THE DEATH PENALTY MEETS SOCIAL SCIENCE:
Deterrence and Jury Behavior Under New Scrutiny

Robert Weisberg
Stanford University, School of Law, Stanford, California 94305;
email: weisberg@stanford.edu

Key Words  capital punishment, econometrics, sentencing, life imprisonment, juries, discrimination

Abstract  Social science has long played a role in examining the efficacy and fairness of the death penalty. Empirical studies of the deterrent effect of capital punishment were cited by the Supreme Court in its landmark cases in the 1970s; most notable was the 1975 Isaac Ehrlich study, which used multivariate regression analysis and purported to show a significant marginal deterrent effect over life imprisonment, but which was soon roundly criticized for methodological flaws. Decades later, new econometric studies have emerged, using panel data techniques, that report striking findings of marginal deterrence, even up to 18 lives saved per execution. Yet the cycle of debate continues, as these new studies face criticism for omitting key potential variables and for the potential distorting effect of one anomalously high-executing state (Texas). Meanwhile, other empiricists, relying mainly on survey questionnaires, have taken a fresh look at the human dynamics of death penalty trials, especially the attitudes and personal background factors that influence capital jurors.

INTRODUCTION

The endlessly recycling debates in the United States over whether we should have capital punishment tend to mix two entirely different types of discourse: a retributivist discourse about whether certain criminals morally deserve the death penalty, and a utilitarian discourse about whether the death penalty serves to reduce murder and, occasionally, about whether it has been imposed fairly. On the latter score, social science research has played a persistent, if fitful, role in influencing jurists and occasionally legislators and even the lay public. This review focuses almost entirely on the most visible category of social science research and the death penalty—the question of deterrence. In the final section, I briefly allude to other areas of capital punishment policy and law where social science research has played a role, such as racial issues and jury behavior.

Perhaps no question relevant to law and social science has been so salient in American public opinion in the last few decades, and yet has so vexed social scientists, as, “Does the death penalty deter murder?”
The general hypothesis of the deterrent justification for capital punishment is straightforward: Although many potential murderers may not rationally reflect on the consequences of their actions, a considerable number do weigh negative consequences, consciously or instinctively. Indeed, one reason why homicide detectives sometimes have to struggle to find the right suspect is that many murderers go to great lengths to conceal their acts or escape detection. Conversely, some scholars argue that it is common sense to expect the death penalty to increase murders. Such an increase might occur because (a) there are suicidal killers out there who will only murder if they think they can then achieve the ultimate penalty, or (b) there is a brutalization effect—i.e., executions or the willingness to execute lead some potential killers to act out their intents, either inspired by the role model of the state or because in some way the state’s willingness to execute cheapens the value of life (Bowers & Pierce 1980). But data supporting any positive correlation between the death penalty and murders are weak. Moreover, for those who doubt the deterrent effect of the death penalty, it may be self-contradictory to claim that criminals are too impulsive to contemplate the negative consequences of their actions but are still sufficiently sensitive to public signals that they are sometimes motivated to kill by social signals modeling killing.

What about direct evidence of deterrence? Certain seemingly direct measures of the deterrent effect of the death penalty are available. For example, once facing prosecution, almost all criminals seek to avoid punishment, and only in about 1% of cases does a capital defendant actually request the death penalty or waive rights of appeal. And then there is the anecdotal evidence, as in the case of one murderer who said he robbed and killed drug dealers in Washington, DC, where he was conscious that there was no death penalty, but specifically chose not to do so in Virginia because he was frightened by memories of Virginia prisoners in the electric chair (Blecker 2005). But such seemingly direct measures are too unsystematic to play an important role in social science research on the death penalty. Thus, the focus of this review is, of course, on more systematic statistical research.

But before examining the undulating history of research on this question, some key qualifications are in order. First, other things being equal, the presence or enforcement of the death penalty obviously will produce fewer homicides than not punishing homicides at all. Thus, the question is one of marginal deterrence—i.e., whether the death penalty deters more homicides than the next most severe penalty, which in all jurisdictions is some form of life imprisonment, and in most is the relatively new sentence of life without the possibility of parole. So it is solely for convenience that throughout this review deterrence stands for marginal deterrence.

Second, deterrence is only one way that the death penalty could reduce murders. Another utilitarian justification for punishment is incapacitation, and one might posit that the death penalty reduces the number of murders not by virtue of sending a deterrent message to other potential murderers but simply by preventing the condemned killer from ever killing again. Of course, that issue only arises if, assuming the alternative punishment is always a true life sentence, a convicted...
murder were able to kill in prison (or perhaps order a killing on the outside). The magnitude of these possibilities lies outside the scope of this review.

Third, the term murder may signify to the lay public any illegal or intentional killing. But of course murder is a complex legal concept covering a number of forms of homicide. Thus, as noted below, statistical analysis of the deterrent effect of capital punishment is somewhat contingent on the question of what type of homicide might be deterred. Most crime data indexes count all homicides explicitly labeled as murder under state law, plus a category called “non-negligent manslaughter” (FBI 2003). In this review, murder is the necessarily imprecise term generally used for the types of homicides covered by the research discussed here.

Fourth, whatever measures of statistical significance one uses, social scientists face a couple of blunt facts about death penalty and deterrence. First, the percentage of people sentenced to death in the United States who actually are executed is minute, so if research is concerned with the actual or perceived likelihood of a death-sentenced murderer suffering the ultimate penalty, the data will always seem insufficient. Second, as discussed below, the executions that do occur are disproportionately centered in a few states—indeed, close to a majority in a single state—so the various effects of skewing hamper sound empirical inference-drawing.

