

# Some Distribution Patterns for the Georgia Death Sentence

*Arnold Barnett\**

*This Article proposes a classification procedure for Georgia murder cases, aimed at illuminating how judges, juries, and prosecutors decide which convicted killers should be sentenced to death. The scheme is empirical in origin, arising from the study of over 600 actual cases. The Article considers what the classification rules imply on such subjects as proportionality review, allegations of racial bias in sentencing, and the relationship between sentencing behavior and statutory guidelines.*

## INTRODUCTION

If proportionality review was at one time among the less prominent issues in the capital punishment debate, its obscurity ended abruptly on October 4, 1983. On that day, J.D. Autry, already strapped down and injected with sedatives, was about to be executed with lethal drugs by the State of Texas. At literally the last moment, the United States Supreme Court halted the execution, explaining that it would have to decide whether Texas was constitutionally required to conduct tests of proportionality of its capital sentences.<sup>1</sup>

The Supreme Court has since held that proportionality reviews,

---

\*Associate Professor, Sloan School of Management, Operations Research Center, Massachusetts Institute of Technology. B.A. 1969, Columbia College; Ph.D. 1973, Massachusetts Institute of Technology. This manuscript is far better because of the detailed and perceptive comments of Richard Van Duizend, Senior Staff Attorney, National Center for State Courts; Professor David Baldus, University of Iowa; Professor Samuel Gross, Stanford; Professor Joseph Kadane, Carnegie Mellon; Dr. Charles Murray, Manhattan Institute; Professor James Q. Wilson, Harvard; and Professor Frank Zimring, University of Chicago and University of California, Berkeley. Mary Elsner of the National Center for State Courts performed the tests of the classification model that are described in Appendix C. This research was performed under the Proportionality Review Project of the National Center for State Courts, Williamsburg, Va., and was partially supported by Grant No. 82-IJ-CX-0027 of the National Institute of Justice, U.S. Department of Justice.

<sup>1</sup> Pulley v. Harris, 460 U.S. 1036 (1983).

while certainly permissible, are not always mandatory.<sup>2</sup> But the decision did not render the issue moot. Over thirty states provide for some kind of proportionality review in their death penalty procedures.<sup>3</sup> And the kinds of comparative judgments embodied in such reviews are important in varied investigations about the workings of the death penalty (for example, whether race affects sentencing).

The notion animating proportionality review — one that has been explicitly endorsed by the Supreme Court<sup>4</sup> — is that death sentences cannot be imposed in an arbitrary manner. It is considered objectionable if a given defendant is put to death while, in adjacent counties (or adjacent courtrooms), defendants in virtually the same situation are given prison terms.

Pursuant to that view, most states have taken steps meant to ensure uniformity in the imposition of the death penalty. Legislatures have prepared lists of aggravating and mitigating factors that judges, juries, and prosecutors must review in their decisions on homicide sentences. As a further precaution, murder trials are often divided into two phases: the first to determine guilt or innocence, and the second to set punishment for those convicted. Proportionality review is a retrospective test of whether such procedures are in fact avoiding capriciousness. When a death sentence is handed down, one looks at the outcomes in a series of similar cases; unless death was the penalty in an appreciable fraction of these, the present sentence is deemed “excessive” (or “disproportionate”).

Defining the word “similar” in this context, however, is a most difficult task. Even two cases that coincide on the primary factual dimension (for example, the robbery-killing of a grocer) might differ substantially on others (for example, the defendant’s prior criminal record). Whether a given death sentence seems disproportionate can depend

---

<sup>2</sup> *Pulley v. Harris*, 104 S. Ct. 871 (1984). The Court stated, however, that proportionality review might be required if alternative checks on arbitrariness in a state’s death sentencing were inadequate. *Id.* at 880.

<sup>3</sup> See Brief for Respondent at App. A, *Pulley v. Harris*, 104 S. Ct. 871 (1984). These states include Alabama, Connecticut, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Carolina, Oklahoma, Pennsylvania, South Carolina, Tennessee, Virginia, Washington, and Wyoming. The relevant statutes are listed in Baldus, Pulaski & Woodworth, *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 J. CRIM. L. & CRIMINOLOGY 661 (1983).

<sup>4</sup> See, e.g., *Zant v. Stephens*, 462 U.S. 862, 879 (1983); *Lockett v. Ohio*, 438 U.S. 586, 601 (1978) (quoting *Gregg v. Georgia*, 428 U.S. 153, 188 (1976)); *Gardner v. Florida* 430 U.S. 349, 361 (1977) (plurality opinion); *Jurek v. Texas*, 428 U.S. 262, 276 (1976) (plurality opinion).

crucially on which cases are held comparable to the one on review.

One could partition homicide cases into "similar" clusters on the basis of some theory of jurisprudence. But the clusters would inevitably reflect value judgments about the function of the death penalty and the relative culpability of various defendants. Thus, two thoughtful individuals could easily devise vastly different systems of categories.

In such a situation, it is useful to examine how capital sentencing guidelines are actually being interpreted. Observing the distinctions that juries and others are making might suggest an operational definition of "similarity" that, if nothing else, would at least have the virtue of reflecting contemporary community standards. And, if one cannot rationally distinguish those cases that evoked death sentences from those that did not, one substantiates the fear that capital punishment cannot be applied consistently.

A major ancillary benefit of such an empirical exercise might be greater understanding of the racial patterns in death sentencing. In overall statistics, the fraction of white-victim slayings that end in death sentences is considerably higher than the comparable fraction for blacks.<sup>5</sup> But does this discrepancy reflect racial prejudice or, instead, legitimate distinctions that are coincidentally correlated with race? This important question might be easiest to consider within a broader analysis of sentencing behavior.<sup>6</sup>

This Article tries to identify the primary stimuli to death sentences in modern-day Georgia. Data and narrative summaries about hundreds of murder cases — all of them tried under Georgia's current death penalty statute — were prepared under the supervision of Professor David Baldus, who made them available to the Proportionality Review Project of the National Center for State Courts.<sup>7</sup> This Article will consider the circumstances and verdicts of about 600 such cases, trying to infer a set of classification rules that, roughly speaking, divide the cases into homogeneous subsets within which all killings are viewed as equally deathworthy.

Our methods of analysis and various findings will be described in

---

<sup>5</sup> See Gross & Mauro, *Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization*, 37 STAN. L. REV. 27, 55 Table 1, 94 Table 30 (1984).

<sup>6</sup> "The question of potential racial bias in sentencing was acknowledged by the United States Supreme Court last December when it granted a stay of execution to a prisoner in Georgia so the issue of discrimination in Georgia's sentencing could be studied," N.Y. Times, July 9, 1984, at A8, col. 3.

