1988

Unpleasant Facts: The Supreme Court's Response to Empirical Research on Capital Punishment

Phoebe C. Ellsworth
University of Michigan Law School, pce@umich.edu

Publication Information & Recommended Citation

Ellsworth, Phoebe C. "Unpleasant Facts: The Supreme Court's Response to Empirical Research on Capital Punishment." In Challenging Capital Punishment: Legal and Social Science Approaches, edited by K. C. Haas and J. A. Inciardi, 177-211. Sage Criminal Justice System Annuals, 24. Newbury Park, Calif.: Sage Publications, 1988.

Volume 24. **Sage** Criminal Justice System Annuals

CHALLENGING CAPITAL PUNISHMENT

Legal and Social Science Approaches

Kenneth C. Haas
James A. Inciardi
Editors

1988

SAGE PUBLICATIONS
The Publishers of Professional Social Science
Newbury Park Beverly Hills London New Delhi

For information address:

SAGE Publications, Inc.
2111 West Hillcrest Drive
Newbury Park, California 91320

SAGE Publications Inc.
275 South Beverly Drive
Beverly Hills
California 90212

SAGE Publications Ltd.
28 Banner Street
London EC1Y 8QE
England

SAGE PUBLICATIONS India Pvt. Ltd.
M-32 Market
Greater Kailash I
New Delhi 110 048 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Challenging capital punishment : legal and social science approachs / edited by Kenneth C. Haas and James A. Inciardi.
p. cm. — (Sage criminal justice systems annual ; v. 24)
Bibliography: p.
Includes index.
ISBN 0-8039-2909-9 : ISBN 0-8039-2910-2 (pbk.)
1. Capital punishment—United States. I. Haas, Kenneth C. II. Inciardi, James A. III. Series: Sage criminal justice systems annuals ; v. 24.
KF9227.C2C46 1988 87-32092
364.6'6'0973—dc19 CIP

FIRST PRINTING 1988

Chapter 7

UNPLEASANT FACTS
The Supreme Court's Response to Empirical Research on Capital Punishment

PHOEBE C. ELLSWORTH

Slowly at first, and then with accelerating frequency, the courts have begun to examine, consider, and sometimes even require empirical data. From 1960 to 1981, for example, use of the terms "statistics" and "statistical" in Federal District and Circuit Court opinions increased by almost 15 times.[1] Of course, citation rates indicate only that a topic is considered worthy of mention, not that it is taken seriously, or even understood. Nonetheless, in a number of areas, such as jury composition and employment discrimination, the courts have come to rely on empirical data as a matter of course.

In the last 25 years, empirical research has been central to most major challenges to the constitutionality of capital punishment. The research involved has covered an enormous range of methods, from surveys to simulations, from econometric analyses to laboratory experiments, and the presentation of the research to the courts has generally been both comprehensive and sophisticated. There is probably no other area of

AUTHOR'S NOTE: *This chapter was written while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. I am grateful to the John D. and Catherine T. MacArthur Foundation and to the James McKeen Cattell Fund for financial support, and to the Center Staff and Fellows for time, space, unsurpassable services, and intellectual stimulation. Joseph B. Kadane and Samuel R. Gross read the first draft of this chapter and made it better.*

criminal law in which the Supreme Court has been faced with such well-organized and wide-ranging empirical demonstrations. It is quite likely that comparable empirical data about some other issue would be persuasive to the Court; when the issue is the death penalty, however, the Court is not persuaded. In case after case, the majority of the Justices have been faced with empirical research that supports an outcome they do not want.

Three major empirical questions have been brought before the Court in relation to capital punishment. The first is the issue of deterrence: Is the death penalty more effective than life imprisonment as a deterrent to murder? The second is the issue of discrimination: Are decisions about which criminals should be executed and which should be allowed to live based in part on the race of the person accused or on the race of the victim? The third is the issue of the fairness of capital juries: Does the common practice of removing strong opponents of capital punishment from the jury create juries that are biased toward a guilty verdict? The Court has dealt with the data on each of these issues somewhat differently—sometimes by calling them inconclusive, sometimes by calling them irrelevant, and sometimes by evading them. I shall examine the Court's use of the data presented on each of these three questions, the first two briefly, and the third at greater length.

DETERRENCE

Most people who favor the death penalty believe that it deters murderers (Ellsworth and Ross, 1983). It seems commonsensical that a person contemplating murder would be deterred by the threat of death. If so, then the existence of a death penalty is morally justified because it saves innocent lives. Opponents of the death penalty reason differently. They argue that few homicides are the product of such rational calculation. Most murderers do not stop to weigh the costs and benefits of their act. Those who do may figure (correctly) that the odds of execution are vanishingly small, so it is worth the risk. At any rate the likelihood of getting killed at the time of committing the murder is considerably greater than the likelihood of getting killed after conviction. Also, an argument for a powerful deterrent effect of capital punishment requires that it be a significantly more effective deterrent than life imprisonment. That is, the hypothetical person contemplating murder would go ahead with it if he considered the prospect of spending

the rest of his life in prison, but not if he considered the prospect of execution.[2]

Until 1974, the empirical research on the deterrent effects of the death penalty was completely consistent: No researcher claimed to have found a significant deterrent effect. The research was dominated by the work of Thorsten Sellin (1959, 1980), who had carried out a comprehensive series of studies covering many different times and places, and using a variety of complementary research designs. He examined neighboring states with and without the death penalty, and found that homicide rates were no lower in the death penalty states (in fact they were usually slightly higher). Over time the levels of homicide in contiguous states tended to rise and fall in synchrony, providing some indication that the states were in fact comparable, that similar forces affected the homicide rate. He also examined states and countries that had a death penalty and then abolished it to see if the homicide levels would rise; and he examined states that instituted a death penalty after a period without it to see if the number of homicides would fall. The death penalty had no noticeable effect in either case. He looked at crimes that are especially likely to be punished by death, such as killing police officers, killing for hire, and killing by prisoners serving life sentences, and again found no differences between the rates of these crimes in jurisdictions with and without a death penalty. One of the great beauties of Sellin's work is that each study answers a question left unanswered or compensates for a methodological shortcoming in the previous works, in a classic demonstration of the incremental elegance of converging research methods. Numerous other researchers, working independently, reached the same conclusion: The death penalty has no detectable long-term effect on homicide. It is not a more effective deterrent than life imprisonment (see, Lempert, 1981; Dike, 1982, for comprehensive reviews).

Another possible type of deterrence might be called "immediate deterrence." Maybe deterrent effects exist, but are very short-lived, so that right after an execution when the death penalty is particularly salient in the mind of the potential murderer, homicide rates will drop. Although there is considerably less research on this topic, there is no evidence whatsoever of such deterrence. In fact there is some evidence that executions stimulate potential killers to kill sooner (Dann, 1935; Bowers and Pierce, 1980).

In 1972, in *Furman v. Georgia,* the constitutionality of the nation's death penalty laws became an open question. The Court held that all the capital punishment statutes then in effect violated the Eighth Amend-

ment prohibition against cruel and unusual punishment, but stopped short of saying that capital punishment itself was unconstitutional. Most of the basic empirical questions that have figured in subsequent cases were raised in the opinions in *Furman,* but none were decided. Each of the Justices wrote an opinion detailing the considerations that he felt were important in deciding whether the death penalty was cruel and unusual. Instead of a coherent majority position, there were five separate one-vote opinions. Although deterrence was mentioned in several of the *Furman* opinions, only Justice Marshall explicitly cited the research, concluding that it proved that the death penalty was excessive, because its deterrent function could be served just as well by other penalties. Chief Justice Burger stated that no one had definitely proven that the death penalty was not a superior deterrent. However, in his view, empirical research on the issue was of no consequence, because deciding that sort of factual issue was not necessary or appropriate in interpreting the Eighth Amendment. Justice White accepted the argument that current death penalty laws had little deterrent effect (without citing studies), but reasoned that they were ineffective because hardly anyone was actually executed.

Although the social science research on deterrence was not discussed in any of the *Furman* opinions besides Justice Marshall's, it was probably influential in preventing any of the Justices from maintaining the simplistic common sense argument that the threat of death must be an effective deterrent, and therefore capital punishment is acceptable because it saves lives. The influence of empirical data cannot be completely evaluated by focusing only on what is said; it is also important to consider what might have been said had there been no data, or different data.

Since *Furman* had not said that capital punishment itself was unconstitutional, state legislatures hurried to enact death penalty laws that would pass constitutional muster. By 1976, when the next major capital punishment case was decided, 35 states had enacted new legislation. It was generally assumed that the main reason the pre-*Furman* laws were declared unconstitutional was that they permitted such unbridled discretion that there was no principled or systematic way of distinguishing among those who were sentenced to death and those who were not. Thus the *arbitrariness* of the death penalty was the central issue, but the issue of deterrence was also raised.

By 1976, however, the empirical record had also changed. Against the unanimity of the pre-*Furman* research on deterrence, a single new study

was presented as part of the solicitor general's brief, purporting to show that executions deterred homicide. The study, by Isaac Ehrlich (1975) used complex statistical analysis and economic theory, and found a small but significant deterrent effect. Unlike Sellin's research, which could be readily understood by any graduate student or Supreme Court Justice who chose to read and consider it thoughtfully, Ehrlich's work could not easily be evaluated without specialized expertise. The brief against capital punishment included several equally complex analyses by other experts discrediting Ehrlich's work. By now it has been established that Ehrlich's claims were unfounded. The research record still shows no evidence that the death penalty is a more effective deterrent than life imprisonment (see, Dike, 1982; Lempert, 1981; Blumstein et al., 1986, for reviews of the empirical evidence).