EARLY RESEARCH: THE SELLIN/ERHLICH STANDOFF

Although the issue of the deterrent effect of the death penalty has a long legacy in criminology generally, it began to play a major role in American legal doctrine most notably in *Furman v. Georgia* (1972) and *Gregg v. Georgia* (1976). In *Furman*, the Supreme Court declared all then-operational death penalty laws of the United States unconstitutional. In *Gregg*, it upheld against Eighth Amendment challenges the new type of guided discretion laws that are now used in three fourths of the states and in the federal system. In the years following *Gregg*, important social science evidence of a deterrent effect to which *Gregg* alluded met widespread skepticism and arguably utter refutation in later social science research. And now, 30 years after the restoration of the death penalty in the United States, the state of the research has become roiled again with new claims of proof of that deterrent effect.

Generally, what research existed before 1972 did little to establish any deterrent effect. Most of the early work was done by criminologists or psychologists whose empirical work relied mainly either on comparisons of homicide rates in states with and without capital punishment, or, within a particular jurisdiction, on comparisons of homicide rates before and after executions. But because this research did not employ the statistical technique of multiple regression, it could not meaningfully distinguish the effect of capital punishment on murder from the effects of other factors. One of the oldest studies (Dann 1935) looked at homicides within 60 days of an execution. Two decades later, another study (Savitz 1958) examined murders eight weeks before and after trials ending in death sentences. Working with sparse data, these studies found no deterrent effect. Another 1950s
study (Schuessler 1952) compared murder rate changes between states that maintained capital punishment and those that never had capital punishment in the year under study and also measured before-and-after murder rates in states that switched in one direction or another. This study concluded that non–death penalty states have murder rates equal to or lower than those of death penalty states. In a key example, the study noted that South Dakota went from a non–death penalty to a death penalty regime in 1939 and saw in the next decade a modest drop in its murder rate of 16%, but that North Dakota, which was a non–death penalty state before and after 1939, enjoyed a drop of 40% for that same decade. This study also attempted to analyze the effects of actual executions and found no evidence of a deterrent effect.

The most notable figure in this early phase of deterrence research was Thorsten Sellin. Sellin, examining the period from 1920 to 1955, found that states retaining the death penalty exhibited murder rates at least as high as those that had abolished it (Sellin 1959). He also did a rough comparison of contiguous jurisdictions and found that on the whole they exhibited similar murder rates and homicide rate trends even where they differed on capital punishment.

However, Sellin could not explain a few contiguous pairs with dramatic differences—especially Ohio/Michigan and Colorado/Kansas—perhaps because he did not address the possibility that some paired states differ significantly along social, economic, or political dimensions that affect murder rates. Sellin also looked at murder rates in a number of states over time as a way to finesse the initial condition problem. That is, he examined murder rates in particular states when they changed from having the death penalty to abolishing it or from not having it to reinstituting it. And, once again, he found no evidence of deterrence. Sellin acknowledged the problem of recursive effect—the possibility that states abolish capital punishment when and because the murder rate is falling, thus raising a problem of reverse causality—and he performed some tests that yielded results inconsistent with this hypothesis. Finally, Sellin studied killings by life prisoners and discerned that a majority of the small number of prison killers were in death penalty states.

The pivotal moment in the history of death penalty deterrence research came in the mid-1970s with the work of University of Chicago economist Isaac Ehrlich. Indeed, the death penalty deterrence debate might be said to be divisible into two eras: before Ehrlich (BE) and after Ehrlich (AE). Ehrlich was the first to study capital punishment’s deterrent effect using multivariate regression analysis (Ehrlich 1975). This approach enabled Ehrlich to distinguish the effects on murder of such different factors as the racial and age composition of the population, average income, unemployment, and the execution rate.

Ehrlich’s famous 1975 paper examined time series data for the period 1933–1969. He tested the effect on national murder rates of various potential deterrent variables (probabilities of arrest, conviction, and execution), demographic variables (size of population, percentage of minorities in the population, percentage of people ages 14–24 in the population), economic variables (unemployment rate, per capita permanent income, per capita government expenditures, and per capita
expenditures on police), and a time variable. Ehrlich concluded that there was a statistically significant negative relationship between the murder rate and execution rate, i.e., a deterrent effect. Specifically, he estimated that each execution resulted in approximately seven or eight fewer murders. This paper was offered to the Supreme Court in draft form by the solicitor general when Gregg was originally litigated. It was then cited in the plurality decision in the Gregg decision itself, in which Justice Stewart cited it as part of a mix of studies that, he inferred, established a scholarly standoff on the question of whether the death penalty deterred murder and so justified treating the deterrence issue as essentially irrelevant to the constitutionality of capital punishment.

Ehrlich’s second paper (1977) studied cross-sectional data from the 50 states from 1940 to 1950. That is, whereas the first paper tested how the total U.S. murder rate changed across time as the execution rate changed, Ehrlich now explored the relationship, during a single year, between a state’s execution rate and its murder rate. Ehrlich again used multivariate regression analysis, including variables similar to those in the 1975 study (for a deterrent variable he added median time spent in prison as well as a dummy variable to distinguish executing states from nonexecuting states, and, for economic variables, median family income and percentage of families with income below half of the median income). Again, he inferred a significant deterrent effect. Ehrlich himself was publicly cautious about trumpeting his conclusions, but immediately his work received both extravagant public praise and sharp academic criticism. It rapidly entered the political sphere as citable proof that each execution could indeed save at least eight innocent lives. But in the ensuing years, many social scientists tried to replicate Ehrlich’s results with differing data and methods, and most were unable to confirm Ehrlich’s conclusions. Indeed, in 1978, a National Academy of Sciences scholarly panel publicly criticized Ehrlich’s work (Blumstein & Cohen 1978).

