

To Tell What We Know or Wait for Godot?*

Phoebe C. Ellsworth

The evidence that death-qualified jurors are more likely than excluded jurors to convict is consistent, robust, and directly relevant to the issues of representativeness and conviction proneness that were before the Supreme Court in *Lockhart v. McCree*. These are exactly the circumstances in which an *amicus* brief from the APA is most appropriate. In science the search for knowledge is never complete; to keep silent until our understanding is perfect is to keep silent forever.

Samuel Johnson, *Rasselas*

To talk in public, to think in solitude, to read and to hear, to inquire and answer inquiries,
is the business of a scholar.

Professor Elliott raises two questions about the American Psychological Association's practice of submitting *amicus* briefs to the courts. First, are our data sufficiently valid, consistent, and generalizable to be applicable to the real world issues? Second, are *amicus* briefs adequate to communicate scientific findings? The first of these is not a general question, but must be addressed anew each time the Association considers a new issue. An evaluation of the quality and sufficiency of scientific knowledge about racial discrimination, for example, tells us nothing at all about the quality and sufficiency of scientific knowledge about sexual abuse. "Are the data adequate?" is an *ad hoc* question. It only becomes a general question if an attempt is made to propose general standards of adequacy. The second question—Are *amicus* briefs adequate to communicate scientific find-

* I would like to thank Reid Hastie, Richard Lempert, and Robert Mauro for their comments on an earlier draft and for their efforts to cheer me up. Requests for reprints and other correspondence should be sent to Phoebe C. Ellsworth, 5242 Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

ings?—is a general question, and it is this question that Professor Elliott claims to address (p. 60).

With the exception of a few brief words at the end of the paper, however, Professor Elliott does not address the issue of the merits of amicus briefs or their alternatives. Instead, the paper is an attack on the research presented in one particular brief, the APA amicus brief in *Lockhart v. McCree* (1985), and thus is of considerably less general interest. He argues that the examination of this particular brief provides “a test of the propriety in general of the APA’s use of scientific briefs” (Elliott, p. 61), but he never makes the connection between the specific case and the general issue. He vaguely argues that this brief provides a test because it is one of APA’s strongest ones, but he provides no comparative evidence to support this statement, except to point out that one of the authors of the brief said it was strong (Bersoff, 1987). Nor is it clear why a “strong brief” is a test of anything. If Elliott really intended to make the general argument that APA should exercise more caution in submitting briefs, a “weak” brief would be at least as valuable an example as a “strong” brief. But no case study is really appropriate to make this argument: In order to demonstrate that the APA has shown a general lack of prudence in submitting briefs, what is required is an analysis of a representative sample of APA’s amicus briefs. Ideally, these would be compared with a sample of amicus briefs submitted by other groups, in order to find out whether APA’s were more or less responsible than the norm.

Thus the real point of Elliott’s article does not involve APA amicus briefs in general, or the *Lockhart* brief as a test case of anything, but the *Lockhart* brief *per se*. In replying, I shall first deal with his criticisms of the research. I believe that on the average, death-qualified jurors are more likely than the excluded jurors to vote for conviction in serious felony cases and that the process of a death-qualifying voir dire exacerbates the tendency to convict. I also believe that over many cases, abolishing the practice of death qualification would result in fewer verdicts of capital murder, both because the voices of the less conviction-prone jurors would be heard on more juries and because the suggestive questioning involved in the death-qualifying voir dire would no longer occur. Obviously there would be many cases, probably a large majority, where the verdict would be the same with or without death qualification. But there would be some—possibly a large minority, possibly even a majority if public opinion on the death penalty shifts back to where it was 25 years ago—where the inclusion of the adamant opponents of the death penalty would change the verdict.

I shall then say a few words about the APA’s role as amicus, both in the *Lockhart* case and in general.

The Concept of Convergent Validity

In the majority opinion in *Lockhart v. McCree* (1986), Justice Rehnquist concluded that the research was still insufficient to form the basis of a constitutional rule. He did so by lining up the 15 studies,¹ identifying some particular flaw

¹ In referring to 15 studies, I include the survey studies of the relation between death penalty attitudes

in all but one of them, and eliminating each one in turn from the set as unworthy of consideration. Elsewhere (Ellsworth, 1988) I criticized this one-by-one elimination of the studies in a consistent body of research as exemplifying an ignorance of the concept of convergent validity. The idea that scientific proof is provided by a single flawless study is generally mistaken. Typically, there are many sources of error and many confounding variables that must be controlled, and typically it is impossible to control them all in a single study. Thus the scientist must approach the truth in steps, ruling out some alternative explanations in some studies and other alternative explanations in other studies until only one explanation is left that can account for the results of *all* the studies. This is the method of convergent validation (Campbell & Fiske, 1959; Ellsworth, 1977), taught in every undergraduate research methods course.