In fact, most of the important criticisms of Ehrlich's work were before the Supreme Court in 1976, along with Ehrlich's study. The briefs on both sides devoted considerable space to equations, technical discussions, and theoretical and methodological disputes. Yet the deterrence question is mentioned in only one of the five cases that the Court considered in deciding whether capital punishment was constitutional, and even in that one it receives only the most superficial attention. Justice Stewart, in his lead opinion in *Gregg v. Georgia*, concluded that the evidence on deterrence is ambiguous, presumably because there was disagreement among the researchers: On the one hand, Ehrlich claimed to have found a deterrent effect; on the other hand, everyone else claimed not to, whether they were analyzing their own data or Ehrlich's. Stewart goes on, illogically, to argue that since the research record is inconclusive, the death penalty must be a significant deterrent to some people, and therefore that capital punishment is not without justification. Since 1976, research on deterrence has not played a major role in any of the numerous Supreme Court decisions on the death penalty.

Had Ehrlich's study not existed, or had its inadequacies been recognized, at least some of the Justices might have felt compelled to deal with the massive body of scientific research indicating that the death penalty could not be justified on the grounds that it prevents murders from occurring, because it does not.

Recognizing the inadequacies of Ehrlich's study would not have been an easy task for a Supreme Court Justice. Statistical analysis was not a form of reasoning that was familiar to the Justices, and the presentations in the briefs were quite abstruse; nonetheless, given a real willingness to try to understand the data, the task would not have been impossible.

Indeed, Justice Marshall's dissenting opinion contains an excellent critical analysis of the research, concluding that no deterrent effect had been found by Ehrlich or anyone else. It still stands as one of the most lucid analyses of complex empirical data to be found in a Supreme Court opinion.

The Justices joining the lead opinion chose a much easier method of dealing with the social science data, simply noting that there was disagreement in the scientific community, and apparently deciding, therefore, that such an "inconclusive" record need not be considered at all. Behind this is a tacit requirement of complete consensus within the social science community before its research on an issue even need be *considered,* a standard far beyond what is required for any other sort of evidence. The issue of deterrence was erased from serious consideration by a *single* dissenting voice from the research community. It is distressing to think that the most influential thing a social scientist can do is to give the Court an excuse to ignore some body of research.

If a similar body of research had been presented on a different issue, such as the relative deterrent efficacy of different sanctions in some area of administrative law, the Court's reaction probably would have been different. But when the death penalty is involved, lay people reason less rationally (Ellsworth and Ross, 1983), and so do Supreme Court Justices. When the data compel a conclusion that might threaten the institution of capital punishment, something must be done about the data. In the case of deterrence, the existence of one flimsy study contradicting a mass of consistent research allowed the Court to avoid considering the evidence on the grounds that "experts disagree."

RACIAL DISCRIMINATION

Although advocates of the death penalty readily incorporated Ehrlich's study into their arguments for capital punishment (e.g., van den Haag, 1975, and the brief for petitioner in *Gregg*), evidence of deterrence is not essential to a principled position in favor of the death penalty. An advocate might argue that even if the utilitarian functions of capital punishment could be served equally well by sentencing murderers to life in prison, the death penalty is justified because it is *just.* Those who commit the very worst crimes deserve the very worst penalty.

Evidence of racial discrimination in the imposition of the death penalty is disastrous to the argument that capital punishment is serving

the ends of justice, because it means that many of those who actually receive the worst penalty are not those who have committed the worst crimes, but those who have the darkest complexions. The deterrence argument is that the death penalty is cruel and unusual because it is excessive—other penalties would work as well; the discrimination argument is that the death penalty is cruel and unusual because it is unfair. And the basis for this unfairness is one that the Court has found completely unacceptable in numerous other contexts. Being born black is no fault of the person, and it generally means a lifetime of slights, risks, and disadvantages, most of which are beyond the reach of the law to correct. The Supreme Court has been stern in combating racial discrimination in other contexts, such as racially imbalanced juries (*Strauder v. West Virginia,* 1880; *Batson v. Kentucky,* 1986), school segregation (*Brown v. Board of Education,* 1954), and employment discrimination (*Griggs v. Duke Power and Light Co.,* 1971). If the death penalty discriminates on the basis of race, not only the Eighth Amendment prohibition against cruel and unusual punishment is involved, but also the Equal Protection clause of the Fourteenth Amendment.

Like research on deterrence, research on discrimination in the application of the death penalty has a long history. In the nineteenth century, several state laws made certain crimes punishable by death if the offender was black, but not if he was white (Bowers, 1974). By the twentieth century the laws were gone, but most studies still showed that when arrested, blacks were more likely than whites to be indicted for capital crimes; when charged with capital crimes, blacks were more likely to be convicted; when convicted, blacks were more likely to be sentenced to death; and when sentenced to death, blacks were more likely to be executed (see Bowers, 1974; Dike, 1982). Racial effects were substantial across a variety of potentially criminal acts, and enormous for the crime of rape. Over the last hundred years hardly anyone has been executed for rape except in the southern and border states, and hardly any white man has been executed for rape anywhere in the country (Bowers, 1984). The most striking data were presented by Wolfgang and Reidel (1973) in a study of rape convictions and executions in 11 southern states from 1945 to 1965. They found that death sentences for rape had virtually disappeared except in cases in which the rapist was a black man and the victim was white: Black men accused of raping white women were 18 times more likely to sentenced to death than white rapists or black men accused of raping black

women. Similar trends have been found in studies of death sentences for homicide (e.g., Zimring et al., 1976), although the disparities are less extreme.

Wolfgang and Reidel's data were considered and rejected by the Eighth Circuit in the case of *Maxwell v. Bishop* (1968) on the grounds that very few of the cases came from the jurisdiction in which Maxwell was tried; some factor other than race or the other variables that Wolfgang and Reidel had considered might have accounted for the disparity; and the study did not show that Maxwell's particular jury had been motivated by racial prejudice when they sentenced him to death. The Supreme Court vacated the judgment in *Maxwell* on other grounds, without touching the discrimination issue.

In 1972 the data on racial discrimination came before the Court again in *Furman*, as part of the more general argument that the death penalty was unconstitutional because the people who were executed were not the people who most deserved to die. If there was any systematic difference between those who were killed and those who were allowed to live, it was an unacceptable one, such as income, class, or, most obviously, race.

What little consensus there was in the *Furman* case seemed to converge on the conclusion that the imposition of the death penalty laws then in effect produced no *legitimate* distinction between the few who were selected to die and the many who were not. At best, death sentences were arbitrary and capricious. Most of the Justices shied away from addressing Wolfgang and Reidel's study directly, although it is difficult to imagine that it had no effect on their decision that the imposition of the death penalty under current laws was unfair. Only Justice Douglas and Justice Marshall explicitly asserted that the death penalty discriminated against the poor and the black. Justice Stewart based his opinion on arbitrariness. Death sentences were "cruel and unusual in the same way that being struck by lightning is cruel and unusual," although he also acknowledged that if there were any factor that particularly characterized those sentenced to death, it was the "constitutionally impermissible basis of race" (*Furman*, 1972, p. 310).

Like the research on deterrence, the research on discrimination was barely mentioned in the *Furman* opinions, but it was probably influential. The Court could have concluded that those who were sentenced to death were those who had committed the most heinous crimes. The Justices could have concluded that those who were sentenced to death were in the same groups that were the victims of

discrimination in other social contexts. Or they could, as they did, conclude that there was no valid basis for distinguishing between those who were sentenced to death and those who were not: The system was arbitrary. The social science data may have persuaded a majority away from the first position, but some Justices may have chosen to emphasize the more neutral problem of arbitrariness rather than endorsing the much more inflammatory idea that black lives had been treated as more dispensable than white lives. After all, the immediate practical outcome was the same. Current death penalty laws were struck down, and current death sentences were canceled.

In *Gregg v. Georgia* (1976) and its companion cases, arbitrariness was the central issue. State laws that provided for mandatory death sentences for certain crimes were struck down, whereas state laws that provided standards by which the unbridled discretion rejected in *Furman* could be "suitably directed and limited" were upheld. Racial discrimination was not mentioned. Presumably a law that prevented arbitrariness by limiting the aggravating factors that could be considered in imposing the death penalty would also prevent racial discrimination. Once again, it is quite likely that the research on racial discrimination was disturbing to the majority Justices, but that they felt that in endorsing the guided discretion statutes they could solve the problem without acknowledging it.

In 1977, a case came before the Supreme Court that many observers felt would force the Court to deal with the issue of racial discrimination directly. It was the case of *Coker v. Georgia,* raising the issue of the constitutionality of capital punishment for the crime of rape. The evidence, as we have seen, indicated that the death penalty for rape had long since been practically abandoned, except for black men who raped white women. Racial prejudice seemed the unavoidable conclusion.