Because Ehrlich’s work was, or seemed, so pivotal in the history of research on the death penalty, a fuller description of the problems in his studies may be useful, and it is supplied by legal sociologist Richard Lempert (1981). As seen by Lempert, Ehrlich could be credited for deploying multivariate regression analysis to study the deterrence hypothesis, but he could be faulted for not using the technique comprehensively or well. First, as Ehrlich himself recognized, his work failed to measure the length of prison sentences in general or the probability of life sentences in particular. Hence, his work is wanting precisely on the question of marginal deterrence. As Lempert suggests, if murderers who would have been executed or sentenced to life imprisonment during periods when execution rates were high often received sentences of less than life when execution rates were low (perhaps reflecting generally more lenient sentences in periods of diminished fear of crime), an association between low homicide rates and high execution rates would not necessarily indicate that executions are a greater deterrent than life sentences. The association might exist because executions and life sentences, or indeed just life sentences, are greater deterrents than sentences of less than life. The marginal deterrence issue, which is whether executions are a greater deterrent
than life sentences, or in today’s sentencing schemes sentences of life without parole, is in these circumstances not necessarily addressed by the data.

Second, replication of Ehrlich’s data shows that if we eliminate the years 1965 through 1969, the deterrent effect is statistically insignificant or even reverses itself. As Lempert (1981) explains this problem, the premise of Ehrlich’s approach is not in the first instance empirical, but theoretical. It posits an economic model of why punishment deters, then uses empirical evidence to test the soundness of this deterrence theory, and then finds the theory confirmed. But the theory cannot explain why the economic model yields different results for different periods of years because this model assumes something fairly essential about human nature. The only way to accommodate the issue of time sensitivity to sustain Ehrlich’s finding is to take account of other factors, such as the state of racial tension, the rate of gun ownership or use, the Vietnam War and political events of other sorts, and then to see if they explained the model’s sensitivity to time period. Not only did Ehrlich not do this, but, Lempert argues, no obvious explanations for the time sensitivity come to mind.

Finally, Lempert (1981) suggests a reversal of Ehrlich’s technique, treating the results as the statement of a hypothesis and then asking whether that hypothesis jibes with data from various jurisdictions. One researcher did a version of this test on several states and found no support for Ehrlich’s hypothesis (Bailey 1978). Lempert himself takes the approach of borrowing from Sellin’s idea that fluctuations in homicide rates over time tend to be similar in contiguous states, so that if one state executes and the other does not, the advantage of the executing state in reducing homicides should increase with each additional execution. Looking back to Sellin through the lens of Ehrlich promises better controls on other relevant variables because each state becomes its own control. Lempert tests Ehrlich’s data along these lines and finds no support for the Ehrlich hypothesis (Lempert 1983).

THE STATE OF RESEARCH AFTER EHRLICH

The newness of Ehrlich’s methods, coupled with his striking findings, brought great attention to his research. Justice Stewart’s inference in Gregg of an empirical standoff on the deterrence issue, prominently citing Ehrlich for the proposition that the death penalty deters, further spurred researchers to examine the deterrent effect of executions. The papers that immediately followed Ehrlich used his data or similar data sets and the same or related statistical methods. Some of this after-Ehrlich, or AE, research found a deterrent effect of capital punishment (Cloninger 1977, Yunker 1976), but others did not (e.g., Bowers & Pierce 1980, Passel & Taylor 1977), while one study came to mixed conclusions depending on the cross-section year used (Leamer 1983).

A second generation of AE econometric studies in the late 1980s and 1990s extended Ehrlich’s national time series data or used more recent cross-sectional data. As before, some papers found deterrence by using, for example, an extension
of Ehrlich’s national time series data covering up to 1977 (e.g., Layson 1985) or national time series data for 1966–1985 (e.g., Chressanthis 1989). Still, others found no deterrent effect by, for example, using daily data for California during 1960–1963 (e.g., Grogger 1990).

Nevertheless, most of these AE studies suffered from their dependence on either national time series or cross-section data. National time series data created a serious aggregation problem. For example, when the murder rate in a state with no executions happens to increase simultaneously with a decrease in the murder rate in a state with a number of executions, the data might mask a true deterrent effect. In contrast, cross-sectional studies, by definition, do not account for changes in criminal behavior and the operations of the criminal justice system over time, nor can they account for cultural factors that might affect the homicide rate in particular regions.

PANEL STUDIES AND THE NEW DETERRENCE CLAIMS

Recently, an impressive new generation of deterrence studies has promised to overcome these difficulties by relying on panel data—that is, data from numerous units (in terms of American criminal justice, the 50 states or all counties in the United States) for numerous time periods. These data sets allow for comparisons across jurisdictions over time; they typically include information on potentially confounding variables; they have enough observations to ensure that analyses based on them will have reasonable statistical power; and they benefit from the increased rate at which executions occurred during the 1980s and 1990s. And, most dramatically, these recent studies, using modern regression techniques, find that executions have not just a significant but a substantial deterrent effect.