Perhaps a Supreme Court Justice may be excused for a failure to understand the concept of convergent validation (although it was clearly explained in the APA amicus brief). It is rather more surprising that Professor Elliott uses much the same logic in evaluating the research. He argues that only three studies use the correct *Witherspoon* question to define the excluded group, and that "three cases . . . are not enough to support a conclusion other than *not proven*" (p. 67). Only three studies involved deliberation. Only a few studies eliminated nullifiers from the sample. He even proposes the "ideal experiment," one that, like almost all "ideal experiments," is doomed to failure. For a person who claims to be an advocate for higher standards of scientific proof, this method of criticism is singularly unscientific. What is needed is an alternative explanation. Why *do* all the studies find that death-qualified jurors show more proprosecution attitudes and greater willingness to convict?² If there are no differences between death-qualified and excludable jurors, one would expect the typical finding to be that they did not differ, with an occasional study finding that death-qualified jurors were relatively

and proprosecution attitudes. An actual vote to convict in a real or simulated case is one measure of conviction proneness in general; a set of beliefs and values that favor the prosecution is another. The total exclusion of all the survey research from consideration is inappropriate. In addition, before the *Lockhart* decision, it was not at all clear that the correlation between death penalty attitudes and jury verdict was the sole or even the major constitutional issue. The Court in *Grigsby* held that the skewed attitudinal composition of death-qualified juries might itself be a violation of the Sixth Amendment. Professor Elliott's narrowing of the issue to conviction proneness and jury verdicts depends on a decision that was handed down after the APA brief was filed.

² Of the four studies that have failed to demonstrate the conviction proneness of death-qualified jurors, three are unpublished studies conducted by Professor Elliott, which were not available at the time the APA brief was written (Nov.-Dec. 1985). The one published study, by Osser and Bernstein (1968), was rejected by the courts and the prosecution experts in *Hovey v. Superior Court* (1980) and in *Grigsby v. Mabry* (1983, 1985). Osser and Bernstein found that juries in robbery and burglary cases were no less likely to convict than juries in murder cases and concluded that death qualification made no difference. Since juries are only death qualified in *capital* murder cases, the comparison groups overlap substantially. More important, murder cases may differ from burglary and robbery cases in a host of other ways (zeal of defense counsel, pressure to go to trial, percentage of repeat offenders, and consequences of conviction, to name just a few of the myriad possibilities), so that uncontrolled comparisons are worthless. This study was reported in the APA brief, and discussed in the evidentiary record. Thus, the suggestion that Professor Elliott makes in his Response that the writers of the APA brief might avoid mention of studies with inconsistent results is ill taken.

prosecution and about an equal number showing that they were relatively prodefense. Instead, out of the 15 published studies that were on point, 15 showed a prosecution bias.

The Use of an Appropriate Witherspoon Standard for Exclusion

One of Professor Elliott's major criticisms of the research is that only three of the studies defined the excluded group according to an appropriate *Witherspoon* standard: those who "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" (*Witherspoon v. Illinois*, 1968, p. 522, footnote 21). Although all three of the studies found the predicted effect, three studies are too few to rely on.³ In fact, the wording of the question is immaterial. It is immaterial from a scientific point of view because *wherever* researchers have drawn the line dividing opponents of the death penalty from other jurors, the results have been the same: Proponents have been more favorable to the prosecution, more likely to vote for conviction. The data themselves demonstrate that drawing the line in exact correspondence to the *Witherspoon* standard makes no difference. If there is a monotonic relationship between death penalty attitudes and conviction proneness, the difference between proponents and opponents will show up wherever the sample is split. The variety of methods used to assess opposition to the death penalty provides strong evidence that the relationship is in fact monotonic and thus strengthens, rather than weakens, the conclusion.

The use of the proper *Witherspoon* wording is immaterial from a legal point of view because the *Witherspoon* standard is *not* the legal standard. In 1985, in the case of *Wainwright v. Witt*, the Court held that the proper standard was whether a juror's attitude toward the death penalty would "prevent or substantially impair the performance of his duty as a juror in accordance with his instructions and oath" (1985, p. 850). All of the *Witherspoon* excludable subjects are included in this definition, since by definition their attitudes prevent them from following the judge's instructions at the *penalty* phase, but the current standard (which was the standard when the APA filed its brief) excludes a larger number of jurors from service in potentially capital cases (cf. Thompson, 1989). The actual line is somewhere between the line defining those who could never impose the death penalty and the line defining those who are generally opposed to the death penalty—nicely bracketed between the two standards that have been most extensively researched. Thus the studies using a "general scruples" standard are not to be discarded, but in fact are particularly useful: If they show an effect, and if the studies using a *Witherspoon* question show an effect, then it becomes very implausible that an