The Court in *Coker* did avoid reaching that conclusion, however. In fact it avoided any mention of the research on racial patterns in capital sentencing for rape. Instead, the Court held that capital punishment for rape was unconstitutional because it was grossly disproportionate to the severity of the crime and therefore excessive. Again, the Court reached the conclusion that would be compelled by the data on racial discrimination, but reached it on other grounds. Curiously enough, the Court actually based its finding that the death penalty was disproportionate on empirical data, of a sort. Justice White, in the lead opinion, argued that public attitudes toward a penalty should be considered in deciding whether that penalty was excessive for a particular crime. As a means of

ascertaining public attitudes, he examined the history of the death penalty for rape, pointing out that few national or international jurisdictions still punished rape by death, and that although after the *Furman* decision many states reenacted death penalty statutes for murder, they did not do so for rape. He also pointed out that even when the death penalty was allowed for rape, juries hardly ever imposed it, opting for a lesser penalty in 9 out of 10 cases, thus indicating that the public felt it was excessive. (Of course he avoided all mention of the racial factors distinguishing those rare cases in which the juries did sentence rapists to death. He also neglected to mention that the jury chooses a penalty less than death in 8 out of 10 murder cases.)[3] He even cited public opinion poll data. In sum, it is hard to resist the conclusion that the Court went out of its way to avoid basing its decision on the powerful empirical evidence of racial discrimination, even to the extent of citing much more fragmentary empirical data on another issue.

After 1972 state legislatures that favored the death penalty drafted new statutes designed to protect against the arbitrary and discriminatory imposition of capital punishment that had been held unconstitutional in *Furman*. In *Gregg v. Georgia* (1976) the Court found acceptable Georgia's new statute, which provided for a bifurcated trial with guilt decided at the first phase, followed by a separate sentencing phase in which new evidence might be introduced to help the jury decide whether the defendant should live or die. Factors that the jurors were permitted to consider as aggravating or mitigating the crime were enumerated, and provision was made for the Georgia Supreme Court to review all cases to make sure that the system was working properly—that the death penalty was reserved for the most aggravated cases. Similar statutes in Texas and Florida were upheld in the companion cases to *Gregg* (*Jurek v. Texas,* 1976; *Proffitt v. Florida*, 1976), and, by implication, all statutes employing comparable methods of limiting the jury's discretion were held to be constitutional.

Because the laws had changed, pre-*Furman* data were no longer sufficient to demonstrate racial discrimination. The question now was whether the new guided discretion statutes did what they were intended: to prevent arbitrariness and discrimination. A crude examination of America's death row population suggested no improvement: In 1971, just before the *Furman* decision, 53% of the people on death row were nonwhite; in 1978, 62% were nonwhite (Reidel, 1976). Raw figures, however, were no longer enough. Research on discrimination had become substantially more sophisticated, and the courts were becoming much more accustomed to complex statistical analyses in other kinds of

discrimination cases. However overwhelming the raw figures, litigators assumed that they would be insufficient to prove discrimination. They had to prove not only that blacks were disproportionately sentenced to death, but that they were sentenced to death *because they were black,* not because they had committed more aggravated murders. If they could not prove that race was the reason, they had at least to prove that no other *legitimate* reason distinguished the blacks on death row from murderers who were sentenced to life imprisonment.

By 1980, eight years after the *Furman* decision, studies of the racial consequences of the post-*Furman* guided discretion statutes began to appear. Bowers and Pierce (1980) found that in Georgia, Florida, Texas, and Ohio (these states accounted for about 70% of the death sentences in the time period studied), a familiar pattern emerged: "Black killers and the killers of whites were more likely to receive the death penalty in all four states, with black killers of white victims overwhelmingly more likely to receive the death sentence" (Dike, 1982, p. 49; see also Zeisel, 1981). With increasingly sophisticated statistical controls, later studies found that the effect of the defendant's race could not conclusively be shown to have an effect independent of the effects of legitimate aggravating factors such as committing murder in the course of another felony or killing a stranger. This means that it is impossible to tell whether the reason blacks are more likely than whites to be sentenced to death is that they are black. The race-of-victim effect, however, continues to show up strongly in the best-controlled studies; killing a white person is much more likely to result in a death sentence than is killing a black person (Radelet, 1981 [Florida]; Radelet and Pierce, 1985 [Florida capital charges]; Jacoby and Paternoster, 1982 [South Carolina capital charges]; Gross and Mauro, 1984 [eight states]). In one study the researchers collected data on over 400 variables in over a thousand Georgia homicide prosecutions and found that nothing could explain away the discrimination against defendants who killed white people (Baldus et al., 1983). Some 10 years later, it is clear that the guided discretion statutes endorsed in *Gregg v. Georgia* have not succeeded in preventing capital sentencing from being contaminated by the "constitutionally impermissible basis of race."

The issue and the new data came before the Court in the recent case of *McCleskey v. Kemp* (1987). The social science research on race-of-victim effects were the core of the case, and this time there was no way the Court could avoid facing the problem of racial discrimination in capital sentencing.

Writing for the majority, Justice Powell made a faint-hearted attempt to question the conclusiveness of the research by arguing that the correlational data presented could not actually *prove* that race was the cause of the sentencing disparities. However, the Baldus study had controlled for over 200 other variables and found that none of them, alone or in combination, could explain the pattern of racial discrimination, and neither Justice Powell nor anyone else was able to come up with any other explanation for the racial effects. History, common sense, and the data all overwhelmingly supported the simple, obvious explanation. The racial effects were due to race. The majority, however, repeatedly referred to the correlation of death sentences with race as "unexplained," for example, "where the discretion that is fundamental to our criminal process is involved, we decline to assume that what is unexplained is invidious" (*McCleskey v. Kemp*, 107 S. Ct. 1756, 1778 [1987]).

The opinion is also written to make the death penalty invulnerable to any future statistical challenges based on race, by holding that proof of racial discrimination requires a demonstration of conscious, intentional racial discrimination in the drafting of the law or the decision in the particular case. That is, in order to succeed with a claim of racial discrimination, a defendant must prove either (1) "that the decision makers in *his* case acted with discriminatory purpose" [emphasis in original] (107 S. Ct. at 1766), or (2) "that the Georgia legislature enacted or maintained the death penalty statute *because of* an anticipated racially discriminatory effect" [emphasis in original] (107 S. Ct. at 1769).

A juror or a prosecutor must come forth and admit that McCleskey was sentenced to death because he was black and he killed a white man, a highly unlikely event because neither jurors nor prosecutors are asked to account for their decisions, and even if they were, they would almost always be unable or unwilling to say that race was a factor. Or, a legislature must come forth and admit that the capital punishment statute was designed to produce racial bias. Otherwise, the Equal Protection clause does not apply.

Once again it is clear that the majority Justices (five of them, in *McCleskey*) value the death penalty so highly that they are willing to ignore the inescapable implications of the social science research and to undermine fundamental constitutional guarantees. The majority admitted that the system for deciding who lives and who dies is imperfect, and conceded that the evidence of discrimination found in the Baldus study would have been persuasive if the case had involved discrimination in

jury selection or employment. Traditionally, the Court has assumed that the decision to impose the death penalty should be safeguarded with higher standards of fairness than other sorts of legal decisions. In *McCleskey*, the Court has reversed this traditional attitude. Empirical evidence that would be sufficient to protect McCleskey's job was not considered sufficient to save his life.

Finally, the majority spent some time describing the sanctity of trial by jury, and their reluctance to tamper with the collective discretion of a representative jury of one's peers. The jury, "representative of a criminal defendant's community, assures a 'diffused impartiality,' in the jury's task of 'expressing the conscience of the community on the ultimate question of life or death'" (107 S. Ct. at 1776, citations omitted). However, as we shall see in the next section, the Court had already held that juries in capital cases are not required to be as fair or as representative of community opinion as juries in other cases.

DEATH-QUALIFIED JURIES

The issue of death qualification also raises questions about the fairness of the administration of capital punishment laws, but it is limited to a procedure—the refusal to allow citizens who adamantly oppose the death penalty to serve as jurors in capital cases. Jurors who state that they cannot consider imposing the death penalty are not allowed to participate in the decision between death and life imprisonment, and hardly anyone has seriously questioned this practice.[4] However, these opponents of the death penalty are also excluded from the jury that decides whether a person charged with a capital crime is guilty or innocent of the crime, even when they state that they would be entirely fair and impartial in weighing the evidence on guilt or innocence. Thus in capital cases, unlike all other cases, the jury that decides guilt or innocence is made up exclusively of people who would be willing to sentence a person to death. The empirical question is: Is such a "death-qualified" jury more likely than an ordinary jury to decide that the defendant is guilty?

Most trial attorneys assume the biasing effects of death qualification as a matter of common sense. Jurors who favor executions tend to favor the arguments of the prosecution, whereas jurors who reject the death penalty are more sympathetic to the defense. Even prosecutors occasionally admit to using death qualification in order to get a jury that is

more likely to convict (Oberer, 1961). But lawyers' intuitions are not evidence. The issue first came before the U.S. Supreme Court in 1968 in the case of *Witherspoon v. Illinois* (1968). Nearly half of the jurors called to hear Witherspoon's case had been excused because of their "conscientious scruples" against the death penalty, and Witherspoon argued that his jury was therefore biased toward conviction. He supported his case with preliminary reports of three unpublished studies showing a relationship between attitudes toward capital punishment and propensity to convict. The Court rejected the empirical data as "too tentative and fragmentary" to demonstrate a bias toward conviction, and it held that a modified form of death qualification, excluding only the most adamant opponents of the death penalty, was constitutionally permissible. However, the Court acknowledged that the question was an empirical one, that its decision was a provisional one, made "in the light of the presently available information" (*Witherspoon v. Illinois,* 1968, p. 518), and that future research might demonstrate that Witherspoon's claims were correct.