For example, one new study by Hashem Dezhbakhsh, Paul H. Rubin, and Joanna Shepherd draws on 20 years of data from 3054 counties nationwide to test the effect of county differences on murder rates (Dezhbakhsh et al. 2002). The authors conclude that all types of homicide are deterred by the death penalty, and they infer from their results that each execution prevents as many as 18 murders. Another study by Shepherd (2004a) uses monthly data from all 50 states over 22 years to test the short-term effect of the death penalty, and also takes the important extra step of examining different gradations of homicide. The gradation factor is important because some might argue that, for example, so-called heat-of-passion killings are impossible to deter or that other types of killings might even be inspired by executions. The Shepherd (2004a) study finds that the combination of death sentences and executions deters all types of homicide, from impassioned intimate killings to stranger killings and robbery-motivated killings, regardless of the race or ethnicity of the killer or victim. It concludes that on the whole each death sentence deters approximately 4.5 homicides and that each execution deters approximately 3 more. Notably, another recent study by Dezhbakhsh and Shepherd focuses on the flipside of deterrence—that is, the effect on the murder rate of delays in or even
moratoria on actual executions (Dezhbakhsh & Shepherd 2003). Delays between
death sentence and execution, of course, depend on the vagaries of the state and
federal appellate systems, and moratoria may result from judicial decisions in
particular jurisdictions that suspend imposition of death sentences or on executive
decisions to suspend actual executions. This study, using state-level panel data
from 1960–2000, compares the murder rate for each state immediately before and
after the state either suspended or reinstated capital punishment. This approach
relies on the fact, or assumption, that many factors that might influence the murder
rate, i.e., social or cultural factors or operational changes in criminal justice, change
only slightly over a short period of time. In addition, and happily for this study,
the various suspensions started and ended in different years in different states and
were of widely differing durations. The study finds that 90% of states manifest
higher murder rates after suspensions, whereas 70% show murder rate drops after
reinstatements. More strikingly, this study concludes that every reduction in the
average wait between death sentence and execution of 2.75 years deters an extra
murder. The authors pronounce, “The results are boldly clear: executions deter
murders and murder rates increase substantially during moratoriums. The results
are consistent across before-and-after comparisons and regressions regardless of
the data’s aggregation level, the time period, or the specific variable to measure
executions” (Dezhbakhsh & Shepherd 2003, p. 27).

Finally, recent research by a Federal Communications Commission economist,
Paul Zimmerman (2004), using state-level panel data from 1978 to 1997, not only
finds a deterrent effect but more boldly seeks to distinguish the effects of particular
methods of execution (this last effort may be legally moot because virtually all
executions now use lethal injection). Using state-level panel data from 1978 to
1997 for all 50 states (excluding Washington, DC), Zimmerman concludes not
only that each execution deters an average of 14 murders but that electrocution
can push the number closer to the mid-20s.

NEW DETERRENCE STUDIES USING
OTHER TECHNIQUES

The new wave of research finding strong evidence of a deterrent effect is not
limited to the panel data studies. For example, Cloninger & Marchesini (2001)
rely on a portfolio analysis in a type of controlled experiment by examining an
unofficial moratorium on executions in Texas during most of 1996. They infer that
this hiatus spared few condemned prisoners but caused a significant net increase
in lives lost to murder. Another cross-sectional study (Brumm & Cloninger 1996),
covering 58 cities in 1985, sought to measure the influence of criminals’ perceived
risk of punishment. It concluded that this perceived risk, including the perceived
probability of execution, is negatively and significantly correlated with the murder
rate. Other studies, including a reentry to the fray by Ehrlich himself (Ehrlich & Liu
1999), use state-level and cross-section analysis to reconfirm that executions have
DEATH PENALTY

a significant deterrent effect. And another study, by Ehrlich’s coresearcher, Liu (2004), finds that legalizing the death penalty not only adds capital punishment as a deterrent but also increases the marginal productivity of other deterrence measures in reducing murder rates.

Finally, Yunker (2002) tests the deterrence hypothesis using two sets of postmoratorium data: state cross-section data from 1976 and 1997 and national time series data from 1930–1997. He finds a strong deterrent effect in the time series data, an effect that disappears when the data are limited to the 1930–1976 period. Therefore, he concludes that postmoratorium data are critical to testing the deterrence hypothesis.

Summarizing this new wave of deterrence research, both the panel-based studies and others, one prolific participant in this research confidently draws even a further conclusion:

[T]he studies that find a deterrent effect of other criminal sanctions give additional support to the deterrent effect of the death penalty, because, if lesser sanctions deter, then we know that more severe sanctions also deter. The studies that find a deterrent effect of 1. increased police presence, or any other levels of security; 2. arrest/arrest rates; 3. criminal sentencing/incarceration terms; and 4. the presence of rules, laws and statutes all provide additional, collateral support for the deterrent effect of the death penalty (Shepherd 2004b).

THE NEW STUDIES UNDER SCRUTINY

The apparent power and unanimity of this new round of studies in proving a deterrent effect has, unsurprisingly, provoked a strong response from skeptics. Although the new round of research has not yet been subjected to the depth and breadth of peer review that ultimately undermined confidence in Ehrlich’s early studies, some general points of attack and some specific criticism of certain components of the new studies have emerged.

As summarized by Jeffrey Fagan (2004), two major criticisms stand out. First, all these studies suffer too much from the statistical risk that their overall findings are driven by a few outlier jurisdictions—most notably Texas. Thus, more fine-tuned comparisons between certain states (say, Texas and California) will be needed to retest the results. Next, the studies do not take account of the most important new legal innovation that has arisen in the post-Gregg era—namely, the availability of life without the possibility of parole (LWOP) sentences in all death penalty states except two (New Mexico and, ironically, Texas, where a new LWOP law has just been enacted). LWOP sentences are far more numerous than death sentences these days, and beyond their obvious incapacitating effect they may well have a powerful deterrent effect as well. Indeed, LWOP may be the key deterrent even when the potential offender might also somewhat fear a death sentence. The data showing that some death row inmates waive their appeals are at least anecdotal evidence
that a criminal code with a maximum sentence of LWOP alone may be more of a deterrent than a code that allows for either LWOP or the death penalty.

Other potential lines of criticism that Fagan suggests include the following:

1. These new studies tend to aggregate several forms of murder, and, as above, the one study that breaks them down purports to find all forms deterrable. This conclusion may be implausible if we believe that heat-of-passion killings are necessarily somewhat harder to deter than other murders. If so, more fine-tuned research will be needed, especially of such specific contextual factors as the availability of guns in certain domestic situations.

2. The new studies do not control for the phenomenon of autoregression, that is, the influence that trends in certain years may exert over longitudinal or time series data covering succeeding years. This problem is especially serious in the context of very rare events like executions.