³ In fact, there are only two; the Jurow (1970) study used a scale question which had a *Witherspoon*-like alternative at one end, but responses to scale questions do not reflect responses to the sort of yes-or-no question asked in court. In pretesting our own research, we discovered that respondents usually use the extreme ends of the scale to reflect that they "strongly oppose" or "strongly favor" the death penalty without paying much attention to the specific label (see also Dawes, 1985).

intermediate criterion would not show an effect. Thus, reducing the population of relevant studies to three is inappropriate.⁴

Finally, it behooves any critic who is so concerned about the applicability of our research results to the real world to take a moment to consider what typically happens in the real world. The process of jury selection does not end once the death-qualification process is over and the most adamant opponents of the death penalty have been excused. The attorneys still have their peremptory challenges. Some research indicates that the prosecutor's peremptory challenges are used to exclude those opponents of the death penalty who survive the challenge for cause (Winick, 1982). Thus, even when the *Witherspoon* standard was in force, those who fit the strict wording of the standard were not the only opponents of the death penalty who could be excluded from the capital jury. Given the low levels of opposition to capital punishment in some jurisdictions, it would not be difficult for a prosecutor to achieve a *pre-Witherspoon* jury—a jury “culled of all who harbor doubts about the wisdom of capital punishment” (*Witherspoon v. Illinois*, 1968, p. 520), a jury in which death penalty opposition and not *Witherspoon* excludable status is the operative criterion for exclusion.

Thus to restrict the universe of relevant studies to those in which the proper *Witherspoon* standard was used shows a disregard for scientific standards of convergent validity, for the current legal standard of exclusion laid down in *Wainwright v. Witt* (1985), and for the realities of the actual practice of jury selection in capital cases.

Significance and Statistics

One of the great problems in explaining social science research to lay people, including judges, is the difficulty of counteracting their tendency to lump the available studies into two groups—“statistically significant” and “nonsignificant,” or worthless—as though on an otherwise continuous scale there were an enormous distance between a probability level of .05 and a probability level of .06. This misperception is undoubtedly our own fault as social scientists; we too frequently fall into the trap of regarding our own arbitrary convention for declaring a result “significant” to be far more meaningful than it is, despite repeated exhortations from within our own ranks to avoid this mistake. The level of significance that matters should depend on the decision to be made. I think none of us would be happy if the Supreme Court, or any other court, decided that a procedure was unfair if the data demonstrating a bias were significant at the .05 level but that there was *no evidence* of unfairness if the results only reached the .06 or the .10 level.

Professor Elliott, of course, does not make this sort of simple-minded claim, but he is always careful to comment on the significance level of the studies, noting

⁴ Legally, since McCree's trial occurred in 1978, the process of death qualification in his case was governed by *Witherspoon*, not by *Witt*. However, the *Witt* decision is relevant to the general scientific issues, and particularly to Professor Elliott's claim that only studies using the *Witherspoon* criterion should be considered, since *Witt* is the standard that has governed all trials since 1985.

those that are marginally significant, attributing the significance of one (Harris, 1971) to its large sample size (Elliott, p. 63), and recalculating the results of other studies so that they come out insignificant. Thus by innuendo he creates the impression that certain studies may no longer be worthy of consideration and that the .05 level is somehow the standard of acceptability.

He points out, for example, that general death penalty attitudes correlate more highly with verdicts than does the attitude defining exclusion under *Witherspoon*, implying that *Witherspoon* excludables may not differ from other jurors. He does not point out that using several points to divide up a linear or monotonic scale automatically yields higher correlations than using just one point.

He makes sure to point out that Zeisel's (1968) data were significant at the .04 level only because a one-tailed test was used, as though if there were a .08 probability that the results were due to chance, the Zeisel study would not show evidence of conviction proneness. He then recategorizes the Zeisel data into three separate subgroups, performing separate chi-squares on each one, conveniently reducing the sample size for each test so that significance is less likely. There are many ways that the Zeisel data can be analyzed, some of which are described and evaluated in the extensive record before the lower courts and the Supreme Court. Most of them find a significant effect. The most recent recalculation uses the arcsine transformation for tests of proportions (Langer & Abelson, 1972), and finds a z of 2.10, for a *two-tailed* probability of .036, consistent with most of the earlier analyses. (Mauro, 1991).

Likewise, Professor Elliott, with no explanation, recalculates the Horowitz and Seguin (1986) results and reduces them to insignificance. Use of the arcsine transformation on these data yields a z of 1.824, with a p value of .068, two-tailed; .034, one-tailed (Mauro, 1991). Given 30 years of consistent findings, use of a one-tailed test hardly seems a scandal.

Finally, Professor Elliott's own studies, both the ones he has already conducted and the "ideal study" he proposes, employ sample sizes so small that they are *guaranteed* to produce nonsignificant results unless the effect of death penalty attitudes on verdicts is far greater than anyone has ever found, or claimed.