By the time the issue came back to the Supreme Court in 1968, there were 15 empirical studies in the record. They had been conducted over a span of three decades, using a variety of samples from different areas of the country, a variety of questions designed to identify the excluded group, and a variety of different research methods. Attitude surveys of random samples of the public were used to discover the size and racial composition of the excluded group and to examine the correlation between attitudes toward the death penalty and predispositions to favor the prosecution or defense. Simulations were used for a direct examination of the relation between death penalty attitudes and verdicts across a range of cases. Interviews with actual jurors were used to establish that attitudes toward the death penalty predicted verdicts in real cases. Another study examined the biasing effects of the death-qualifying voir dire itself, and still others looked at the effects of death qualification on the quality of jury deliberation. Without exception, the studies found that death-qualified jurors were more favorable to the prosecution and more likely to vote guilty than the citizens who are excluded from the jury.

For many reasons, death qualification is a particularly interesting context in which to examine the Court's response to empirical data related to capital punishment. First, the empirical question was posed by the Court itself, in *Witherspoon.* Thus the Court ought to be unusually receptive to empirical research specifically designed to answer

that question. And the Court not only posed the question, but stated that the holding in *Witherspoon* might be reversed if new evidence established that death-qualified juries were "less than neutral with respect to *guilt*" (*Witherspoon v. Illinois,* 1968, p. 520, n. 18, emphasis in original).

Second, the constitutional issue was a Sixth Amendment issue, one involving the fairness of the jury, rather than an Eighth Amendment issue challenging the constitutionality of capital punishment per se. In the past, the Court had shown considerable willingness to consider social science data in cases involving jury composition (cf. *Williams v. Florida,* (1968); *Colgrove v. Battin,* (1973); *Ballew v. Georgia,* (1978).[5]

Third, the studies themselves were easy to understand. In most of the studies two groups of people were compared, and the main data took the form of percentages of each group favoring proprosecution attitudes or voting guilty. The basic argument did not hinge on complex regression equations or mathematical models. And of course the data agreed with the "common sense" of most people familiar with the criminal justice system.

Fourth, the social science data were immune from the criticism that the data were introduced into the decision making process too late to have received careful critical evaluation from the lower courts. This criticism was clearly stated by Justice Powell in his concurring opinion in *Ballew v. Georgia* (1978):

> I have reservations as to the wisdom—as well as the necessity—of Mr. Justice Blackmun's heavy reliance on numerology derived from statistical studies. Moreover, neither the validity nor the methodology employed by the studies cited was subjected to the traditional testing mechanisms of the adversary process. The studies relied on merely represent unexamined findings of persons interested in the jury system [p. 246].

The new data on death qualification had been introduced in an evidentiary hearing that lasted several weeks, with testimony by five expert witnesses for the defense and two for the prosecution (*People v. Moore,* August-September, 1979). The California Supreme Court, in *Hovey v. Superior Court* (1980), carried out an extraordinarily thorough evaluation of the empirical record, perhaps the most knowledgeable analysis of social science data yet to appear in an appellate court opinion. The California Supreme Court concluded that excluding

jurors according to the *Witherspoon* standard did create juries that were unconstitutionally biased toward guilty verdicts, but questioned the applicability of the research to California juries, because in California, jurors at the opposite end of the attitudinal spectrum could also be excluded for cause. These are the jurors who state that they would be unable to consider a penalty less than death for anyone convicted of a potential capital crime, regardless of the circumstances.

Subsequent research indicated that the number of people so bloodthirsty that they would automatically vote for death without regard for the evidence is minuscule—only 1% of the population (Louis Harris, 1981), far too few to correct the bias toward guilty verdicts created by death qualification (Kadane, 1984). In 1981 an evidentiary hearing was conducted before the U.S. District Court for the Eastern District of Arkansas, when all the *Hovey* research was presented in addition to the new evidence on "automatic death penalty" jurors. Again, experts testified on both sides, and the *Hovey* record was introduced in evidence. Again, the court evaluated the research very carefully and concluded that the practice of death qualification created juries that were unconstitutionally biased against the defendant (*Grigsby v. Mabry,* 1983). The state appealed but the Eighth Circuit affirmed (*Grigsby v. Mabry,* 1985). The *Grigsby* record, which included the *Hovey* record, reached the U.S. Supreme Court in the case of *Lockhart v. McCree* (1986).[6] Thus the data on death qualification had abundantly fulfilled Justice Powell's requirement that they be subjected to the "traditional testing mechanisms of the adversary system," having been scrutinized in 26 days of evidentiary hearings and evaluated in 145 pages of appellate court opinions.

Fifth, the Justices could not avoid evaluating the data, as they had avoided the data on deterrence, by concluding that there was a lack of consensus among the experts. Although the state tried to argue that the social scientists' use of different research *methods* revealed a lack of consensus, all 15 of the studies reached the same *result*. As the American Psychological Association (APA) argued in its amicus brief, "the use of diverse subjects, stimulus materials, and empirical methods does not reveal a 'lack of consensus' but comports fully with the goal of 'generalization,' the accepted rubric for evaluating how far beyond the specific facts of any one particular study one can apply its findings." Few bodies of social science research have shown such consistency across times, places, groups, and methods. The only "study" that failed to find a difference in conviction rates is one that simply compared conviction

rates in murder trials with those in robbery and burglary trials. No prosecution expert was willing to present this study as evidence, and no lower court found it worthy of consideration. In general, the prosecution experts made minor methodological criticisms and conceded many of the defense's empirical claims. No studies were introduced that contradicted those claims; the research record was uncontroverted. The state's brief attacked the research primarily with vague global statements about the "pseudoscientific" nature of social science research in general.

Finally, it would be very difficult for the Court to do what it did with the issue of racial discrimination in the pre-*McCleskey* cases—make a decision consistent with the research without mentioning the research itself. The whole case was built on the empirical proposition, raised in *Witherspoon*, that death qualification creates juries that are unlike all other juries in that they are "less than neutral with respect to guilt."[7]

However, by 1986 there were over 1,500 people on death rows across the country, most of whom had been convicted by death-qualified juries. There were many more people serving life sentences who had also been convicted by death-qualified juries. If the Court admitted that death qualification unfairly prejudices the jury against capital defendants, what would happen to these people? A decision in line with the evidence would raise enormous legal and moral questions about retroactivity. It seems quite possible, even probable, that if the practice of death qualification had not existed before, and some state had attempted to introduce it as a new restriction on the normal process of jury selection, the Court would have found the data persuasive, and declared the procedure unconstitutional. But the political and practical consequences of such a decision in 1986 were extremely unpalatable. Thus, as in *McCleskey*, the Court was faced with a substantial, consistent, and highly persuasive body of data that pointed to a conclusion opposite to the one the majority wanted to reach.

Given that the Court was determined to come out in favor of death qualification, its decision was bound to be disappointing to the social science research community. The actual decision was worse than disappointing; it was lamentable. Justice Rehnquist, writing for the majority, first attacked the research in ways that suggested that the majority Justices had either not understood it or not read it, or that they just didn't care. He then declared that it didn't matter how compelling the data might be, because it is constitutionally permissible to try capital cases before juries that are biased toward guilty verdicts. In short, "we don't believe the data, but if we did it wouldn't matter."

In the discussion that follows, I shall review the *Lockhart* opinion's criticisms of the empirical data. In most instances, the same issues had been reviewed in the *Grigsby* and *Hovey* opinions, and, when appropriate, I shall compare the three courts' methods of dealing with them.

1. *The data are insufficient.* In 1968, the Court in *Witherspoon* rejected the existing studies as "too tentative and fragmentary to establish that jurors not opposed to the death penalty tend to favor the prosecution in the determination of guilt" (p. 527). Most social scientists would have agreed with this assessment. The Court had before it incomplete, preliminary versions of three studies. Most social scientists would recognize, however, that as new investigations accumulated, all confirming the original conclusions, the research became less tentative and fragmentary, and correspondingly more definite and complete. Courts, however, have sometimes seemed to regard the phrase "tentative and fragmentary" as having a binding precedential force on all research that might be done on death qualification (Gross, 1984).

In *Lockhart*, Justice Rehnquist also concluded that the data were insufficient to form the basis of a constitutional rule, but he did so in a slightly more sophisticated manner. The 15 studies were lined up, and each of them was found to have a flaw; each time a flaw was discovered, the study was eliminated from the set until only one study was left. "Surely," concluded Justice Rehnquist, "a 'per se constitutional rule' as far-reaching as the one McCree proposes should not be based on the results of the lone study that avoids this fundamental flaw" (*Lockhart,* 1986, p. 1764).