3. The new studies are only sporadically successful, at best, in accounting for controls supplied by the various operations of the criminal justice system, including such essential factors as the success of police in even identifying offenders. It is a virtual cliche of criminal deterrence that the certainty of punishment, of any type, is a more effective deterrent than the possibility of severe punishment, contingent on apprehension and conviction. If the cliche is true, then initial police success in catching offenders should be a more effective deterrent than the rarer death sentences or still rarer executions. If high-executing states also have higher-than-average homicide clearance or arrest rates, this fact could explain the apparent deterrent effects. Some of the newer studies try to control for murder or homicide arrest rates (Dezhbaksh et al. 2002, Mocan & Gittings 2005). But because arrest rates are likely to be particularly high for homicides that are not death-eligible or for which the death penalty is seldom given (e.g., fights between friends, crimes of passion), the adequacy of the control is questionable unless arrest rates as well as homicide rates are broken down by the death-eligibility of the crime. Unfortunately, none of the new studies attempts that breakdown.

4. The studies ignore large amounts of missing data in important states such as Florida, thus potentially biasing their conclusions. Fagan suggests that different techniques for restoring missing data should be used to determine whether the lack of available data can explain findings of deterrence.

Finally, to those who tell a deterrence story by using new data and quantitative methods to shore up their findings, Fagan suggests that these researchers look for confirming evidence in actual mechanisms by which deterrence may operate. For example, the deterrence case could find support in evidence that violence-prone people are aware of executions and the relative likelihood of executions in their own states. Studies that use only national data improbably assume that violence-prone people are aware of execution rates in faraway states, and this assumption needs scrutiny. Similarly, researchers might offer an explanation of
why executions should deter non-death-eligible homicides. Do potential offenders attend to punishments enough to know that murderers may be executed but not enough to know that only certain kinds of murders are death-eligible? Do they know that executions for murder occur but not know how rare they are in many states? Is there any evidence that murderers rationally decide to forego homicide and use less lethal forms of violence? On all these questions about deterrence, argues Fagan, empiricists should consider the contemporary social science research on the generally bounded rationality of human decision making and attempt to apply it to the even more dubiously rational thinking of violent offenders.

One new deterrence study is worth further attention because its findings have been closely examined, and its data reanalyzed, by a leading quantitative sociologist, Richard Berk. Berk’s (2005) article is a sharp critique of work by H. Naci Mocan and R. Kaj Gittings (2005). Mocan & Gittings use state-level panel data from 1977 to 1997 (including information on all 6143 death sentences in this period) to examine the relationship between executions, commutations, and murder. Their study finds a significant deterrent effect, suggesting that each execution deters an average of five murders. More strikingly, it concludes that each commutation results in approximately five extra murders and that each removal from death row generates an additional murder. Finally, it infers that every set of three additional pardons (i.e., commutations of the death sentence) causes 1 to 1.5 additional murders.

Berk’s (2005) sharp critique, one with implications beyond the Mocan-Gittings study that is his focus, argues that these new studies show flaws ranging from the “conceptual leap of treating observational data as an experiment to a large number of nuts-and-bolts statistical difficulties.” Berk points out that in the Mocan-Gittings study, the mean for the number of executions per state per year is 0.35, implying that each state executes about one prisoner every three years, but he then notes that because the standard deviation is 1.35, skewing is a serious concern. Moreover, Berk notes that in the Mocan-Gittings study, the median is zero, with the mean dominated by a few extreme values (i.e., 29 for Texas in 1997, the last year studied). Thus, statistical leverage becomes a serious problem because extreme values of an explanatory variable are paired with extreme values of the response variable. As Berk says, “The potential impact of leverage on a model’s fit becomes a reality,” and the problem is especially severe when an extreme value is not just atypical, but also an outlier—that is, located a great distance from the mass of the data. (This is true with respect to the number of homicides per year, for which the mean is 420 but the standard deviation is 607.)

Further, argues Berk, after controlling for potential confounders, we can conclude that once one knows the large number of homicides in a particular state during the 20-year period, knowing the number of executions adds virtually nothing to the analysis. (This phenomenon is especially evident in situations of five executions or fewer.) As for the temporal dimension, when indicator variables for years rather than states are used to account for national trends in the number of homicides and the homicide rate, we gain no new knowledge of any deviance.
Reanalyzing the data, Berk finds that the relationship Mocan & Gittings report seems entirely dependent on the small number of states with more than five executions. If we exclude the 11 observations out of 1000 that involve more than five executions from the analysis, then we find no systematic difference in the average homicide rate between states that have had no executions and those that have had one; we find a slight negative relationship as we move from states executing one individual to states executing three; and we find a slight positive relationship as we move from states with three to those with five.

Berk acknowledges the possible counterargument that states may differ in their inclination and ability to seek the death penalty, so that controlling for state differences is misleading. If so, using just the 1977 homicide rate as a predictor could adjust for that because the factors affecting the homicide rate were hardly stationary over the following 20 years. Nevertheless, Berk concludes that using this factor as a predictor affects the outcome hardly at all. Unsurprisingly, then, if instead of using the number of executions one uses a binary indicator of zero executions versus one or more executions, evidence of a deterrence effect disappears. Equally unsurprisingly, notes Berk, little evidence of deterrence appears when Texas is removed from the analysis. Put another way, if we shuffle the number of executions for all states other than Texas randomly, so that the number is unrelated to any of the other variables for those states, and we then add Texas back into the mix, the result is an apparent deterrent effect similar to the one found in the real data. Still more skeptically, Berk suggests that using monthly time units or county area units cannot solve any of these problems because the number will be zero for the great majority of them. Finally, he questions whether the data from Texas are ample enough to prove a deterrent effect to the death penalty even there.