<sup>7</sup> In addition, Professor Baldus sent me scores of other case records so that I had a complete data file.

detail in the remainder of this Article. Briefly, I devised a classification scheme that seems fairly well to describe Georgia death-sentencing patterns. But to summarize the scheme in a few words risks oversimplifying those patterns; thus, I will not try to do so here. Nor will I risk diminishing the problems of this endeavor by racing through them in an Introduction.

A quick summary of this Article goes as follows: the discussion starts with a brief review of both some recent literature and the Georgia homicide statute. Then, part III presents the rationale for and details of some "scoring rules" for Georgia murder cases. Part IV shows the relationship between case scores and the chances of death verdicts. Thereafter, the Article devotes four sections to "Category 3," the subset of cases in which sentencing behavior seems least consistent. In particular, part VI explores whether racial bias is the source of the inconsistency.

The Article ends with a discussion of the limitations of this analysis, with a comparison of this study and the related work of Baldus, Pulaski, and Woodworth,<sup>8</sup> and with various tests of the viability of the classification model (Appendix C). Appendix A presents the scoring rules in their entirety; Appendix B provides several illustrations of their use.

## I. RECENT RESEARCH

This Article is not the first on death sentencing behavior, a point underscored by the fact that I am about to discuss seven others. Without exception, those other studies are both thoughtful and worthwhile; to the extent that these brief summaries tend to emphasize imperfections, it is as part of the explanation why the present paper is not superfluous.

Broadly speaking, empirical studies of capital sentencing can be described as either "classical" or "exploratory." Classical efforts begin by specifying in advance some possible determinants of sentencing behavior. Then, through some statistical method, they calculate the frequency of capital verdicts as a function of these determinants. Implicitly, such studies estimate the effect on the sentencing outcomes of each of the explanatory factors considered.

The exploratory studies give the data more latitude to "speak for themselves." In an initially unstructured manner, their authors peruse the summaries of a variety of homicide cases. Then, they try to characterize how the relatively few trials that led to death sentences differed

---

<sup>8</sup> Baldus, Pulaski & Woodworth, *supra* note 3.

in their details from the others.

The noteworthy classical papers on recent death-sentencing patterns include those of Radelet,<sup>9</sup> Zeisel,<sup>10</sup> Bowers and Pierce,<sup>11</sup> and Gross and Mauro.<sup>12</sup> All these scholars were concerned primarily with racial disparities in the imposition of the punishment. Statistics from Florida were present in all four articles; Bowers and Pierce also examined data from Georgia, Ohio, and Texas, while Gross and Mauro studied a total of eight geographically dispersed states.

Beyond considering the races of the defendant and the victim, the authors subdivided their cases on a few aggregate dimensions. Bowers and Pierce<sup>13</sup> and Zeisel<sup>14</sup> only inquired whether the slaying was a felony killing (one committed in connection with a separate felony). Radelet employed a victim-status dichotomy, based on whether the person killed was a close friend, lover, ex-lover, or family member of the accused.<sup>15</sup> After several cross-tabulations of death-sentencing rates by race and one other variable, Gross and Mauro culminated their work with a logit regression analysis that included five nonracial factors:<sup>16</sup>

- (1) Whether the homicide was committed in conjunction with another felony;
- (2) Whether the victim and defendant knew one another;
- (3) Whether there were one or several victims;
- (4) Whether there were any females among the victims;
- (5) Whether the killing was committed by gun.

These papers uniformly suggest that, as applied, capital punishment is largely restricted to killers whose victims are white. The "race-of-victim" effect depicted is both dramatically large and statistically significant. Certain secondary racial patterns arise in some of the papers more than in others. Two articles, for example, imply that black defendants faced higher death risks than similar white killers.<sup>17</sup>

While reading such studies is most stimulating, one is struck by the

---

<sup>9</sup> Radelet, *Racial Characteristics and the Imposition of the Death Penalty*, 46 AM. SOC. REV. 918 (1981).

<sup>10</sup> Zeisel, *Race Bias in the Administration of the Death Penalty: The Florida Experience*, 95 HARV. L. REV. 456 (1981).

<sup>11</sup> Bowers & Pierce, *Arbitrariness and Discrimination under Post-Furman Capital Statutes*, 26 CRIME & DELINQ. 563 (1980).

<sup>12</sup> Gross & Mauro, *supra* note 5.

<sup>13</sup> Bowers & Pierce, *supra* note 11, at 589.

<sup>14</sup> Zeisel, *supra* note 10, at 459.

<sup>15</sup> Radelet, *supra* note 9, at 921.

<sup>16</sup> Gross & Mauro, *supra* note 5, at 78 n.127.

<sup>17</sup> Bowers & Pierce, *supra* note 11, at 595; Zeisel, *supra* note 10, at 466.

sweep of their simplifying assumptions. Perhaps it was unavoidable, but some distinctions of potentially great significance are entirely absent from the calculations. Even the comparatively detailed Gross and Mauro piece, for example, gives a case the same rating whether the felony in a felony killing was robbery or firebombing, whether the victim was the defendant's spouse or the defendant's clergyman, or whether the deceased was stabbed with a knife or mutilated with an icepick. None of the researchers considered the defendant's prior record or the possibility of self-defense. And strength-of-evidence is deemphasized to the point that even cases in which the defendant was acquitted remain with unknown consequence in the data bases.

One wonders whether, in their cumulative effect, such omissions compromise the conclusions of the studies. Gross and Mauro concede that killings against whites are on average more "aggravated" than those against blacks.<sup>18</sup> Is it not conceivable that, within the broad categories of homicides they lump together as similar, those with white victims are likewise more aggravated? Radelet acknowledges that the "strength of the racial disparities observed in this study will fluctuate as other potentially relevant variables are introduced."<sup>19</sup> But until such factors have been given appropriate weight, how can one be sure that their effects are mere fluctuations? These studies, in short, raise serious and troubling questions about racial bias in sentencing. But that they answer them unequivocally seems less clear.

Among the exploratory studies, one of the most distinguished was performed by Kalven and Zeisel<sup>20</sup> two decades ago. They considered thirty-five homicide cases from the early 1960's in which the judge or the jury favored a death sentence. In fully twenty-one of the thirty-five, the judge and jury split on whether the defendant should be executed.

Case-by-case reviews suggested that when a killing involved multiple victims, sexual torture, or a bizarre weapon, a consensus for death was especially likely. In the twenty-one disputed cases, the factors that evoked leniency included the defendant's not being the actual killer, a lover's triangle aspect, an earlier noncapital verdict for another participant in the slaying, and a "worthless" victim. But to Kalven and Zeisel, the patterns just cited are less important than another. When judges and juries hearing the same case under the same statute so often disagree on the life-or-death decision, consistency in the selection of those to be executed might be a farfetched concept.