We may look at statistical significance in another way, one that Professor Elliott avoids mentioning, and that is to ask, What is the probability that the combined results of all 15 studies presented to the Supreme Court in *Lockhart v. McCree* were due to chance? Or, what is the probability that the results of the subset of studies using verdicts as a dependent variable were due to chance? These analyses, using the technique described by Mosteller and Bush (1954), were in the record that was before the Lockhart court and that was evaluated by the drafters of the APA amicus brief, and of course they indicate that the likelihood of this many studies coming out with effects this large by chance is infinitesimal.

Professor Elliott also reports that he himself has thrice failed to replicate the relationship between death qualification and tendency to convict. It is difficult to understand why someone who has the methodological sophistication to identify so many flaws in the existing research would choose to carry out the particular studies that Elliott contributes to the corpus of research on death qualification. Rather than studying jury-eligible citizens, he studies college students; he allows

the amount of time between the trial and the predeliberation verdict to range from hours to days; he mentions no attempts to control communication among the subjects in the interval; and he allows the size of the deliberating group to vary. He also runs too few subjects to make statistical significance a realistic possibility. There may be other flaws in the studies; they are only described in a footnote, and the description of the method is far too sketchy to permit an adequate evaluation. I fail to see the purpose of conducting noisy, low-powered studies on a topic that has already been studied in a far more rigorous fashion.

Courts often look for ways to find a body of data “inconclusive” when the data support an unpalatable holding, and one of the easiest ways to do so is to point to disagreement in the scientific community. Sometimes, as in the case of research on the deterrent effects of the death penalty, a *single* dissenting voice from the scientific community provides an opportunity to erase an empirical issue from serious consideration on the grounds that “experts disagree” (Ellsworth, 1988). Given the courts’ lack of methodological expertise, a bad study with findings that are inconsistent with the rest of the research on a topic can take on an authority that is incommensurate with its merits. It is distressing to think that the most influential thing a social scientist can do is to give the Court an excuse to reject some body of research. I do not think Professor Elliott conducted his studies in order to provide ammunition to the proponents of death qualification—that is hardly necessary after *Lockhart*—but I do think he should have been aware of the possible consequences of conducting and reporting research that falls so far short of the current standards of the field and reporting his conclusions before the research itself has been reviewed and published. The possible consequences are matters of life and death.

Professor Elliott also suggests an “ideal” study (p. 69) with a death-qualifying voir dire, a proper *Witherspoon* question, capital cases, and 10 juries per cell. Of course no study is the ideal study, and this one is no exception. The *Witherspoon* question is no longer appropriate, and given that the *Witt* standard allows a far wider scope for judicial discretion in deciding who can be excluded, designing an appropriate death-qualifying voir dire and identifying the excludable group becomes a much more difficult enterprise, one that cannot, I think, be accomplished in a single study (cf. Thompson, 1989). Simulations of *capital* cases also maximize the objection that a simulation cannot capture the “felt responsibility” of real jurors. I am reasonably satisfied that my research captures much of what jurors in real cases go through in deciding guilt or innocence, especially as the results are consistent with those of studies of real jurors (Zeisel, 1968; Moran & Comfort, 1986; Luginbuhl, Kadane, & Powers, 1990). I would be less confident that I could simulate a decision where the defendant’s life was at stake, and I would not be at all confident that any court would accept the validity of such a simulation. Like Professor Elliott, I think it is highly plausible that death penalty attitudes would have a more powerful effect in cases where the death penalty is a salient issue. However, I preferred to use a case where the effect might be weaker because I felt that no court would believe a “simulation” of a capital case, and I was not sure that I would either. Adding such a study to the record at this point might be useful, but I would certainly not consider it to be a

better test of the hypothesis. Finally, with 10 juries per cell, the undertaking would not only be large, but would also be doomed. Statistical significance could not be achieved for a guilty–innocent split in a given case unless over two thirds of the mixed juries arrived at not-guilty verdicts and *none* of the death-qualified juries did. Thus the ideal study is ideally designed to produce nonsignificant findings.

In addition to his emphasis on statistical significance, Professor Elliott also repeatedly mentions the low correlation between death penalty attitudes and verdicts and the small percentage of the variance accounted for. If there is an effect of death penalty attitudes on conviction rates, it is probably not very large, he argues. Like Justice Rehnquist, he cites with approval Finch and Ferraro's (1986) statement that "no definitive conclusions can be stated as to the frequency or the magnitude of the effects of death qualification" (p. 66). Like Justice Rehnquist, he misrepresents Finch and Ferraro's position. They refer to the conviction proneness of death-qualified juries as a "confirmed phenomenon" (p. 70), and in discussing the frequency and the magnitude of the effects of death qualification, they suggest that "extant research findings may actually *understate* the magnitude of the problem raised by death qualification" (p. 62, emphasis in original).