The first "flaw," accounting for the elimination of 8 of the 15 studies, was that the studies "dealt solely with generalized attitudes and beliefs about the death penalty and other aspects of the criminal justice system" (*Lockhart* 1986, p. 1762). In fact, several of these studies measured attitudes about the death penalty precisely in the terms defined by *Witherspoon,* and in all of them the attitudes about "other aspects of the criminal justice system" directly reflected a proprosecution (and therefore proconviction) bias. Among these attitudes, for example, are the belief that defense attorneys are less trustworthy than prosecutors are, that defendants who do not testify are guilty, that the insanity defense is a loophole for guilty defendants, and that a confession should be considered even if the judge rules it inadmissible. The emergence of the same relation between death penalty and proprosecution attitudes in every study indicates that the exact wording of the question is not important. Wherever the line between death-qualified and excludable

citizens is drawn, the former group will hold attitudes more favorable to the prosecution. The relevance of studies involving attitudes toward the criminal justice system is strongly suggested by (1) the direct connection between the attitudes measured and the jurors' task, (2) the argument that the excluded jurors share a perspective that is different from that of the qualified jurors, and (3) the fact that the attitudinal studies and the more direct studies of conviction rates reach identical conclusions. Both the *Hovey* opinion and the *Grigsby* opinion explicitly considered the relevance of the attitudinal studies, and found that, *taken in conjunction with the other studies*, they were important. Here, as elsewhere, the *Lockhart* opinion chose to ignore the conclusions reached by the "traditional testing mechanisms of the adversary system."

One study (Haney, 1984) showed that the very process of questioning jurors intensively about their attitudes toward capital punishment before the trial suggested to the jurors that the defendant was probably guilty. Thus added to the bias caused by the composition of the jury was a bias caused by the experience of the death-qualifying *voir dire*. This study was dismissed from consideration because it "would not, standing alone, give rise to a constitutional violation," (*Lockhart*, 1986, p. 1763). Again the Court's logic seems to be that if each study is insufficient evidence on its own, then all 15 put together must also be insufficient. If one payment, standing alone, is insufficient to buy a house, then it should not be considered in conjunction with others in assessing the means of the purchaser.

Three studies were rejected because they were the studies that had been before the Court in *Witherspoon*, albeit in preliminary form: "It goes almost without saying that if these studies were 'too tentative and fragmentary' to make out a claim of constitutional error in 1968, the same studies, unchanged but for having aged some 18 years, are still insufficient to make out such a claim in this case" (*Lockhart*, 1986, p. 1763). In fact, the studies were not "unchanged"; preliminary drafts had become completed, fully-documented reports. The briefs and the lower court opinions were clear on this point. More important, the conclusions of the first study on a new research topic are nearly always "tentative." It takes replication to establish the truth and generality of a scientific finding and in turn to establish the validity of the early work. Having been replicated, the early conclusions were no longer tentative. The Supreme Court's assessment of the early research is analogous to claiming that Alexander Fleming's discovery of penicillin is of trivial medical importance because in 1930 the scientific community felt that

the prospects for using it to cure infections in humans were discouraging.

Two of the studies demonstrating conviction-proneness were rejected because they did not include jury deliberation and because they did not identify and exclude jurors who might admit to being unable to make a fair decision on the guilt of a capital defendant. No explanation was given for why deliberation is essential, and indeed, research on juries has consistently shown that the distribution of individual first-ballot verdicts is a very good predictor of the final jury verdict (Kalven and Zeisel, 1966; Hastie et al., 1983). Several of the studies rejected on other grounds did identify and exclude people who could not judge guilt fairly, and the overall bias was demonstrated in these studies as strongly or more strongly than in the studies that did not take this precaution.

The "flaws" identified in individual studies were in no instance fatal flaws. But more important, the one-by-one elimination of studies from a consistent body of research shows an ignorance of the principle of convergent validity. The idea that new scientific truths are proven by means of a single, perfect, definitive experiment is generally mistaken. Typically there are numerous sources of error and numerous confounding variables that must be controlled, and typically it is impossible to control them all in a single study. Thus the scientist must triangulate in on the truth by ruling out some alternative explanations in some experiments and other alternative explanations in other experiments until only one explanation is left that can account for the results of all the experiments. This is the method of convergent validation. If all the studies share the same shortcoming, then that shortcoming may be a possible alternative explanation that has not yet been ruled out. But if the experiments have different limitations, then none of the limitations can serve as an explanation for a result that is consistent across all the experiments.

For example, in order to show that death-qualified jurors vote guilty more often than the jurors who are normally excluded, it is important to demonstrate that in response to the *very same trial* more of the death-qualified jurors vote guilty. Since in the "real world" each case is tried only once, before a single jury, the only way to expose many juries to the very same trial is to conduct a simulation, as Cowan et al. (1984) did. Simulations can always be criticized, however, on the grounds that people may behave differently on real juries.

In this case, the argument might be that on real juries people's verdicts are not influenced by their attitudes toward the death penalty. Zeisel's (1968) study of actual jurors, though lacking the control of the

Cowan et al. study, showed that the correlation between death penalty attitudes and guilty votes is not restricted to simulated juries. The fact that both studies reached the same result shows that this result could not be due to any "flaw" unique to either study. If 10 studies were done to test the effectiveness of a new drug, five on samples of women and five on samples of men, and all showed that the drug cured the disease, the logic of the *Lockhart* opinion would compel the conclusion that we have no information about the drug's effectiveness, since five of the studies should be dropped from consideration because they included no women, and the other five because they included no men, leaving not a single study. The lower court opinions in *Hovey* and in *Grigsby* contained extensive discussions of the principle of convergent validation and the necessity of viewing the studies "as a whole, not in isolation" (*Hovey,* 1980, p. 1343). The APA brief in *Lockhart* contained a clear account of the value of convergent validation, as did the dissenting opinion:

> The chief strength of respondent's evidence lies in the essential unanimity of the results obtained by researchers using diverse subjects and varied methodologies. Even the Court's haphazard jibes cannot obscure the power of the array. (*Lockhart,* 1986, Marshall, J., dissenting, p. 1773).

Nonetheless, the majority managed to miss the point.

2. *No bias exists because none of the jurors who tried McCree was biased.*

Having completed its excursion into the unfamiliar realm of social science criticism, the Court returned to its own turf—the close analysis of the individual case—and focused on McCree's particular jury. Since the 12 people who served on McCree's jury had all said they could judge the case impartially, the Court concluded that it was therefore a neutral jury. The social scientists argued that across many trials it could be demonstrated that juries comprised only of death-qualified jurors were not neutral, because the proportion of guilty verdicts would be higher than it would be in juries that included the whole spectrum of death penalty attitudes. The probability of guilty verdicts is affected by the group that is not there—just as it would be if Catholics (or atheists) were never allowed on juries in abortion cases. Social scientists are trained to think in terms of comparison. Death-qualified juries are not biased or unbiased in some absolute sense; rather they are biased toward guilty

verdicts relative to the juries that try all other cases in American courts. Following the Court's reasoning, it would be constitutional to exclude all Republicans, or all people with a college education, or any other group, as long as 12 impartial jurors remained.

Although aggregate, comparative statistical reasoning is uncommon for courts, it is by no means beyond their capacity. In *Ballew v. Georgia* (1978) the Supreme Court considered a variety of empirical studies of jury size and concluded that juries composed of only five people were unconstitutional; the Court did not examine the competence of the particular group of five people who tried Ballew, but reached a general conclusion based on general data. More to the point, the Court in *Witherspoon* found that by excluding all opponents of the death penalty from the penalty decision, "the state crossed the line of neutrality. In its quest for a jury capable of imposing the death penalty, the state produced a jury uncommonly willing to condemn a man to die" (*Witherspoon,* 1968, p. 520-521). The fact that the 12 people who did try Witherspoon's case all said that they could be fair and impartial was irrelevant. The logic of the Court's decision on bias at the penalty phase of *Witherspoon* was exactly the same as the logic of the argument on bias at the guilt phase in McCree's case.

The *Lockhart* Court bolstered its faulty reasoning with two other spurious arguments. The majority opinion claimed that McCree was asking that every single jury be perfectly representative of the community, with "the proper number of Democrats and Republicans, young persons and old persons, white-collar executives and blue-collar laborers, and so on" (*Lockhart*, 1986, p. 1767). McCree's brief is quite explicit on this point, admitting that "no defendant can demand that all elements of the community be included on his or her jury," but arguing that no state has the "right to remove an entire class of jurors from every capital jury." (*McCree*, brief, p. 88). For example, if I were on trial, I could not claim unfairness if, by the luck of the draw, there were no female professors on my jury. But I could reasonably claim unfairness if, at the beginning of jury selection, the judge told all the female professors in the room that they were disqualified from serving in my case. The *Lockhart* Court confused these two very different positions.

Finally, the Court argued that there is no problem because McCree might have gotten the very same jury by chance, even if death qualification did not exist: "[I]t is hard for us to understand the logic of the argument that a given jury is unconstitutionally partial when it results from a state-ordained process, yet impartial when exactly the

same jury results from mere chance" (*Lockhart,* 1986, p. 1767). Leaving aside the fact that the likelihood that McCree would have gotten "exactly the same jury" by chance is minuscule, the idea that any practice is constitutional as long as any particular consequence of that practice could have occurred by chance is astounding. All-white juries are possible by chance; therefore blacks may be banned from jury service. A defendant might have confessed to a crime inadvertently; therefore defendants may be forced to confess. Or Mr. McCree could argue that the woman he was accused of killing might by chance have been struck by lightning that day, so why was his act illegal?