---

<sup>18</sup> Gross & Mauro, *supra* note 5, at 100.

<sup>19</sup> Radelet, *supra* note 9, at 926.

<sup>20</sup> H. KALVEN, JR. & H. ZEISEL, *THE AMERICAN JURY* 434 (1966).

Zimring, Eigen, and O'Malley<sup>21</sup> studied Philadelphia homicides in the first few months of 1970. Only three such killings led to death sentences; all involved the slaying of whites by black strangers. Two of these killings were felony-murders, one was committed by grotesque means (a hacksaw and a sledgehammer), and one was clearly a manifestation of racial hatred. The authors recognized that these murders were different from most others, but noted an inconsistency between their death verdicts and the outcomes for other aggravated killings (for example, multiple murders) as well as those for other defendants in the three cases themselves.<sup>22</sup>

In the post-*Furman* period, an exploratory study of Dade County, Florida was performed by Arkin.<sup>23</sup> He compared ten felony killings that led to capital sentences (three of them given to the same defendant) with forty-four felony killings that resulted in prison terms. The death cases, he argued, were especially aggravated; they were clearly distinguishable from a majority of the others and at least somewhat different from thirty-eight of the forty-four. Thus, Arkin saw considerable (though not perfect) selectivity in Dade County's death sentencing and stated that any inference of racial discrimination "collapses" under scrutiny.<sup>24</sup>

A potential problem with these exploratory studies relates to their small sample sizes. The danger is that the discussion of the cases, however illuminating, could be construed as more speculative than systematic. For example, Gross and Mauro, having noted that Zimring, Eigen, and O'Malley had "only three death sentences," went on to contradict Arkin, contending that "the small size of his sample — ten death penalties in all — precludes definite conclusion on the existence of racial discrimination."<sup>25</sup>

Taken together, these seven classical and exploratory efforts suggest a clear pattern in this literature. When the number of cases is small — and statistical procedures would founder on the paucity of data — the

---

<sup>21</sup> Zimring, Eigen & O'Malley, *Punishing Homicide in Philadelphia: Perspectives on the Death Penalty*, 43 U. CHI. L. REV. 227 (1976).

<sup>22</sup> The authors also compared cases that led to life imprisonment with the more numerous others that yielded far shorter terms. They discerned a race-of-victim effect, and contrasted the great disparities in sentence lengths with the seemingly lesser differences in the homicides themselves.

<sup>23</sup> Note, *Discrimination and Arbitrariness in Capital Punishment: An Analysis of Post-Furman Murder Cases in Dade County, Florida 1973-1976*, 33 STAN. L. REV. 75 (1980).

<sup>24</sup> *Id.* at 88.

<sup>25</sup> Gross & Mauro, *supra* note 5, at 43.

intuitive exploratory approach is viewed as legitimate. But as the sample size reaches into the hundreds, more formal methods involving broad, prespecified categories have usually been brought to bear. The result is something of an inverse relationship between the number of cases studied and the level of detail per case, which can lead to the persistent sense that something is missing.

Against the backdrop of the research just cited, the work of Baldus, Pulaski, and Woodworth<sup>26</sup> is most important. For various Georgia killings from the 1970's that resulted in murder trials, the authors gathered both narrative summaries and data that concern roughly 200 variables. And this vast information was obtained not for a few but for several hundred cases. This exceptional data set provides the "raw material" for the present analysis.

More precisely, this Article presents a reanalysis, for Baldus, Pulaski, and Woodworth made extensive calculations with the data that they collected. Of the two studies, as we shall see, theirs is closer to the classical tradition while this one is more in the exploratory spirit. Forthcoming sections will discuss both efforts in some detail; prior to any such endeavor, however, it would seem wise to review the Georgia statute that is the basis of all discussion.

## II. SOME HISTORICAL BACKGROUND

The present era in death sentencing began on June 29, 1972 when the United States Supreme Court, in a five to four decision, struck down the capital statutes of Georgia and Texas (and, by implication, those of all other states).<sup>27</sup> While two Justices found that capital punishment is inherently unconstitutional,<sup>28</sup> the three others in the majority found that, as applied, the death penalty violated the eighth amendment prohibition against cruel and unusual punishment.<sup>29</sup> The relatively few murderers put to death were chosen in so "freakish" and "arbitrary" a manner, and the discretion allowed to judges and juries seemed so unbridled that, to Justice Potter Stewart, "this death penalty is cruel and unusual in the same way that being struck by lightning is cruel and unusual."<sup>30</sup>

In the next four years, over thirty states revised their capital punish-

---

<sup>26</sup> Baldus, Pulaski & Woodworth, *supra* note 3.

<sup>27</sup> *Furman v. Georgia*, 408 U.S. 238 (1972).

<sup>28</sup> *Id.* at 257 (Brennan, J., concurring); *id.* at 314 (Marshall, J., concurring).

<sup>29</sup> *Id.* at 240 (Douglas, J., concurring); *id.* at 306 (Stewart, J., concurring); *id.* at 310 (White, J., concurring).

<sup>30</sup> *Id.* at 306-10 (Stewart, J., concurring).



ment laws so as to reduce the degree of randomness in the distribution of death sentences. But on July 2, 1976, the Supreme Court in *Woodson v. North Carolina*<sup>31</sup> struck down a North Carolina statute that provided mandatory executions of convicted first-degree killers. The Court held that the law was too inflexible to be just, because it barred juries from considering "compassionate or mitigating factors" stemming from the "diverse frailties of humankind."<sup>32</sup> The Court speculated that, rather than sentence certain defendants to death, juries would acquit them even if convinced of their guilt.

Having ruled out both too little and too much discretion, the Court indicated what kind of "intermediate" statute it would find acceptable. In *Gregg v. Georgia*,<sup>33</sup> decided on the same day as *Woodson*, the Court upheld Georgia's revised death-sentencing law. With approval, the Court noted that juries would be "permitted to consider any aggravating or mitigating circumstances," yet would be required to "find and identify at least one statutory aggravating factor" before sentencing a murderer to death.<sup>34</sup>

The Georgia law that the Court upheld states that a defendant is "death-eligible" only if at least one of the certain aggravating conditions was present.<sup>35</sup> Unlike many other states, Georgia does not accom-

---

<sup>31</sup> 428 U.S. 280 (1976).

<sup>32</sup> *Id.* at 304.

<sup>33</sup> 428 U.S. 153 (1976).

<sup>34</sup> *Id.* at 206.