Elliott's arguments about the size of the effect, by the way, do not constitute a criticism of the APA brief, which states clearly that the research record does not permit us to draw conclusions about the magnitude of the effects of death qualification:

Inevitably, juries composed of a mixture of death-qualified and [*Witherspoon* excludable] jurors will fall somewhere between a jury composed solely of [*Witherspoon* excludables] or [death-qualified jurors]. Death qualified jurors are more conviction prone than such mixed juries would be. Thus, the size of the difference between the pure [*Witherspoon* excludable] and pure [death-qualified] groups in the research studies cannot be taken as an accurate estimate of the size of the difference between jury verdicts, although the existence and direction of the effect is clear from the studies. The precise magnitude of the difference in a given case will change as a result of many variables, e.g., quality of lawyering, first ballot verdict, number of WE's on mixed juries, strength of the evidence.

The research demonstrates that the composition of juries in terms of death penalty attitudes is an important variable and that over the long run eliminating [*Witherspoon* excludable jurors] will increase the number of guilty verdicts." (1985, p. 28)

Thus it is hard to fault the APA brief for failing to point out that we cannot estimate the size of the effect on the basis of the research. It may be smaller than the effects found in the studies. It may be larger, as Finch and Ferraro (1986) suggest. Craig Haney (1984) found that the very *process* of going through a death-qualifying voir dire suggests to jurors that the defendant is probably guilty. In the real world, this "process" effect would be *added* to the selection effect. In the real world, especially following *Witt*, but even before, more opponents of the death penalty are excluded than in the studies using a proper *Witherspoon* question. In the real world, most of the trials involving death qualification are capital trials, so the attitudinal effects may be more striking. We simply do not know. Had the Court held that the *magnitude* of the effect was the issue of constitutional significance and that the research was insufficient to decide that issue, the deci-

sion would have reflected an appropriate use of the empirical data (though perhaps not of the U.S. Constitution).

I would also like to point out that I am often suspicious of attempts to dismiss variables as trivial because they account for very small proportions of the variance. Interesting social situations, including most situations involving complex decisions, are complicated, and the outcomes are determined by many forces. Single variables rarely account for much of the variance unless the situation is very simple, or the variable that predicts is very similar to the outcome. Thus reinforcement accounts for a lot of the variance in the pecking of starving pigeons in Skinner boxes, but much less of the variance in the good behavior of prisoners. The way a child greets a returning parent in a controlled laboratory situation at the age of 18 months accounts for a lot of the variance in the way the same child greets the same parent in the same situation at age 6, but it does not explain much of the variance in other affectionate behaviors or in greeting other people. Many of the variables we consider most important in life account for very small amounts of the variance in people's reports of how happy they are. Andrews and Withey (1976) found that age, income, education, race, and gender, *taken all together*, account for only about 8% of the variance in life satisfaction. Jury trials are complex events, and a person's attitude toward the death penalty is only one of hundreds of factors that may affect his or her decision in a given case. It differs from most of the other factors, however, in that it is the one that survives from one murder case to another. Its power is its consistency. Over many trials, the addition of excludable jurors may have an effect that is not only statistically significant, but important as well.

Abelson (1985) makes this argument far more cogently than I can and uses a far more vivid example: a major league ball player coming up to bat. An "at bat," like a juror's decision, is a complex event. The outcome is predicted by many things. One of these is the player's batting average. How much of the variance in whether or not he gets a hit is explained by his batting average? The answer is about one third of one percent, less even than Professor Elliott's gloomiest estimate of the variance in verdicts explained by death penalty attitudes. Should we conclude that the relationship between batting average and likelihood of getting a hit is "not proven"? Of course not. Over the long run the effects are substantial—over repeated ball games, and over "repeated decisions by ideologically similar policy makers" (Abelson, 1985, p. 133), such as juries.

But Can We Trust the Excluded Jurors?

After spending the bulk of the paper arguing that social science has failed to demonstrate any substantial difference between death-qualified and excludable jurors, Professor Elliott turns around and argues that inclusion of the *Witherspoon* excludables on capital cases might unfairly bias the jury *against* the state. Removing all of these people does not create a noticeable bias against the defendant, but including them could create a bias against the state. How can this be? Elliott's theory is that in a *capital* case, all the prodefendant leanings that have failed to show themselves in previous research will emerge in powerful form. This argu-

ment assumes that we cannot believe these jurors when they say they can be fair and impartial. Nor, I suppose, can we trust the normal process of voir dire to identify their partiality.

Legally, this argument rests on vapors. In sustaining a challenge for cause, an attorney must show, one by one, that each individual juror is biased, rather than imputing bias to a broad category of jurors as a group. Usually the juror has to admit to the bias before he or she can be challenged for cause. The exclusion of a whole group of people who say they are impartial on the grounds that some of them may be deceiving the court has been explicitly rejected in every death-qualification case heard by the Supreme Court (*Witherspoon*, 1968; *Adams v. Texas*, 1980; *Witt*, 1985). In *Adams*, the Court held that it was impermissible to exclude jurors whose evaluation of the evidence on guilt would be affected by their death penalty attitudes. Fairness does *not* require that a person be untouched by the possibility of death.