Again, the lower court opinions avoided these mistakes, recognizing that it was necessary to compare one type of jury with the other, not simply to examine the 12 jurors in a particular case:

> The issue is not whether non-death-qualified jurors are acquittal prone or death-qualification jurors are conviction prone. The real issue is whether a death-qualified jury is more prone to convict than the juries used in non-capital criminal cases—juries which include the full spectrum of attitudes and perspectives regarding capital punishment [*Grigsby v. Mabry,* 1985, p. 241, note 31].
>
> It can be confidently asserted that, over time, some persons accused of capital crimes will be convicted of offenses—and to a higher degree—who would not be so convicted if the jury were more representative of the populace [*Hovey v. Superior Court,* 1980, p. 1314, note 57].

3. *The excluded jurors are not a distinctive group.* The majority opinion held that death-qualified jurors were not a "distinctive group" whose elimination would impair jury representativeness. From the perspective of social science, the excluded jurors are obviously a distinctive group. They can easily be distinguished from other jurors on the basis of their answers to a few simple questions asked during the voir dire and in study after study their attitudes and behavior have been significantly different from those of other jurors.

In this case the obvious social science argument is insufficient to answer the legal question. Most previous cases on jury representativeness had dealt with the exclusion of blacks, women, or Mexican Americans, groups that have enjoyed special constitutional protection in recent years. Although one of the justifications for including these groups on juries was that the absence of their distinctive attitudes and

perspectives would result in a jury that did not reflect the common-sense judgment of the whole community, it was not the only justification. There was also the need to eliminate formal discrimination against these historically disadvantaged groups.

During oral argument in *Lockhart,* McCree's counsel was asked numerous questions about other attitudinal groups that might claim a constitutional right to representation on juries. Some Justices may have feared that the case could set a dangerous precedent. McCree's counsel argued first of all, that no other attitudinal group of fair and impartial jurors was automatically excluded from jury service, and second, that no other attitudes had been found to affect verdicts as powerfully as attitudes toward the death penalty. Judge Eisele, in the District Court opinion in *Grigsby v. Mabry* (1983) concluded that shared attitudes were an insufficient criterion to define a constitutionally "distinctive" group unless it could be shown that the presence or absence of those attitudes actually affected the functioning of the jury. He went on to conclude that the attitudinal exclusion created by death qualification had a substantial adverse effect on the functioning of the jury and thus was unconstitutional. He also pointed out that the representation of unimpeachably cognizable groups was impaired by the process of death qualification because blacks and women are more likely to be excluded are than white males.

Justice Rehnquist chose not to follow this reasoning, holding for the majority in *Lockhart* that a group defined solely in terms of shared attitudes is not a constitutionally distinctive group that needs to be considered in assessing jury representativeness. "Unlike blacks, women, and Mexican-Americans, '*Witherspoon*-excludables' are singled out for exclusion in capital cases on the basis of an attribute that is within the individual's control" (*Lockhart,* 1986, p. 1766). This analysis may also lead some observers to fear a dangerous precedent, suggesting that citizens may be banned from jury service on the basis of unpopular religious or other beliefs.

4. *The constitutional requirement of representativeness is satisfied if the jury pool is representative.* The majority opinion in *Lockhart* cited several precedents holding that the right to a representative jury did not mean that every single group of 12 citizens who sit as a jury must proportionally reproduce the characteristics of the population: "The point at which an accused is entitled to a fair cross-section of the community is when the names are put in the box from which the panels are drawn" (Justice Blackmun, *Pope v. United States* (1971), cited in

Lockhart, (1986, p. 1765). McCree, of course, was not arguing that his jury must include people adamantly opposed to the death penalty, but that such a person must have the same *chance* of inclusion as a death-qualified juror. To say that the fair cross-section requirement was met because the names of these people were put in the box from which the panels were drawn is a childish effort at sleight-of-hand. One doesn't have to have any statistical training to realize that in order to achieve representativeness, the names must not only be put in the box, they must still be in the box when the names of the actual panel members are drawn. True, the names of the death-penalty opponents are put in the box, but then they are all taken out again, and then the actual jury is drawn. Obviously, their chances of serving on a capital jury are exactly the same as if they had never had their names in the box at all: zero.

Again, the courts in *Hovey* and *Grigsby* avoided this error. The opinion in *Grigsby,* which was reversed by the *Lockhart* decision, explicitly stated:

> [1] Any implication by the dissent suggesting the majority requires representation of all cognizable groups on each petit jury is misplaced. [cites omitted] Our holding simply forbids the systematic exclusion of any cognizable group from a petit jury" [*Grigsby v. Mabry*, (1985) p. 230, n. 6].

and

> [2] There is no functional difference between excluding a particular group of eligible citizens from the "jury wheels, pools of names, panels or venires from which juries are drawn" and systematically excluding them from sitting on a petit jury [cites omitted]. The result is the same in either case: a distinct group of the citizenry is prevented from being considered for service on petit juries [*Grigsby,* 1985, p. 230, n. 7].

5. *The state has a legitimate interest in a single jury to decide both guilt and penalty.* The *Witherspoon* opinion explicitly recognized that if future research established that death-qualified juries were

> less than neutral with respect to *guilt* . . . the question would then arise whether the state's interest in submitting the penalty issue to a jury capable of imposing capital punishment may be vindicated at the expense of the defendant's interest in a completely fair

> determination of guilt or innocence—given the possibility of accommodating both interests by means of a bifurcated trial [*Witherspoon*, 1968, p. 520, n. 18].

Justice Rehnquist did not admit that the bias of death-qualified juries had been established, but he did go on to argue that the state had a legitimate interest in "obtaining a single jury that could impartially decide all of the issues in McCree's case" (*Lockhart*, 1986, p. 1768). Neither the state's brief nor the opinion in *Lockhart* is very clear in explaining why a single jury system serves an important state interest.

One argument that was repeatedly raised in the brief and at oral argument was that the excluded jurors were people who would not follow state law, and the state had a legitimate interest in keeping lawbreakers off the jury. Since the excluded jurors were lawbreakers only in the sense that they would not vote to execute anyone, this argument simply restates the basic issue. The group in question included only people who said they *could* follow state law in deciding guilt or innocence, and they were asking only to be included in the guilt/innocence stage of the trial. Thus the state's interest in a single jury is not really addressed by this argument.

A second argument was that having two separate juries would be costly and inefficient. The Court seemed to assume that the only remedy would be to have two entirely different groups of jurors decide guilt and penalty, although various other methods, less expensive and less cumbersome, had been suggested in the lower court opinion (*Grigsby v. Mabry*, 1985). In any case the cost/efficiency argument was never very explicit, perhaps because it seemed tasteless to weigh such considerations against the fairness of the trial.

Finally, the Court reiterated an argument that had first been made in *Smith v. Balkcom* (1981) and had become standard in court decisions that found that death qualification was constitutional. Even though a juror is convinced beyond a reasonable doubt when he or she votes to convict a defendant, there still may be some residual, "whimsical" doubts about the evidence. These residual doubts might cause the juror to hesitate to execute the defendant; thus it is in the *defendant's* best interest to have the same jurors decide both guilt and penalty in order to profit from the mercy associated with whimsical doubts.

Whether such residual doubts are common and whether they result in a reluctance to vote for death are, of course, empirical questions, just as surely as the conviction-proneness of death-qualified jurors is an

empirical question. The difference is that there is not a scrap of empirical evidence on the residual doubt question.[8] Yet the Court rejects the evidence on conviction-proneness because 14 of the 15 studies have some flaw, and endorses the residual doubt argument based on no evidence whatsoever. And of course the argument that a single jury increases one's chances for a sentence of life imprisonment rather than death is likely to seem cold comfort to a defendant who believes he was unfairly convicted in the first place. As Finch and Ferraro (1986) argue, "The trade-off suggested by the court—a more conviction-prone jury for a less death-prone jury—seems one that the accused is unlikely to make" (p. 69-70, n. 169).

6. *None of the research examined real jurors' decisions in real capital cases.* There is one criticism made by the majority in *Lockhart* that applies to all of the empirical research. The studies investigated the behavior of people "who were not actual jurors sworn under oath to apply the law to the facts of an actual case involving the fate of an actual capital defendant" (*Lockhart,* 1986, p. 1763). Zeisel's study had investigated the behavior of sworn jurors in actual felony cases, but they were not capital cases. Of course the very existence of the questionable practice makes it impossible to carry out the "perfect" research to condemn or exonerate the practice. If no *Witherspoon*-excludable citizens are allowed on the jury in any capital case, then it is impossible to discover how their behavior might differ from that of the death-qualified jurors, and the suggestion in *Witherspoon* that future research might establish a bias toward guilt in death-qualified juries becomes meaningless. The *Witherspoon* court assumed that the question could be answered empirically; the *Lockhart* decision seemed to assume that it can not, at least not in this world.

The decisions in *Hovey* and *Grigsby* recognized that the "perfect" experiment would have two juries hearing each of a large number of capital cases, one jury empaneled with the usual method of death qualification, the other questioned only about their ability to be fair and impartial in deciding guilt, each jury believing that it was responsible for the decision. The lower court decisions also recognized that "such an experiment is legally impossible." Thus, "the hypothesis must be tested indirectly" (*Hovey,* 1980, p. 1315). "[I]t is the courts who have often stood in the way of surveys involving real jurors and we should not now reject a study because of this deficiency" (*Grigsby,* 1985, p. 237).