<sup>35</sup> The Georgia statute allows a death sentence to be imposed if one of the following circumstances is present:

- (1) The offense of murder was committed by a person with a prior record of convictions for murder, armed robbery, kidnapping or rape, or the offense of murder was committed by a person who has a substantial history of serious assaultive criminal convictions.
- (2) The offense of murder was committed while the offender was engaged in the commission of rape, armed robbery, or kidnapping, or aggravated battery, or the offense of murder was committed while the offender was engaged in the commission of burglary or arson in the first degree.
- (3) The offender by his act of murder, knowingly created a great risk of death to more than one person in a public place by means of a weapon or device which would normally be hazardous to the lives of more than one person.
- (4) The offender committed the offense of murder for himself or another, for the purpose of receiving money or any other thing of monetary value.
- (5) The murder of a judicial officer, former judicial officer, district attorney or solicitor or former district attorney or solicitor during or

pany these aggravating factors with a list of statutory mitigating factors such as the defendant's youth or the absence of a criminal record. Thus, a Georgia jury<sup>36</sup> is under no obligation to treat any aspect of a case as mitigating. On the other hand, the jury is not compelled to give a death sentence even if many aggravating circumstances are present.

In other words, the new law does not foreclose the possibility that the same case could elicit different responses from different juries. In upcoming sections, this Article explores how the Georgia statute is actually being implemented.

### III. FRAMEWORK OF THE STUDY

#### A. *Choosing a Methodology*

The focus of this study is the general question: given the circumstances of a particular murder case, what is the probability the perpetrator will be sentenced to death? We are striving for an empirical answer based on the details and outcomes of a large number of Georgia cases. This Article's approach deviates from the most familiar social science paradigm, a circumstance that should be discussed at the outset.

Traditionally, those with such rich data as we would create  $N$  numerical variables ( $x_1, x_2, \dots, x_N$ ) that, taken together, summarize the facts of the case. For example,  $x_j$  could be the defendant's number

---

because of the exercise of his official duty.

- (6) The offender caused or directed another to commit murder or committed murder as an agent or employee of another person.
- (7) The offense of murder was outrageously or wantonly vile, horrible or inhuman in that it involved torture, depravity of mind, or aggravated battery to the victim.
- (8) The offense of murder was committed against any peace officer, corrections employee or fireman while engaged in the performance of his official duties.
- (9) The offense of murder was committed by a person in, or who has escaped from, the lawful custody of a peace officer or place of lawful confinement.
- (10) The murder was committed for the purpose of avoiding, interfering with, or preventing a lawful arrest or custody in a place of lawful confinement, of himself or another.

GA. CODE ANN. § 27-2534-1 (1983).

The statute describes treason and hijacking as always susceptible to the death penalty. But no cases involving either are considered in this Article.

<sup>36</sup> Throughout this Article, the word "jury" is used as a generic term for the sentencing authorities. In reality, Georgia prosecutors are unusually influential in the process; in a considerable fraction of cases, the state simply waives the death penalty. And judges obviously participate in the decisions through their sentencing instructions.

of prior convictions for violent crimes, while  $x_k$  could be an indicator variable equal to one if the defendant raped the victim and zero if he did not. Some of the factors would presumably pertain to the strength of the evidence presented in court.

Using some multivariate statistical technique (for example, logit regression), one could then develop a mathematical formula that estimates  $p$ , the probability the defendant gets a death sentence, from the values of the  $x_i$ 's that describe the case. The formula would be calibrated from the collective data about the actual trials. Through scrutinizing the mathematical expression that arises, one could, in theory, infer how juries are affected by the presence (or absence) of any given circumstance.

Such an approach is appealing in the abstract, but qualms can develop as one's thinking gets more specific. Any multivariate method for which a well-developed theory exists — meaning any of those on the standard computer packages — entails a series of strong assumptions. Unless these assumptions are accurate, computations that depend on them can yield highly misleading results. One could wind up discarding variables of real importance, while embracing others that are actually irrelevant.<sup>37</sup>

While some potential problems are rather technical, others are not at all abstruse. Two of particular interest here concern statistically correlated variables and the assumption of independent effects. Correlation arises because, in a series of homicide cases, certain features might tend to arise in tandem. Every deliberate drowning, for example, might be preceded by the kidnapping of the victim. In every slaying of a bank teller, the victim might be white. When two variables tend to "move the same way" within the data set, it is hard to tell whether one of them is responsible for combined effect and, if so, which, or whether both of them contribute and, if so, in what proportions.

Moreover, simple models assume that the various factors in a case independently exert influence on the jury's decision. But the very weight accorded a particular circumstance could well depend on which others are present. Killing by strangulation, for example, might in general increase the chance of a death verdict. But if a man, arriving home to find his wife in bed with another, proceeds to strangle the intruder,

---

<sup>37</sup> The author has argued that such problems could well have afflicted studies on job discrimination and on the deterrent effect of capital punishment. See Barnett, *An Underestimated Threat to Multiple Regression Analyses Used in Job Discrimination Cases*, 5 *INDUS. REL. L.J.* 156 (1982); Barnett, *The Deterrent Effect of Capital Punishment: A Test of Some Recent Studies*, 29 *OPERATIONS RESEARCH* 346 (1981).

his mode of killing might heighten the jury's belief that he acted in the thrall of uncontrollable passion. Hence, the very same factor could be deemed aggravating in some situations and mitigating in others.

While there are multivariate techniques aimed at coping with such problems, their success in the present context cannot be guaranteed. If two variables are correlated, one can try to partition their joint effect between them. But the procedure can give unstable answers and is imperiled by the violation of any of several technical assumptions. And if the effect of two variables taken together differs from the sum of their individual effects (for example, strangulation and lovers' triangle), the use of "interaction terms" in the model could reflect this. But if there are (say) forty original variables, there are 780 possible pairwise interactions and 9880 possible three-way interactions. Thus, the model could become unwieldy and its data requirements enormous.

For these reasons (and many others like them), the formal statistical methods might not be ideally suited to the problem we are studying. We could be in a situation in which mathematical complexity and deeper understanding, far from being synonymous, could be negatively correlated. Thus, it is not to favor the horse over the locomotive to conjecture that a person, working with actual case summaries and using common sense, *might* gain greater insight into jury behavior than a computer that processes the data mechanically. With this possibility in mind, I begin discussing in the next section a fairly simple approach for analyzing the case records.