Nor in our view would the Constitution permit the exclusion of jurors . . . who frankly concede that the prospects of the death penalty may affect what their honest judgment of the facts will be or what they may deem to be a reasonable doubt. Such assessments and judgments by jurors are inherent in the jury system, and to exclude all jurors who would be in the slightest way affected by the prospect of the death penalty or by their views about such a penalty would be to deprive the defendant of the impartial jury to which he or she is entitled under the law. (*Adams v. Texas*, 1979, p. 50)

Empirically, this argument rests on thin air. Elliott cites a study by Lee Ross and myself (Ellsworth & Ross, 1983) in which we found that people thought they would require more evidence to convict given the prospect of a *mandatory* death penalty, not a “possible” death penalty, as Elliott states. If voting for guilt *is* voting for death, of course we would expect death penalty attitudes to play a role. His second source of support is a midterm exam question asked to the very same subjects he used to show that death penalty attitudes make no difference in verdict. He finds that the *Witherspoon* excludables are more likely than the death-qualified subjects to say that the prospect of a death penalty would affect their determination of guilt, a response specifically approved in *Adams*.

These data amount to little more than speculation—a question about mandatory capital punishment and an exam question for a college class. In places, Elliott admits that we know nothing about this issue at this time; in places, he uses phrases like “it appears from the limited data available,” evincing far more tolerance towards these data than he does to the 15 studies in the APA brief. As Finch and Ferraro (1986) put it, “it seems somewhat disingenuous to grasp at the unproven assumption that excludable jurors are latent nullifiers as a means of rebutting the confirmed phenomenon of conviction proneness” (p. 70).

Should the APA Have Submitted a Brief?

Witherspoon v. Illinois (1968) was an unusual decision in that the Supreme Court explicitly left open the question of the relationship between death qualification and conviction proneness. The Court itself invited the research that was done. By 1980 (*Hovey v. Superior Court*), it was clear that the empirical research

would be a central issue in the higher courts' reconsideration of the constitutionality of death qualification. The brief for petitioner in *Lockhart* (the State of Arkansas) would inevitably argue that the research was worthless. The brief for respondent (McCree) would inevitably argue that it was excellent. In such a case, where a major issue before the court is the validity of psychological research, it seems to me that the APA would be remiss in its duties if it did *not* submit a brief. The APA was not "giving psychology away," thrusting it upon parties who would not otherwise have considered the research. The APA could either trust the adversarial system to sort out the evidentiary claims of petitioner and respondent, or it could comment. It did not have the option of keeping the research out of the Court's view.

But as I said earlier, I do not think Professor Elliott is really writing about amicus briefs in general; I think he is writing about *this* brief. Even in this case, he might have looked with approval at an APA brief that attacked the research, for example. Such a brief would certainly have appeared far more influential than this one was.⁵ The majority could have written a very brief opinion stating that the claim that death-qualified juries are conviction prone "cannot be said to be supported" (Elliott p. 74). The dissenters would have had little to say if the APA had decided to sabotage the research. Such a brief would have been misleading and irresponsible.

Elliott implies that by the usual standards of peer evaluation, the APA brief in *Lockhart v. McCree* would have been rejected. This is a misleading analogy. A brief is not a free-standing document. It is a comment on the record in the case, and its aim is to provide information that is helpful to the Court in reviewing that record. The record in the case included the *Grigsby v. Mabry* record (1980; 1985), involving the testimony of three experts, and a transcript of the testimony of a fourth from *Hovey v. Superior Court* (1980). These witnesses presented some 270 evidentiary exhibits and were extensively cross-examined. Several statistical analyses were done on many of the studies. Numerous supplementary studies were discussed in the record that did not appear in the APA brief. Three expert witnesses testified for the state, commenting on the research. Many of the points Professor Elliott raises were extensively discussed in the evidentiary record and in the lower court opinions. Amicus briefs are strictly limited to 30 pages. Peer review of the brief without a review of the record (thousands of pages) would make little sense, given that the brief is a comment on the record.