THE UNCERTAIN FUTURE OF DETERRENCE RESEARCH

Intuitively, the notion that the death penalty deters finds support in some very rough empiricism about recent events in the United States: Murder rates plummeted over the last decade and a half, and some might argue that a steady diet of executions must have played a role. From 1966–1980, the murder rate nearly doubled, from 5.6 to 10.2 per 100,000. During that 1966–1980 period, the United States averaged only one execution every three years, with a maximum of two executions per year, most obviously because that period covers the last national moratorium on executions (June 1967 to January 1976). Conversely, between 1995 and 2000 the national murder rate dropped 46%, from a high of 10.2 per 100,000 to 5.5 per 100,000, while executions for that period averaged 71 per year. But of course those figures, however convincing they might look to some, tell us little about the real relationship, if any, between these parallel rates. Similarly, one of the most striking correlations comes in the nation’s most active death penalty jurisdiction, Harris...
County (Houston) Texas, where the murder rate dropped 73% between 1982, when executions were resumed, and 2000. But that correlation may just signal that Texas is too anomalous to tell us much about the nation. More generally, while homicide rates were dropping dramatically across the nation, so were rates for other crimes not punishable by death (Levitt 2004).

Data from the multiple regression studies already mentioned are less intuitive than the simple statistics given above but are obviously potentially more meaningful. Nevertheless, some have questioned whether the modern econometric approaches, which most of the new research employ, are as powerful or sophisticated as they might appear. As one harsh critic summarized the econometric approaches, “There is simply too little data and too many ways to manipulate it” (Goertzel 2004). That is, there are too many ways to select model specifications. As a technical matter, notes another critic, to obtain a significant deterrent effect many new studies take the questionable approach of adding a set of data with no executions to a time series and including an executing/nonexecuting dummy in the cross-sectional analysis (Cameron 1994). Thus, we see that the proper specification of econometric approaches is open to controversy, and choices made by researchers, even if defensible, have an inevitable subjective element and could conceivably be more important than the data in determining a study’s conclusions.

Recall Justice Stewart’s remark in Gregg that there is an empirical standoff on the matter of the death penalty’s deterrent value. That remark may thus now be true even if, when uttered, it distorted the apparent weight of the studies that had been done to that point. It is certainly difficult for the uninvolved observer to be confident of where the truth lies. The claims of the latest round of empirical research appear strong, and the work is not vulnerable to the relatively simple and convincing refutations that followed Ehrlich’s initial foray into these matters. Fagan’s critique and Berk’s close look at the Mocan-Gittings data suggest, however, that the results of even these sophisticated studies will have to be qualified as the analyses of the capital punishment deterrence data become yet more refined. Whether the effects of further scrutiny will be to support the deterrence hypothesis, while perhaps putting it in more precise context, or will be to provide further evidence of the null hypothesis of no effect remains impossible to say. We can, however, conclude with more confidence, now that critics have begun to weigh in on the most recent research, that the relationship between executions and murders still lacks clear proof.

SOCIAL SCIENCE RESEARCH ON OTHER DEATH PENALTY ISSUES

Some of the most incisive and insightful interventions of social science into the operation of the death penalty have addressed issues other than deterrence. I briefly discuss some of these studies here.
Victim-Race Discrimination

When the Supreme Court temporarily suspended the use of the death penalty in the United States in *Furman* in 1972, a major issue was whether capital punishment was being imposed in a racially discriminatory manner. Many noted that the penalty was imposed on minorities, most notably black Americans, in numbers several times greater than their proportion in the population. But because the disproportion was similar in terms of convictions for murder, it was difficult to argue that the death penalty by itself was inflicted disproportionately on black defendants. The belief that the death penalty did indeed discriminate on the basis of the race of the defendant may have influenced the *Furman* outcome, but soon thereafter in *Gregg* the Court held that the new post-*Furman* death penalty statutes were well designed to prevent any such effects.

After the Court reimposed the death penalty in *Gregg*, the focus of the death penalty discrimination research shifted more substantially from the race of the defendant to the race of the victim. A major study, first published in 1983, by David C. Baldus, George Woodworth, and Charles Pulaski (Baldus et al. 1990) examined over 2000 murder cases that occurred in Georgia during the 1970s. It inferred that defendants charged with killing white persons received the death penalty in 11% of the cases, whereas those charged with killing blacks received the death penalty in only 1% of the cases. Ironically but significantly (given that most killings occur between members of the same race or group), the race-of-defendant numbers revealed a reverse disparity—4% of the black defendants but 7% of white defendants received the death penalty.

Baldus et al. (1990) also divided the cases according to the combination of the race of the defendant and the race of the victim. They found that capital punishment was imposed in 22% of the cases with black defendants and white victims; 8% of the cases with white defendants and white victims; 1% of the cases with black defendants and black victims; and 3% of the cases with white defendants and black victims. As for prosecutorial penalty-seeking decisions, prosecutors asked for capital punishment in 70% of the black-kills-white cases; 32% of the white-kills-white cases; 15% of the black-kills-black cases; and 19% of the white-kills-black cases.

The Baldus study initially used multiple regression techniques to control for 230 variables; its subsequent regressions included only the theoretically or substantively most important explanatory variables. The study concluded that people killing white victims were more than four times as likely to receive death as those killing black victims. A parallel study using newer data and covering other states (Gross & Mauro 1989) came to very similar conclusions. No later research has seriously questioned these results.