### B. *A Procedure for Classifying Georgia Homicide Cases*

The available data consist of narrative summaries of over 600 homicide cases, accompanied by computer-coded information about the defendant, the victim(s), the circumstances of the killing, and the verdict reached.<sup>38</sup> These cases concern all offenders who were arrested and charged with murder in Georgia between March 28, 1973 and June 30, 1978 and who later received a life or death sentence after trial or were sentenced to death after pleading guilty to murder.<sup>39</sup> About one-

---

<sup>38</sup> I focus on the verdict in the initial trial, believing that the most useful reflection of prevailing community standards. However, virtually all the death sentences have been appealed; some have been reversed, and almost all others are still being considered. Although the appeals and reversals clearly raise troubling questions, I will not consider them in this study.

<sup>39</sup> We will not be considering the "filtering" process by which only a minority of solved killings lead to murder charges. That process could in itself be somewhat arbitrary, but it is not the subject of this inquiry.

sixth of the offenders received death sentences.

To get a sense of how the minority of cases that led to death penalties differed from the majority that did not, I started reading the summaries of blocks of cases of each type. In the initial overview, no attention was paid to the race of the defendant or the victim, to details about the victim's criminal history, or to the characteristics of the county in which the trial took place. However, these factors later entered the analysis.

The summaries made clear that homicide cases show immense variety. While certain elements are present in numerous records, the full constellation of circumstances in a given case is only rarely reproduced. Proportionality review, therefore, cannot realistically be based on comparisons across cases that have nearly identical facts.

In terms of the sentences meted out, however, certain consistencies did seem to come through. The death cases appeared to differ from the others on three primary dimensions:

- (1) The Certainty the Defendant is a Deliberate Killer;
- (2) The "Status" of the Victim;
- (3) The "Heinousness" of the Killing.

These dimensions will be defined shortly.

It seemed useful to classify the various cases under a simple numerical scheme, in which three integers — one for each of the dimensions — reflect what appeared to be the case's most salient elements. To make the procedure as objective as possible, a detailed set of classification rules was prepared; they are presented in their entirety in Appendix A.

Readers will naturally want to know how the objectivity of the classification procedure was established, what predictive power it achieved, and whether its underlying assumptions were proved viable. But first they will want to know just what the procedure is. The remainder of this section is devoted to that concern; issues of reliability are discussed in some detail in Appendix C.

### 1. The Certainty the Defendant is a Deliberate Killer

The word "certainty" refers to the degree of assurance that the accused was, in fact, the killer of the victim. (If substantial doubts existed, the defendant would presumably have been acquitted; the notion is, however, that the threshold of certainty needed for a death sentence is higher than that for a guilty verdict.) "Deliberateness" pertains to

whether, even assuming the defendant performed the killing, he<sup>40</sup> acted knowingly to cause the victim's demise.

On this dimension, the case is rated either zero (unusually low), one (average), or two (unusually high). Zero reflects a relatively weak case in terms of certainty and/or deliberateness. A case based solely on circumstantial evidence, for example, would deserve this rating. Zero would also apply if the defendant was not the triggerman, if he clearly suffered from mental illness, or if the details of the slaying hint that it was an accident. (As noted, the precise criteria for this and other ratings are set forth in Appendix A.)

A score of two, by contrast, signifies exceptionally strong evidence that the killing was not an isolated, aberrant act of passion or panic. If the defendant plotted the murder extensively, had previously tried to kill the victim, or was implicated in other killings, the case would be classified two.

If neither a zero nor a two is justified under the scoring rules, the case receives a rating of one. If, as happens rather rarely, criteria for *both* zero and two are satisfied, a score of one is also given. Among recent murder cases in Georgia, most seem to warrant this intermediate classification: of the 606 the author rated, sixty-eight percent scored one. (Twenty-six percent scored zero, and six percent scored two.)

## 2. The "Status" of the Victim

The "status" of the victim relates primarily to the relationship between the victim and the accused. Its presence in the classification scheme reflects an observed pattern under which, all other factors being equal, stranger-to-stranger killings are more "prone" to death sentences than those in which the victim knew the defendant.<sup>41</sup> The cases are scored either zero or one on the dimension of "status," with the latter number suggesting a higher chance of a death verdict.

The zero/one dichotomy is close but not identical to the stranger/nonstranger split. A stranger who acted in a highly provocative way just prior to his killing, or was engaged in an illegal enterprise, would call forth a rating of zero. And even if the defendant and victim were acquaintances, the score one would be appropriate if the person was slain in his official capacity (for example, as the defendant's supervisor in a factory, or a teller in a bank being robbed).

---

<sup>40</sup> The pronoun "he" is used because the heavy majority of both killers and victims are male.

<sup>41</sup> This notion is hardly original with this study. The point is that the case records are in accord with an intuitive and widely recognized pattern.

This status dimension has little explicit basis in law. It is hard to avoid speculating that, in killings in which jurors can imagine themselves or their loved ones as victims, death penalties are more likely to be imposed. To say this is not to impute cynicism to the juries; when a case evokes genuine fear, considerations of deterrence may more greatly affect the sentencing decision than otherwise.

### 3. The Heinousness of the Killing

On this dimension, the case is scored zero, one, or two, depending on the answers to the questions: "Could the killing be construed as an act of self-defense?" and "Was the killing vile?" The criteria for self-defense are quite stringent, requiring a clear mortal threat to the defendant or his loved ones. Among vile slayings are those with multiple victims, those preceded by psychological torture or sexual abuse, and those involving bizarre weapons or mutilated bodies.

A killing in self-defense that is not vile scores a zero, while a vile murder unrelated to self-defense scores a two. All other homicides are assigned the rating one; a one generally reflects the absence of both vileness and self-defense.

In summary, each case is classified with three separate numerical ratings. This study will use the notation (i, j, k) with "i" the score on "certainty," "j" on "status," and "k" on "heinousness." "i" and "k" can take on any of three values, while "j" takes on two; thus there is a total of  $3 \times 2 \times 3 = 18$  possible classifications. The individual ratings are arranged so that, the higher the score on a given dimension, the greater seems the empirical risk of a death sentence.

The least "deathworthy" score would be (0,0,0), which could arise if the defendant knew the victim and killed him in self-defense in a manner seeming somewhat accidental. At the other extreme is a (2,1,2) case, such as the murder-for-hire of three police officers. Most robbery killings of a merchant would be classified (1,1,1); if, however, the victim took out a gun and fired at the defendant, (1,1,0) would probably be appropriate.<sup>42</sup>

The next section presents the possibility of a death verdict as a function of a case's classification. But various details, implications, and limits of the classification scheme are not apparent from the preceding brief description; for continuity, I postpone discussing them to parts

---

<sup>42</sup> Appendix B illustrates how four actual cases were rated. The reader might do well at least to skim Appendices A and B now, lest the classification procedure seem needlessly obscure.