Secondly, I think it is presumptuous to claim that "most scientists of the Association" would judge the research described in the *Lockhart* brief to be "inadequate" (p. 74). First of all, most of the studies were reviewed (probably by APA members) and deemed worthy of publication. Secondly, even the

⁵ It is interesting that Professor Elliott never discusses the Supreme Court's irresponsible treatment of the empirical issues in *Lockhart v. McCree* or the treatment of the research in the Arkansas brief. To question the advisability of submitting briefs without looking at how the data have been distorted in other briefs before the Court, or how the Court has actually decided the cases, is to create an ideal, imaginary context for the evaluation of APA policy. The research record may fall short of perfection, but the Court's opinion is embarrassing.

prosecution experts in the lower courts had few serious criticisms of the research. Third, aside from Professor Elliott's review, the published reviews of the death-qualification literature have praised the research and criticized the Court's use of it (Hans & Vidmar, 1986; Mauro, in press; Monahan & Walker, 1980; White, 1987). There is no hint that these authors are failing to apply the usual standards of peer evaluation. Finally, new studies continue to find the effect (Horowitz & Seguin, 1986; Luginbuhl, Kadane, & Powers, 1990).

The District Court in *Grigsby v. Mabry*, having reviewed the whole record, held that the process of death qualification created conviction-prone juries, and the Court of Appeals affirmed. The usual legal standard for reversing a district court's finding of fact is that the finding be "clearly erroneous" (Rule 52(a), Federal Rules of Civil Procedure). The Supreme Court has stated that it "cannot undertake to review concurrent findings of fact by two courts below in the absence of a very obvious and exceptional showing of error" (*Graver Tank and Mfg. Co. v. Linde*, 1949, at 275). The brief for Arkansas attempted to show that the lower courts' interpretation of the social science research was "clearly erroneous." McCree's brief argued that it was not. Several states, as amici, joined in a brief that also criticized the research. The criticisms of the research were mostly unsophisticated and included sweeping attacks on social science research and statistics in general. This was the context in which the APA decided to submit a brief, and much of its content was devoted to answering the inappropriate criticisms raised by Arkansas and the amici states. When the value of psychological research is clearly at issue, I think the Association has a duty to comment.

As for Professor Elliott's proposed alternatives, it is hard to see that they would have made much difference in this case, given the consensus among the commentators on the value of the research. Practically, we might ask, who would be asked to serve on a National Academy Panel? Who would the courts appoint as experts? The APA brief was written by Craig Haney, Reid Hastie, John Monahan, Michael Saks, and myself, along with APA counsel (Bersoff and Ogden). Other experts might include Hans Zeisel, William Thompson, Joseph B. Kadane, Richard Lempert, Neil Vidmar, Valerie Hans, Steven Penrod, Tom Tyler, and perhaps a couple of dozen others. It is hard to imagine a group of experts that would seriously disagree with the APA's conclusions unless it were hand picked to do so.

A Final Word

Social science evidence is increasingly prevalent in the legal system in general and in appellate court decisions in particular. It is not a question of our "giving psychology away"; it has already been taken. We cannot take it back. Research about human behavior, as well as statements about human behavior that have no scientific basis whatsoever, will continue to appear in briefs and in court opinions. Rather than "giving psychology away," the APA's role is to put it in perspective. The APA cannot possibly keep up with all the uses and misuses of psychological research and psychological speculation in the legal system, but it has an obligation to do so in some cases.

First and foremost among these are cases where the psychological research is a central issue, as it was in *Lockhart v. McCree*. Sometimes our role will be to point out the weaknesses of psychology: We cannot reliably predict who will commit a future violent act however much the courts or legal standards may “need” this information. Sometimes our role will be to point out the insufficiency of psychology: We do not have enough research *yet* to say which forms of questioning elicit the most accurate answers from child witnesses, but in the future we may. Sometimes our role will be to endorse a body of research presented in the lower court records. Endorsement should not be reserved for cases where the research is perfect, or where there is nothing more to be done, because that day never comes in science.⁶ Instead, the APA’s decision should be made in the context of the decision facing the court. The Supreme Court had to decide whether death qualification makes no difference, or whether it does make a difference in the likelihood of guilty verdicts. The APA evaluated the record and wrote a brief saying that it makes a difference; although we cannot be certain about the size of that difference in the real world, we can be quite confident about its reality and its direction. This seems to me far more responsible than writing a brief saying that death qualification probably makes no difference, or that the difference is “not proven,” or than telling ourselves that it is somehow more scientific to say nothing at all.

REFERENCES

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133.
- Adams v. Texas, 448 U.S. 38 (1980).
- American Psychological Association (1985). In the Supreme Court of the United States: *Lockhart v. McCree*: Amicus curiae brief for the American Psychological Association.
- Andrews, F. M., & Withey, S. D. (1976). *Social indicators of well-being: American perceptions of life quality*. New York: Plenum.
- Bersoff, D. N. (1987). Social science data and the Supreme Court: *Lockhart* as a case in point. *American Psychologist*, *42*, 52–58.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (vol. 1, pp. 509–566). New York: Random.