The Baldus study was the main evidence used by the defendant in *McCleskey v. Kemp* (1987), a case of a black man charged with killing a white, to argue for a reversal of the death sentence on the basis of racial discrimination. When *McCleskey* went to the Supreme Court, many predicted that the Court would dodge the issue by finding the empirical research unconvincing or not yet sufficient.
To the surprise of many, the Court effectively mooted this line of research by its apparent willingness to concede that the Baldus study was accurate. But on McCleskey's claim that his death sentence therefore violated the equal protection clause, the Court followed established doctrine in requiring proof of intentional discrimination. And the statistical evidence in the Baldus study was not designed to show intent, and except in the case of extreme intentional discrimination in which the inference is inescapable, statistical evidence of discrimination will usually be consistent with a variety of causes. Although discriminatory intent cannot be ruled out in some death penalty cases, the discriminatory effects documented by Baldus may well have resulted from a complex mixture of half-conscious or unconscious decisions by various legal system actors, including judges, juries, and prosecutors. And as for McCleskey's alternative Eighth Amendment claim that the effects themselves rendered the death penalty unconstitutional, McCleskey lost because, paradoxically, the Court found that the implications of the Baldus study were too great for the system to bear: The Court assumed the causes of the discriminatory effect Baldus reported probably infected the entire law enforcement system, but the justices feared that if they recognized the defense in a capital case, the logic of the decision would effectively require that they put our entire criminal justice system into receivership.

Prosecution-Prone Jurors

A year before McCleskey, the Court heard and rejected another very dramatic social science–based claim about the death penalty in Lockhart v. McCree (1986). The claim specifically rested in large part on the new type of death penalty procedure legislated after Furman and approved by Gregg. As the post-Gregg statutes operate, a jury first determines whether the defendant is guilty of the highest degree of murder. Then, if the prosecutor seeks the death penalty, a second, separate penalty trial occurs, and in the great majority of the cases the decision maker at this second trial is the very same jury that has already convicted the defendant. This jury is always death qualified. That is, in helping choose the jury, the prosecutor can challenge for cause those jurors who state in voir dire that they categorically oppose and would refuse to impose the death penalty. The rationale for such challenges is that a juror who admits to such views implicitly admits that he or she will be unable or unwilling to obey jury instructions on the penalty decision because the instructions will require the juror to at least entertain the possibility of a death sentence.

Several decades ago, defendants began to argue that automatically excluding opponents of the death penalty denied the defendant a fair cross section of eligible jurors and also ensured that the jury would be biased in favor of the prosecution at the guilt phase. Of course, with a separate decision on penalty, rendered after the guilty verdict and following the presentation of evidence bearing solely on the penalty, the court could reconfigure the jury after the guilty verdict to exclude and replace those jurors who earlier stated their categorical refusal or inability to impose death.
The defendant in *Witherspoon v. Illinois* (1968) had made this very argument to the Supreme Court with the support of some empirical evidence. The *Witherspoon* Court, treating a pre-*Gregg* death penalty statute, acknowledged that the claim was plausible in theory, but the Court found insufficient evidence that death-qualified juries would be more prone to convict defendants than juries that had been seated without regard to whether its members supported the death penalty. But rather than shut the door on the issue, the Court seemed to invite death penalty opponents to reopen the issue if they developed more and better evidence.

When the death penalty was reinstated under new statutes that clearly separated the guilt phase of the trial from the penalty phase, death penalty opponents could again argue that there was no need for death qualification of the jury that decided guilt or innocence, and they set about making a firmer social science case for the conviction-proneness of death-qualified juries than the case they been able to present in *Witherspoon*.

Relying on a number of new empirical studies done with sophisticated survey and jury simulation techniques (e.g., Cowan et al. 1984), the defendant in *Lockhart v. McCree* made a compelling claim that jurors who would not impose the death penalty tended on more general questions of guilt or innocence to lean more heavily to the side of leniency or not-guilty verdicts. Moreover, he presented evidence that the very process of death qualification might itself incline jurors toward guilty verdicts (Haney 1984).

But despite the quality of the empirical data, the Court rejected the argument that this phenomenon violated the Sixth Amendment requirement of a fairly selected and impartial jury. The Court criticized most of the research cited by the defendant for problems peculiar to each study but never acknowledged the power of the body of the research taken as a whole nor the fact that all the research presented to it pointed in the same direction. Not only was the Court not persuaded by the claims of unfairness, but it also took the view that the only solution under the new death penalty statutes would be to convene a second, separate jury for the penalty phase, a procedure it regarded as too cumbersome to impose on the states. Since then, the prosecution-proneness argument has itself become less visible; one federal trial judge recently held that newer studies had strengthened the earlier prosecution-proneness arguments enough to warrant requiring a second, separate jury in federal death penalty cases, although this ruling was then reversed by the Court of Appeals (*U.S. v. Green* 2005).

The Vagaries of Jury Behavior

The newest category of application of social science concerns some of the subtler processes by which capital jurors decide whether a defendant shall receive a life or death verdict. This research, even more clearly than the research in *Lockhart*, is a byproduct of the particular procedures mandated by the new post-*Gregg* laws. It also results from a now-healthy amount of data available about the operation of the hundreds of capital trials we have in the United States each year. A decade of
new studies, especially a remarkable set of papers done by a group at Cornell Law School (Theodore Eisenberg, Stephen Garvey, and Martin Wells), takes advantage of the opportunity to identify and interview large numbers of people who have actually served on capital juries, most notably in South Carolina, where the information has been developed by the Capital Jury Project. The Cornell researchers have constructed survey/questionnaire instruments to examine a variety of important post-\textit{Gregg} questions, such as how well jurors follow jury instructions and what conscious and perhaps unconscious factors influence their votes in capital cases. The jurors’ responses are translated into complex factor-coding, followed by advanced multiple regression analysis.

One paper (Eisenberg et al. 2001a) shows that the personal characteristics of jurors strongly influence their votes on the death penalty, with the dominant factors being race, religion, and general attitudes about the propriety of the death penalty for murder. Moreover, those factors play an especially dominant role on the first vote the jury takes, and that first vote usually determines the final one.

Another study (Eisenberg et al. 2001b) examines jurors who have indicated a willingness to impose the death penalty; it discovers that some of these jurors nevertheless harbor considerable concerns about its potential unfairness and would be prone to decline to impose the death penalty in favor of LWOP. The authors find these concerns especially evident among black jurors and Southern Baptists of both races.