## VIII-X.

## IV. DEATH SENTENCING RATES BY CASE CLASSIFICATIONS

Some initial death sentencing rates appear in Table 1, where the cases are partitioned based on their (i, j, k) vectors. For convenience, I place in the same column all classes of cases with the same value of  $i+j+k$ . This arrangement is consistent with a general expectation that, as  $i+j+k$  goes up, so does the chance of a death sentence. But I am *not* suggesting that the content of the three ratings can be represented by their simple sum. A (1,1,1) killing differs in substantial respects from a (1,0,2), and no theoretical reason exists for assuming that their death sentencing rates will be equal.

Table 1 reflects only an initial "sort" of the murder cases. As we will see, other data allow us to refine and clarify certain provisional numbers. Even without such elaboration, however, the table conveys a good deal of information.

TABLE 1: DEATH-SENTENCING RATE AS A FUNCTION OF CERTAIN DETAILS OF THE KILLING

		(1,0,1) .01 (184)	(2,0,1) .09 (11)		
(0,0,0) 0 (8)	(1,0,0) .02 (43)	(1,1,0) 0 (10)	(1,1,1) .25 (60)	(1,1,2) .81 (73)	(2,1,2) .88 (8)
	(0,0,1) 0 (55)	(0,1,1) .02 (48)	(0,1,2) .26 (21)	(2,0,2) .56 (18)	
		(0,0,2) .04 (23)	(1,0,2) .29 (42)		

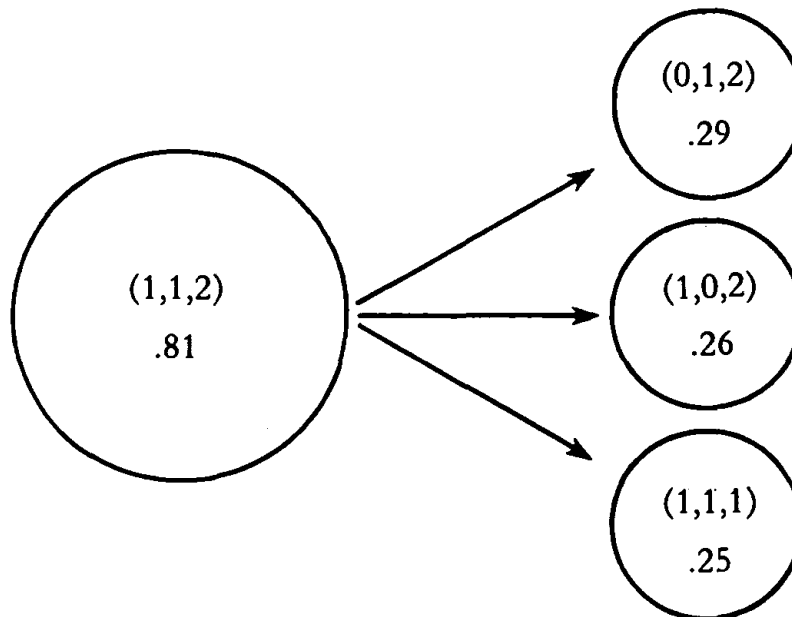
**NOTE:** Numbers in parentheses refer to the number of cases in the category. In two possible categories — (2,1,0) and (2,1,1) — there were no cases at all. (2,0,0) and (0,1,0) had only one case apiece; neither ended in a death verdict.



If the classification procedure had no discriminatory power, one would expect the death rates in all categories to hover around 19 percent (the overall rate for the cases considered). But Table 1 indicates that, as anticipated, there is clear positive relationship between  $i+j+k$  and the risk of a death verdict. Of the 373 cases in which  $i+j+k$  (hereafter defined as  $s$ ) does not exceed two, a mere *four* of them elicited death verdicts. The death rate rose to 25 percent for the cases for which  $s=3$ , and when  $s \geq 4$ , fully 78 percent of the defendants were condemned to death.

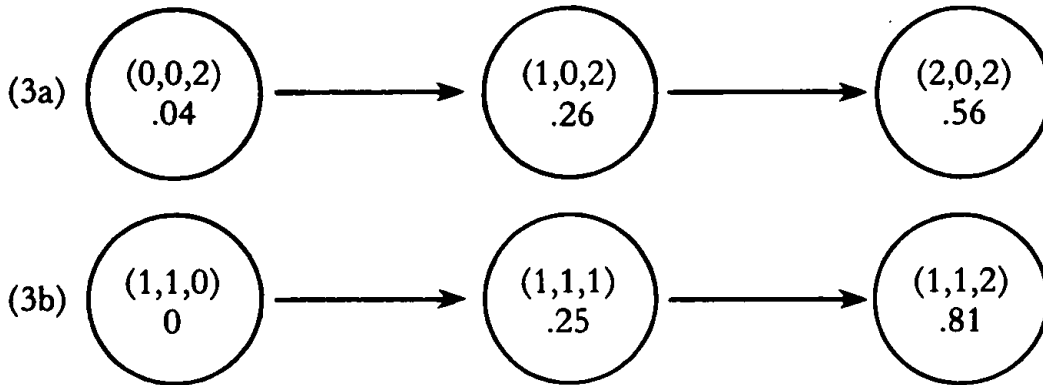
Indeed, it is instructive to start at the category (1,1,2) and then, in three separate maneuvers, to reduce each of  $i$ ,  $j$ , and  $k$  by one while holding constant the other two variables. The result, reiterated in Table 2, is a consistently drastic drop in the capital punishment rate. Such statistics lend strong support to our hypothesis that each of the three dimensions is of major importance in its own right.

TABLE 2: DROP IN THE DEATH-SENTENCING RATE WHEN ONE OF THREE CASE RATINGS IS REDUCED



In Table 3a,  $j$  and  $k$  are held constant while  $i$  spans the values from zero to two. Table 3b displays some analogous data when  $k$  is varied across its range. The steady rises in death rates justify the decision to let  $i$  and  $k$  take on three values rather than just two.

TABLE 3: THE EFFECT OF VARYING  $i$  AND THEN  $k$  WHILE HOLDING FIXED THE OTHER TWO PARAMETERS.



Of course, the classification rules were largely developed from the very data on which they were tested. Thus, their success identifying key patterns is not altogether surprising. But it was not at all foreordained that a simple numerical scheme would prove so effective. That outcome was contingent upon (and, indeed, shows the existence of) a considerable degree of regularity in the Georgia sentencing decisions.