⁶ In this response, Professor Elliott says that I do not harbor even the “seed of some fugitive doubt” about the relationship between death qualification and verdicts, and that it is this absolute certainty that makes me think filing the brief was the right thing to do. That is not so. I think we should file briefs when we believe that we have something to say that would improve the quality of the Courts’ decision making. I think we should say what we know when the issue arises, without waiting for absolute certainty. Nor does the idea that we have to be absolutely certain that a procedure is unfair before rejecting it square with the legal standards for capital cases. It is the *fairness* of the procedure that has to be proven, not the unfairness, if its use is to be permitted in capital cases. “When a defendants’ life is at stake the Court has been particularly sensitive to insure that every safeguard is observed” (*Gregg v. Georgia*, 1976, at 187).

- Elliott, R. (1991). Social science and the APA: The *Lockhart* brief as a case in point. This issue. *Law and Human Behavior*, 15.
- Ellsworth, P. C. (1977). From abstract ideas to concrete instances: Some guidelines for choosing natural research setting. *American Psychologist*, 32, 604–615.
- Ellsworth, P. C. (1988). Unpleasant facts: The Supreme Court's response to empirical research on capital punishment. In K. C. Haas & J. Inciardi (Eds.), *Adjudicating death: Moral and legal perspectives on capital punishment* (pp. 177–211). Beverly Hills: Sage.
- Ellsworth, P. C., & Ross, L. D. (1983). Public opinion and the death penalty: A close examination of the views of abolitionists and retentionists. *Crime and Delinquency*, January, 116–169.
- Federal Rules of Civil Procedure 52(a).
- Finch, F., & Ferraro, M. (1986). The empirical challenge to death-qualified juries: On further examination. *Nebraska Law Review*, 65, 21–74.
- Graver Tank and Manufacturing Company v. Linde, 336 U.S. 271 (1949).
- Gregg v. Georgia, 428 U.S. 153 (1976).
- Grigsby v. Mabry, 483 F.Supp. 1372 (E.D. Ark. 1980) aff'd as modified, 637 F.2d (8th Cir. 1980).
- Grigsby v. Mabry, 569 F.Supp. 1273 (E.D. Ark. 1983), aff'd, 758 F.2d (8th Cir. 1985) (en banc).
- Haney, C. (1984). On the selection of capital juries: The biasing effects of the death-qualification process. *Law and Human Behavior*, 8, 121–132.
- Hans, V., & Vidmar, N. (1986). *Judging the jury*. New York: Plenum.
- Harris & Associates, Inc. (1971). Study No. 2016. Reported in White, W. The constitutional invalidity of convictions imposed by death-qualified juries. *Cornell Law Review*, 1973, 58, 1176–1220.
- Horowitz, I., & Seguin, D. (1986). The effects of bifurcation and death qualification on the guilt determination process. *Journal of Applied Social Psychology*, 16, 165–185.
- Hovey v. Superior Court, 28 Cal.3d 1, 616 F.2d 1301 (1980).
- Jurov, G. (1971). New data on the effect of a death-qualified jury on the guilt determination process. *Harvard Law Review*, 84, 567–611.
- Langer, E., & Abelson, R. P. (1972). The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology*, 24, 26–32.
- Lockhart v. McCree, (1986). 476 U.S. 162.
- Luginbuhl, J., Kadane, J. B., & Powers, T. (1990). Death scrupled jurors: They are not always more lenient. Technical Report, Department of Statistics, Carnegie Mellon University.
- Mauro, R. (1991). Tipping the scales toward death: The biasing effects of death qualification. In P. Suedfeld & P. Tetlock, *Psychology and social policy*, New York, Hemisphere Press.
- Monahan, J., & Walker, L. (1990). *Social science in law: Cases and materials*. Westbury, NY: Foundation Press.
- Moran, G., & Comfort, J. (1986). Neither "tentative" nor "fragmentary": Verdict preference of impaneled felony jurors as a function of attitude toward capital punishment. *Journal of Applied Psychology*, 71, 146–155.
- Mosteller, F., & Bush, R. R. (1954) Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology* (pp. 289–334). Cambridge, MA: Addison-Wesley.
- Osser, A. I., & Bernstein, B. (1968). The death-oriented jury shall live. Note, 1 *San Fernando Valley Law Review*, 253.
- Thompson, W. C. (1989). Death qualification after *Wainwright v. Witt* and *Lockhart v. McCree*. *Law and Human Behavior*, 13, 185–215.
- Wainwright v. Witt, 105 S.Ct. 884, (1985).
- White, W. S. (1987). *The death penalty in the eighties: An examination of the modern system of capital punishment*. Ann Arbor: University of Michigan Press.
- Winick, B. (1982). Prosecutorial peremptory challenges in capital cases: An empirical study and a constitutional analysis. *Michigan Law Review*, 81, 1–98.
- Witherspoon v. Illinois, 391 U.S. 510 (1968).
- Zeisel, H. (1968). Some data on juror attitudes towards capital punishment. Monograph. Center for Studies in Criminal Justice, University of Chicago Law School.