth, P. C. Public opinion and the death penalty. *Stanford Law Review,* 1974, **26,** 1245–1270.

White, W. S. The constitutional invalidity of convictions imposed by death-qualified juries. *Cornell Law Review,* 1973, **58,** 1176.

Wilson, W. C. Belief in capital punishment and jury performance. Unpublished manuscript. University of Texas, 1964.

Zeisel, H. Some data on juror attitudes toward capital punishment. Monograph. Center for Studies in Criminal Justice, University of Chicago Law School, 1968.

Cases

Apodaca v. Oregon, 406 U.S. 408 (1972).
Ballew v. Georgia, 435 U.S. 223 (1978).
Duren v. Missouri, 439 U.S. 357 (1979).
Hovey v. Superior Court, 616 Pac. 2d, 1301 (1980).
Norris v. Alabama, 294 U.S. 587 (1935).
Taylor v. Louisiana, 419 U.S. 522 (1975).
Witherspoon v. Illinois, 391 U.S. 510 (1968).