An earlier study by the Cornell group (Eisenberg et al. 1996) examines whether jurors accept responsibility for their decisions. On the one hand, jurors may take the view that when they vote for a death sentence they are, in effect, merely affirming a result that naturally follows from the defendant’s egregious actions. On the other hand, they may believe that their vote is just one step, and not necessarily a major step, in a complicated legal process, so that the real responsibility for any ultimate death sentence rests more with, say, a judge or appellate court that will review their vote. But the law contemplates that jurors have primary responsibility and substantial discretion to determine whether the defendant should live or die, and this study sets out to measure jurors’ attitudes and understandings about their mandate. This study concludes somewhat optimistically that jurors on the whole appreciate the degree to which the criminal law actually does impose responsibility on them, although it suggests some reforms in the mechanisms by which that message is conveyed.

Still earlier research (Eisenberg & Wells 1993) examined juror understanding of the jury instructions they had received. It concludes that the jurors often do not appreciate the alternatives to a death sentence that state law provides them, and also that they do not understand the special burden-of-proof rules applied at the sentencing phase. On both these scores, this study suggests that the misunderstanding disfavors the capital defendant.

Conversely, a paper examining so-called victim-impact statements and their role in death sentencing (Eisenberg et al. 2003) offers somewhat more reassuring conclusions, at least for those who are dubious about the fairness of these statements.
Although victim-impact statements cause jurors to rate the murder victim higher on an admirability scale, the study finds little evidence that a higher rating on this scale increases the chances of a death sentence for the defendant. Rather, the possible effect of a victim-impact statement is largely mooted by the jurors’ reaction to more facts about the killing itself. A parallel study by Sundby (2003), using California data, found somewhat stronger evidence that the perceived character of the victim influences the sentence. When asked fairly abstract questions, surveyed jurors tended to deny that the character of the victim would make much difference. But their descriptions of actual jury deliberations suggested an important victim-evidence effect along two specific dimensions. Jurors are significantly more inclined toward a death verdict when the victim was engaged in ordinary activity and was chosen randomly by the killer, whereas they tilted toward leniency if the victim had exhibited antisocial or deviant behavior, even if that behavior in no way mitigated the killer’s culpability for the crime by providing some defense.

Even if these and similar studies do not lead to global constitutional litigation over the legitimacy of the death penalty, they offer the prospect of usefully educating judges and even legislators into ensuring that the promise of the so-called guided discretion statutes approved in Gregg is fulfilled.

But regarding the perceived legitimacy of capital punishment, the increasing number of DNA acquittals during the past decade may be more important than these studies. These DNA acquittals include acquittals of many prisoners on death row and prisoners serving life sentences who would have been on death row had their alleged crimes not been committed during the period of the death penalty moratorium (Scheck et al. 2000). It is the sheer number of these cases that is so striking. Although death penalty supporters may have acknowledged, in principle, that mistakes can happen in death penalty cases, the general assumption before these DNA acquittals seems to have been that the legal system took its greatest care in cases in which execution was possible, and the chance of error in capital cases was in fact minimal. We now know this is not so, and we can put faces on people who, but for the system’s slowness and their own good luck, might well have been executed for murders they did not commit. This human reality may prove to be more important than masses of social science evidence on deterrence and other issues in determining support for the death penalty in the long run.

The Annual Review of Law and Social Science is online at http://lawsocsci.annualreviews.org

LITERATURE CITED


Furman v. Georgia 408 U.S. 238 (1972)
Leamer E. 1983. Let’s take the con out of econometrics. Am. Econ. Rev. 73:31–43
Levitt S. 2004. Understanding why crime fell in the 1990s: four factors that explain the decline and six that do not. J. Econ. Perspect. 18:163–90
U.S. v. Green, U.S. App. LEXIS 8236 (1st Cir. 2005)
Witherspoon v. Illinois, 391 U.S. 510 (1968)
CONTENTS

COMING OF AGE: LAW AND SOCIETY ENTERS AN EXCLUSIVE CLUB,
Lawrence M. Friedman 1

THE COMPARATIVE STUDY OF CRIMINAL PUNISHMENT,
James Q. Whitman 17

ECONOMIC THEORIES OF SETTLEMENT BARGAINING,
Andrew F. Daughety and Jennifer F. Reinganum 35

LAW AND CORPORATE GOVERNANCE, Neil Fligstein and Jennifer Choo 61

TRANSCATIONAL HUMAN RIGHTS: EXPLORING THE PERSISTENCE AND
GLOBALIZATION OF HUMAN RIGHTS, Heinz Klug 85

EXPERT EVIDENCE AFTER DAUBERT, Michael J. Saks
and David L. Faigman 105

Plea Bargaining and the Eclipse of the Jury, Bruce P. Smith 131

The Death Penalty Meets Social Science: Deterrence and
Jury Behavior Under New Scrutiny, Robert Weisberg 151

Voice, Control, and Belonging: The Double-Edged Sword of
Procedural Fairness, Robert J. MacCoun 171

Law, Race, and Education in the United States, Samuel R. Lucas
and Marcel Paret 203

Law Facts, Arthur L. Stinchcombe 233

Real Juries, Shari Seidman Diamond and Mary R. Rose 255

Feminism, Fairness, and Welfare: An Invitation to Feminist
Law and Economics, Gillian K. Hadfield 285

Criminal Disenfranchise ment, Christopher Uggen, Angela Behrens,
and Jeff Manza 307

After Legal Consciousness, Susan S. Silbey 323

Why Law, Economics, and Organization? Oliver E. Williamson 369

Reversal of Fortune: The Resurgence of Individual Risk
Assessment in Criminal Justice, Jonathan Simon 397

INDEX

Subject Index 423