The entries in Table 1 serve to focus our attention on Category 3: those classes of cases for which  $s=3$ . There the death rates, although low, are nonetheless well above zero. Before discussing this category further, we might do well to remind ourselves what kinds of killings it contains. Typical cases in the subdivision of Category 3 might be:

- (1,1,1) The killing of an unarmed grocer with a single shot during a robbery.
- (2,0,1) The extensively plotted — though not especially sanguinary — killing of one's spouse for economic motives.
- (1,0,2) The killing of a long-term personal enemy through holding her head below water.
- (0,1,2) The kidnap-murder of a stranger, in which the defendant beat the victim, but did not fire the shot that killed him.

Given the differences just outlined, it was not obvious that the death-sentencing rate would be fairly stable across Category 3. Yet, the observed rates in (1,1,1), (0,1,2), and (1,0,2) are very close together and, while the (2,0,1) rate is lower, its deviation from the category-wide average of 25 percent is nowhere close to statistically significant. Thus, in terms of actual sentencing outcomes, Category 3 can be viewed as quite

homogeneous. This circumstance is something of a "stroke of mathematical luck"; it is an empirical finding I did not anticipate and *not* a condition imposed in advance on the analysis.

More important, an obvious question arises about Category 3. Why is it that, for every such killing that leads to a death sentence, there are three others that do not? The 25 percent figure could reflect oversimplifications in the classification rules, genuinely inconsistent behavior by different juries, or invidious distinctions (for example, by race) that effectively divide otherwise homogeneous classes of murders. But the explanation could also be more benign, tied to the defendant's prior record or to regional differences in the adjudication of capital cases.<sup>43</sup> These varied possibilities are explored in the next few sections, starting with geography and criminal history.

#### V. THE ROLE OF REGION AND OF CRIMINAL RECORD

From the Census Bureau's characterization of the county in which it took place, Professor Baldus classified each murder trial as either "urban" or "rural." The underlying idea was that, if attitudes on the death penalty do vary across localities, it would most likely be apparent on an urban/rural dimension. The differing viewpoints between such regions have been prominent in many a statewide election; one might suspect a similar divergence on a punishment issue that, in recent years, has often been correlated with more general political views.

One could measure the defendant's prior record by some complex mathematical function of the number and nature of his past offenses, as well as the time he spent in prison. But trying to devise an appropriate formula is uninviting, especially because it is hard to imagine that any juries actually used it. Therefore, a simpler approach is adopted here.

For the present purposes, the issue of prior record is reduced to a simple yes/no question. The accused is said to have a serious prior record if he had been convicted for any felonies, or if, while his only convictions were for minor offenses, he served time in a Georgia prison. The rationale for the latter condition is that, given the lenient treatment generally accorded early offenders, those incarcerated for minor crimes presumably committed a great number of them. The defendant is said to have no serious prior record if (as happens quite frequently) he has no convictions or if, although he does, they involve neither felonies nor

---

<sup>43</sup> In *Williams v. Maggio*, 679 F.2d 381 (5th Cir. 1982), the Fifth Circuit sustained the constitutionality of Louisiana's system of circuitwide — rather than statewide — proportionality review. See also *Maggio v. Williams*, 104 S. Ct. 311 (1983) (order vacating stay of execution).

imprisonment.

With such definitions in hand, we can try to assess the role of region and criminal history in the death sentencing decision. Neither of these variables could be pivotal when  $s=i+j+k$  falls below two because, regardless of how such a case stands in these respects, a death verdict is exceptionally rare. More detailed analysis is necessary in Categories 3 and 4 (that is,  $s=3$  or 4), which have both appreciable death sentencing rates and nontrivial sample sizes.

As noted, the death rate shows no significant variation across the subdivisions of Category 3. Thus, it is not to blur salient distinctions to use overall Category 3 statistics in this discussion. The same cannot be said for Category 4 since there is a large difference in the death-sentencing rates for (1,1,2) and (2,0,2) killings. The latter class has too few cases for the present purpose (there are only *two* urban/prior record (2,0,2) killings); hence the Category 4 study is restricted to the (1,1,2) class.

TABLE 4: DEATH SENTENCING RATES BY LOCALITY AND DEFENDANT'S HISTORY: CATEGORY 3 HOMICIDES

	<u>URBAN</u>	<u>RURAL</u>
PRIOR RECORD	.29(28)	.36(28)
NO PRIOR RECORD	.17(30)	.21(48)

(Sample sizes in parenthesis.)

TABLE 5: DEATH SENTENCING RATES BY LOCALITY AND DEFENDANT'S HISTORY: CLASS (1,1,2) HOMICIDES

	<u>URBAN</u>	<u>RURAL</u>
PRIOR RECORD	.75(8)	1.00(28)
NO PRIOR RECORD	.58(19)	.72(18)

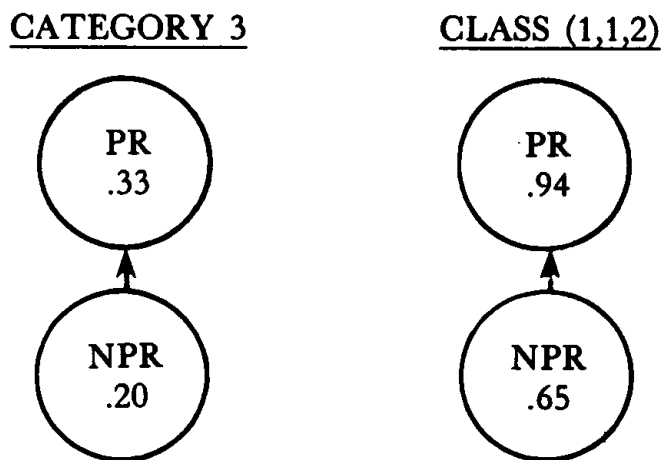
Tables 4 and 5 present the relevant data. In all of the four geographic comparisons made (record/no record, Category 3, class (1,1,2)), the urban areas had lower death sentencing rates than their rural counterparts. But the rural rates did not dwarf those from urban areas: when  $x$  is the urban death rate, the corresponding rural rate averages about  $1.25x$ . Hence, even if rural juries are "tougher" than those in cities, they only appear so to a limited extent.

Prior record seems to play a somewhat larger role in the sentencing

decision. For a given type of killing and type of county, a serious record raises a defendant's death risk by a factor averaging over 1.5. As Table 6 points out, a death sentence is almost certain in a (1,1,2) case unless the defendant had no record of serious trouble. From the entries in the table we can infer that, among the twelve (1,1,2) killers not sentenced to death, at least two-thirds have their "clean" prior records to thank.

Although of interest on their own, these findings only partially resolve the "puzzle" of Category 3. In every contingency depicted in Table 4, somewhere between two and six times as many defendants avoid death sentences as receive them. The next part will consider whether race might explain the discrepant outcomes.