In his dissenting opinion in *Lockhart,* Justice Marshall argued that the absence of the impossible "perfect" study is compensated by the

convergence of findings in all the studies conducted: "Where studies have identified and corrected apparent flaws in prior investigations, the results of the subsequent work have only corroborated the conclusions drawn in the earlier efforts" (Marshall, J., dissenting *Lockhart,* 1986, p. 1773). When the bulk of the data (or in this case, all of the data) points to one conclusion across many settings the scientific "burden of proof" shifts to those who would discredit the data to propose some hypothesis to explain why the findings will not apply in a new setting—in this case, to explain why the excluded jurors would be expected to behave like the death-qualified jurors if they sat on a real capital case.

A more respectable version of the majority's concern is that the research can not predict the *magnitude* of the bias created by the practice of death qualification. How many cases would come out differently? The opinion quoted Finch and Ferraro, who attempted to estimate the size of the biasing effect of death qualification, and concluded that "no definite conclusions can be stated as to the frequency or magnitude of the effects of death qualification" (1986, p. 66). The opinion did not quote Finch and Ferraro's statement that "extant research findings may actually *understate* the magnitude of the problem raised by death qualification" (1986, p. 62). It is quite true that estimates of the magnitude of the bias are likely to be highly unreliable. It could be more serious or less serious than the studies suggest. There is no question, however, that there is a bias.

THE STANDARD OF REVIEW

Ordinarily, and especially recently, the Supreme Court has been reluctant to reconsider the factual conclusions of the lower courts, preferring to rely on the "traditional testing mechanisms of the adversary process" to reveal the facts most clearly to the judge who actually witnesses this testing. In this case, the court of appeals affirmed the district court's findings, and previous Supreme Court decisions had held that the Supreme Court "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 & Manufacturing Co. v. Linde,* 1949, at 275; see also *United States v. Doe,* 1984).[9] Given that the lower courts had dealt extensively with all of the considerations raised in the majority's analysis of the empirical research, where was the obvious error? The answer is that despite the Court's criticisms of the research,

the quality of the research and the lower court's factual findings were not at issue, "[b]ecause we do not ultimately base our decision today on the invalidity of the lower courts' 'factual' findings." (*Lockhart* 1986, p. 1762, n. 3). The Court was unwilling to assert that the studies are invalid, but unwilling to admit that they are valid. With masterful condescension, the Court revealed that it was willing to set aside its better judgment and accept the empirical conclusions. Nevertheless, the practice of death qualification remains constitutional.

> Having identified some of the more serious problems with McCree's studies, however, we will assume for purposes of this opinion that the studies are both methodologically valid and adequate to establish that "death-qualification" in fact produces juries somewhat more "conviction-prone" than "non-death-qualified" juries. We hold, nonetheless, that the Constitution does not prohibit the states from "death-qualifying" juries in capital cases [*Lockhart*, 1986, p. 1764].

Basically, the majority held that a jury of 12 impartial people was an impartial jury for constitutional purposes, as long as no traditionally protected group, such as blacks or women, is excluded.

It seems that the Court's only plausible course was to decide that the basic question was not an empirical question after all. Even if it were possible to demonstrate the biasing effects of death qualification on the outcome of real capital cases, it would make no difference. In capital cases it is constitutional to decide guilt or innocence with juries that are biased toward conviction. In other words, despite the explicit question raised in *Witherspoon,* death qualification is immune to empirical challenge.

CONCLUSION

In the 1980s the public strongly favors capital punishment. In most cases this attitude is based on very little information. The desire to acquire more information is negligible, and when new information contrary to one's attitude is learned, the attitude remains unchanged (Ellsworth and Ross, 1983). The Supreme Court and the public have something in common in this regard.

The majority of the Court has been ideologically committed to the constitutionality of capital punishment since 1976. In addition to the

ideological commitment, there is the practical problem of the ever-increasing number of prisoners under sentence of death. If the death penalty, or some common feature of its administration such as racial bias or death qualification, were to be declared unconstitutional, hundreds of convicted murderers might have to be dealt with all over again. Since *Furman,* the Court's opinions on the death penalty have often been plurality opinions, confusing and contradictory. When a particular practice has been held unconstitutional, the decision has typically affected very few cases. As Gross and Mauro remarked, "*Furman* became the fountainhead of an expanding swamp of uncertain rules and confusing opinions" (1984).

In part, this messy line of constitutional doctrine is a function of the Court's attempts to circumvent the empirical data. Opponents of capital punishment, in particular the NAACP Legal Defense and Educational Fund, have relied heavily on empirical arguments in their challenges to the death penalty. Their use of data has been extraordinarily sophisticated, and their preparation of empirical challenges exceptionally thorough. Therefore the Court's evasion of the data is bound to blur the issues, and it makes the Court's handling of death penalty cases look less competent than its handling of other issues.

An alternative hypothesis worth considering is that the Court is *generally* reluctant to give serious consideration to social science data, and the recent history of opinions on capital punishment has more to do with the Court's attitude toward social science than with its position on the death penalty. Although it is true that the Court has been skeptical of data in other contexts (cf. Ellsworth and Getman, 1987), I do not think this general reluctance can explain the Court's response to data that challenge capital punishment or its administration. First, the Court has considered data on racial discrimination and jury composition in other contexts, and has ruled in accordance with weaker data than those presented in death cases (*Griggs v. Duke Power and Light Co.,* 1971; *Ballew v. Georgia,* 1978).

Second, the Court has cited empirical evidence when it supports the constitutionality of the death penalty. A good example is the use of public opinion data. The Eighth Amendment's prohibition of cruel and unusual punishment has been interpreted as a prohibition against punishments that are seen as excessive or barbaric in contemporary society. Thus punishments that were acceptable when the Constitution was written, such as mutilation, may be found to be cruel and unusual by later generations (*Weems v. United States,* 1910; *Trop v. Dulles,* 1958).

In the years just preceding *Furman,* public attitudes toward capital punishment were evenly split, and in some polls opponents outnumbered proponents. These polls were used by those who challenged capital punishment to argue that the death penalty was cruel and unusual by contemporary standards.

Since then, however, polls have shown increasing levels of support for the death penalty. Since 1976 the Court has often referred to evidence from opinion surveys and state referenda, new legislation, and jury verdicts—"objective indicia that reflect the public attitude" to support its conclusion that "a large proportion of American society continues to regard death as an appropriate and necessary criminal sanction" (*Gregg v. Georgia,* 1976, p. 878), and thus not cruel and unusual by contemporary standards. Likewise, in 1977, when the Supreme Court held that the death penalty for rape was disproportionate and excessive, and therefore unconstitutional, the opinion relied heavily on empirical data indicating that the public rejected the death penalty for rape (*Coker v. Georgia,* 1977).

The most telling evidence that the Court's opinions are a product of the Justices' attitudes toward capital punishment rather than their attitudes toward social science is the case of *Barefoot v. Estelle* (1983). In order to sentence a person to death in Texas, the jury must find that the convicted person is likely to be a continuing threat to society, that is, the jury must predict that he or she will commit dangerous acts in the future. There is substantial consensus among social scientists that the field has not advanced to the stage at which predictions of dangerousness can be made with any accuracy. In *Barefoot,* then, the American Psychiatric Association argued against the use of expert testimony on future dangerousness, because the data indicated that the predictions of experts are usually wrong. The Court, however, held that the Texas law was constitutional, and that experts should be allowed to testify about whether or not the defendant would pose a continuing threat to society. In Texas, the prediction of future dangerousness is usually the only issue the jury must decide in determining whether to execute a defendant. One would expect that the criteria for evaluating the psychiatric evidence would thus be particularly stringent. Instead, the majority found such evidence satisfactory because "Neither petitioner nor the [American Psychiatric] Association suggests that psychiatrists are always wrong with respect to future dangerousness, only most of the time" (*Barefoot v. Estelle* 1983, p. 901).

In *Barefoot* the Court accepted a form of social science "evidence" that was strongly repudiated by the scientific community in order to uphold the constitutionality of the administration of the death penalty in Texas. In *Lockhart* and in *McCleskey* the Court rejected empirical arguments that were strongly endorsed by the scientific community in order to uphold the constitutionality of the administration of the death penalty. The parsimonious explanation for the failure of social science data to influence the Court in death penalty cases seems to be that the outcome of these cases is frequently a foregone conclusion.

NOTES

1. These figures were derived from a computerized search of the LEXIS (Mead data) library.

2. An alternative argument in favor of deterrence is that by punishing certain crimes by death, a government communicates to its citizens that these crimes are completely intolerable. The existence of the punishment teaches the people that murder is evil. Opponents of the death penalty would argue that most people learn that murder is unacceptable whether or not their state or country executes murderers. Some opponents argue that capital punishment actually teaches people the opposite—that killing human beings is sometimes the right thing to do.

3. Of course there are many different ways of determining the proportion of cases in which defendants are sentenced to death because there are many different ways of defining the relevant universe of cases. The figure of 8 out of 10 was cited by the Court itself in the plurality opinions in *Gregg v. Georgia* and *Woodson v. North Carolina* handed down a year before *Coker*. It was cited in support of the propositions (a) that the public accepted capital punishment but felt that is was only appropriate for a few extreme cases, and (b) that the public would therefore reject mandatory death penalty laws. The figure of 8 out of 10 agrees fairly well with the 1983 findings of Baldus et al.

4. Justice Douglas, in his dissenting opinion in *Witherspoon v. Illinois* (1968) argued that the jury would not be truly representative of the conscience of the community unless all members of the community had a chance of serving at the penalty trial as well as the guilt/innocence trial.

5. The Court has been criticized for relying on bad research and for misinterpreting some of the social science evidence in the earlier jury cases. I do not claim that the Justices used the data well. My only point is that the Court has sought out social science research in Sixth Amendment cases, and has incorporated it into several major decisions.

6. Grigsby died in prison; McCree's case had been joined to his; thus the name change. Two lower courts had previously held that the factual evidence was immaterial to the constitutionality of death-qualified juries (*Smith v. Balckom,* 1982; *Keeten v. Garrison,* 1984), but the courts that actually considered the evidence had all concluded that death qualification biases juries against capital defendants on the issue of guilt (*Hovey, Grigsby* and *Keeten v. Garrison* (W.D.N.C. 1984).

7. The one possible alternative might have been to condemn the practice on the ground that excluding any group, however defined, violated the representativeness of the

jury. This was the position Justice Douglas took in his dissent in *Witherspoon.*

8. The opposite hypothesis, that a new set of jurors would be *less* convinced of the rightness of the guilty verdict than the jurors who were responsible for that verdict, seems at least as plausible.

9. Justice Rehnquist argued that the Court is not bound by lower court decisions with regard to this sort of legislative fact finding. However, the fact that they are not *bound* by the lower court's review hardly justifies the majority Justices' complete failure to *consider* the lower court analyses of exactly the same facts that were discussed—and misunderstood—in the Supreme Court opinion.

REFERENCES

BALDUS, D. C., C. A. PULASKI, and G. WOODWORTH. (1983) "Comparative review of death sentences: An empirical study of the Georgia experience." Journal of Criminal Law & Criminology 74: 661-753.

BLUMSTEIN, A., J. COHEN, and D. NAGIN [eds.] (1986) "Deterrence and incapacitation: Estimating the effects of criminal sanctions on crime rates." Panel on Research on Deterrent and Incapacitative Effects, Committee on Research on Law Enforcement and Criminal Justice, Assembly of Behavioral and Social Sciences, National Research Council. Washington, DC: National Academy of Sciences.

BOWERS, W. J. (1974) Executions in America. Lexington, MA: D. D. Heath.

BOWERS, W. J. (1984) Legal Homicide: Death as Punishment in America, 1864-1982. Boston: Northeastern University Press.

BOWERS, W. J. and G. L. PIERCE. (1980) "Arbitrariness and discrimination under post-*Furman* capital statutes." Crime and Delinquency 26: 563-635.

COWAN, C. L., W. C. THOMPSON, and P. C. ELLSWORTH. (1984) "The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation." Law and Human Behavior 8: 53-79.

DANN, R. H. (1935) "The deterrent effect of capital punishment." Friends Social Series, 29: 1-20.

DIKE, S. T. (1982) Capital Punishment in the United States: A Consideration of the Evidence. New York: National Council on Crime and Delinquency.

EHRLICH, I. (1975) "The deterrent effect of capital punishment: A question of life and death." American Economic Review 65: 397-417.

ELLSWORTH, P. C. and J. G. GETMAN (1987) "The use of social science and legal decision making." In L. Lipson and S. Wheeler (eds.) Law and the Social Sciences. New York: Russell Sage.

ELLSWORTH, P. C. and L. D. ROSS. (1983) "Public opinion and the death penalty: A close examination of the views of abolitionists and retentionists." Crime and Delinquency (January): 116-169.

FINCH, M. and M. FERRARO (1986) "The empirical challenge to death-qualified juries: On further examination." Nebraska Law Review, 65: 21-74.

GROSS, S. R. (1984) "Determining the neutrality of death-qualified juries: Judicial appraisal of empirical data." Law and Human Behavior 8: 7-30.

GROSS, S. R. (1985) "Race and death: The judicial evaluation of evidence of discrimination in capital sentencing." U.C. Davis Law Review 18: 1275-1325.

GROSS, S. R. and R. MAURO. (1984) "Patterns of death: An analysis of racial disparities in capital sentencing and homicide victimization." Stanford Law Review 37: 27-153.

HANEY, C. (1984) "On the selection of capital juries: The biasing effects of the death-qualification process." Law and Human Behavior 8: 121-132.
HASTIE, R., S. D. PENROD, and N. PENNINGTON. (1983) Inside the Jury. Cambridge, MA: Harvard University Press.
JACOBY, J. E. and R. PATERNOSTER. (1982) "Sentencing disparity and jury packing: Further challenges to the death penalty." Journal of Criminal Law & Criminology 73: 379-387.
KADANE, J. B. (1984) "After *Hovey*: A note taking into account automatic death penalty jurors." Law and Human Behavior 8: 115-120.
KALVEN, H. and H. ZEISEL. (1966) The American Jury. Boston: Little, Brown.
LEMPERT, R. O. (1981) "Desert and deterrence: An assessment of the moral bases of the case for capital punishment." Michigan Law Review 79: 1177-1231.
Louis Harris & Associates, Inc. (1981) Study No. 814022.
OBERER, W. E. (1961) "Does disqualification of jurors for scruples against capital punishment constitute denial of fair trial on the issue of guilt?" University of Texas Law Review 39, 545-573.
RADELET, M. (1981) "Racial characteristics and the imposition of the death penalty." American Sociological Review 46: 918-927.
RADELET, M. and G. L. PIERCE. (1985) "Race and prosecutorial discretion in homicide cases." Law and Society Review 19: 587-621.
RIEDEL, M. (1976) "Death row 1975: A study of offenders sentenced under post-Furman statutes." Temple Law Quarterly 49: 261-290.
SELLIN, T. (1959) The Death Penalty. Philadelphia: American Law Institute.
SELLIN, T. (1980) The Penalty of Death. Beverly Hills, CA: Sage.
van den HAAG, E. (1975) Punishing Criminals: Concerning a Very Old and Painful Question. New York: Basic Books.
WOLFGANG, M. E. and M. RIEDEL. (1973) "Rape, judicial discretion, and the death penalty." Annals of the American Academy of Political and Social Science 407: 119-133.
ZEISEL, H. (1968) Some data on juror attitudes toward capital punishment. Monograph. Center for Studies in Criminal Justice, University of Chicago Law School.
ZEISEL, H. (1981) "Race bias in the administration of the death penalty: The Florida experience. Harvard Law Review 95: 456-468.
ZIMRING, F. E., J. EIGEN, AND S. O'MALLEY. (1976) "Punishing homicide in Philadelphia: Perspectives on the death penalty." University of Chicago Law Review 43: 227-252.

CASES

Ballew v. Georgia, 435 U.S. 223 (1978).
Barefoot v. Estelle, 463 U.S. 880 (1983).
Batson v. Kentucky, 467 U.S. 106 S. Ct. 1712 (1986).
Brown v. Board of Education of Topeka, 347 U.S. 483 (1954).
Coker v. Georgia, 433 U.S. 584 (1977).
Colgrove v. Battin, 413 U.S. 149 (1973).
Furman v. Georgia, 408 U.S. 238 (1972).
Graver Tank & Manufacturing Co. v. Linde, 336 U.S. 271 (1949).

Gregg v. Georgia, 428 U.S. 153 (1976).
Griggs v. Duke Power and Light Co., 401 U.S. 424 (1971).
Grigsby v. Mabry, 569 F. Supp. 1273 (E.D. Ark. 1983).
Grigsby v. Mabry, 758 F. 2d 226 (8th Cir. 1985).
Hovey v. Superior Court, 28 Cal. 3d 1 (1980).
Jurek v. Texas, 428 U.S. 262 (1976).
Keeten v. Garrison, 742 F. 2d 129 (4th Cir. 1984).
Keeten v. Garrison, 578 F. Supp. 1164 (W.D. N.C. 1984).
Lockhart v. McCree, 106 S. Ct. 1758 (1986).
Maxwell v. Bishop, 398 F. 2d (8th Cir. 1968), *rev'd on other grounds*, 398 U.S. 262 (1970).
McCleskey v. Kemp, 107 S. Ct. 1756 (1987).
People v. Moore, Alameda County Superior Court (California), No. 67113, August-September, 1979.
Pope v. United States, 372 F. 2d 710 (8th Cir. 1967) *cert denied* 401 U.S. 949 (1971).
Proffitt v. Florida, 428 U.S. 242 (1976).
Smith v. Balkcom, 651 F. 2d 858 (5th Cir. 1982).
Strauder v. West Virginia, 100 U.S. 303 (1880).
Trop v. Dulles, 356 U.S. 86 (1958).
United States v. Doe, 104 S. Ct. 1237 (1984).
Weems v. United States, 217 U.S. 349 (1910).
Williams v. Florida, 399 U.S. 510 (1968).
Witherspoon v. Illinois, 391 U.S. 510 (1968).
Woodson v. North Carolina, 428 U.S. 304 (1976).