TABLE 6: THE EFFECT OF PRIOR RECORD ON THE DEATH-SENTENCING RATE IN OTHERWISE COMPARABLE CASES



PR = Prior Record; NPR = No Prior Record

(Table 6 combines the urban and rural data from Tables 4 and 5.)

## VI. THE ROLE OF RACE

As noted earlier, almost all recent studies<sup>44</sup> on capital sentencing treat the race of the victim as a major explanatory factor. Some of these investigations concern the same period and state examined in this study. Gross and Mauro have estimated that, in post-*Furman* Georgia, killers of whites face 7.2 times the death sentencing risk of the killers of blacks.<sup>45</sup> Baldus, Pulaski, and Woodworth concluded that:

<sup>44</sup> See *supra* text accompanying notes 9-26.

<sup>45</sup> Gross & Mauro, *supra* note 5, at 79. This multiplier of 7.2 is already adjusted for

Georgia is operating a dual system, based on the race of the victim, for processing homicide cases. Georgia juries appear to tolerate greater levels of aggravation without imposing the death penalty in black victim cases; and . . . the level of aggravation in black victim cases must be substantially greater before the prosecutor will even seek a death sentence.<sup>46</sup>

The present review starts with some macroscopic statistics. In the 606 cases examined, the defendant was black in 59.7 percent. This fraction slightly exceeds the proportion of blacks among those defendants sentenced to death (53.2 percent). Thus, there is no immediate evidence of bias against black defendants. But, while 40.6 percent of the cases involved black victims, only 15.9 percent of the capital cases did so. From these last two figures, we deduce that a factor of 3.6 separates the death sentencing rate in white-victim cases from that in the others.

It is useful to disaggregate the murder cases according to their s-values, which, as we have seen, are clearly related to the rate of death verdicts. Starting with a partition suggested by Table 1, we observe the following patterns:

TABLE 7: DISTRIBUTION OF MURDER CASE CLASSIFICATIONS BY RACE OF *DEFENDANT*

<u>s-VALUE</u>	<u>PERCENTAGE OF BLACK DEFENDANT CASES</u>	<u>PERCENTAGE OF WHITE DEFENDANT CASES</u>
0-2	62.2	60.6
3	24.0	19.3
4-5	13.8	20.1

TABLE 8: DISTRIBUTION OF MURDER CASE CLASSIFICATION BY RACE OF *VICTIM*<sup>47</sup>

<u>s-VALUE</u>	<u>PERCENTAGE OF BLACK- VICTIM CASES</u>	<u>PERCENTAGE OF WHITE- VICTIM CASES</u>
0-2	75.2	52.2
3	18.3	24.7
4-5	6.5	23.1

---

certain differences in case characteristics.

<sup>46</sup> Baldus, Pulaski & Woodworth, *supra* note 3, at 710.

Note that the distribution of cases by s-value is almost independent of the *defendant's* race. While there is a small excess of black defendants in Category 3, there is a small discrepancy of the opposite kind in Categories 4 and 5. Thus, we would not have expected much overall correlation between the chance of a death sentence and the race of the accused and, indeed, there is not much.

For race-of-victim, however, the situation is different. As proportions of their respective total numbers, there are 1.4 times more white than black victims in Category 3 killings. And in Categories 4 and 5, the corresponding multiplier jumps to 3.6. Therefore, whites are disproportionately victimized by those kinds of killings that most often evoke death verdicts. Perhaps these statistics form the embryo of a *nonracial* explanation of the apparent importance of the victim's race.

Before pursuing this line of thought further, however, there is a possible parallel to consider. In some job discrimination cases, questions have been raised about the value of performance ratings devised by employers. The objection was made that such ratings, far from being neutral measures of employee achievement, might be reflections of the very bias that was the subject of inquiry. In the present context, the various distinctions made by judges, prosecutors, and juries, even if expressed in terms unrelated to race, could still be manifestations of conscious or unconscious racism.

The author is no more qualified to make judgments on this matter than are readers. I would simply suggest a careful review of Appendix A, with special attention to whether the various criteria induce race-related effects, or vice versa.

Delving into the data a bit further yields the following chart:

---

<sup>47</sup> While the records include some multiple killings, there were none with victims of both races.

CHART A  
DEATH SENTENCING RATE BY RACE OF VICTIM:<sup>48</sup>

	BLACK VICTIM	WHITE VICTIM
CATEGORIES 0-2:	.01(184)	.01(189)
CATEGORY 3:	.11(45)	.32(89)
CATEGORY 4:		
(1,1,2)	.75(8)	.84(64)
(2,0,2)	.57(7)	.55(11)
CATEGORY 5:	1.00(1)	.86(7)

The victim's race seems rather unimportant outside Category 3, but in that category the situation seems less clear. There the killers of whites are about three times as likely as the killers of blacks to get capital sentences. A more detailed breakdown of the Category 3 murders yields the following statistics:

TABLE 9: DEATH SENTENCING RATES BY RACE AND DEFENDANT'S RECORD: CATEGORY 3 HOMICIDES

	BLACK KILLS BLACK	WHITE KILLS BLACK	WHITE KILLS WHITE	BLACK KILLS WHITE
PRIOR RECORD	.11 (18)	0 (1)	.31 (13)	.50 (24)
NO PRIOR RECORD	.13 (24)	0 (2)	.29 (31)	.14 (21)

These numbers demonstrate that the higher death rates for the killers of whites are not explained by their worse prior records. Once normalized for the defendant's history, *all* rates in the table for black-victim slayings fall below their counterparts for white-victim cases.

There is some value to paying particular heed to the half of Category 3 murders that are classified (1,1,1). As murders go, such crimes are not especially bestial; nor do their perpetrators seem unusually cold-blooded. A typical (1,1,1) killing is of the homeowner during a bur-

---

<sup>48</sup> Case D-01 (Shavers) is deleted from this Table because the two computer variables in the file provided contradictory information about the race of the victim.



glary, of the taxi-driver during a robbery, or of the policeman trying to arrest the defendant. Self-defense, narrowly construed, is rarely an element in such cases, and rarely are the murders vile. (Vile robbery-killings of course exist, but they would generally be classified (1,1,2).)

The race-of-victim effect is more pronounced in the (1,1,1) cases than in Category 3 as a whole. And in (1,1,1) killings committed by prior felons, a factor of 5.5 separates the death sentencing rates in the white and black victim cases. To be more specific, there were twenty white-victim (1,1,1) cases in which the defendant had a serious prior record; ten of them led to death sentences. But of the eleven such black-victim cases, only *one* led to the death penalty. Despite the small sample sizes, this disparity is statistically significant if viewed in isolation.

While this last finding heightens the sense that race is important, its message is not utterly definitive. A racial disparity is only visible at one of six *s*-values,<sup>49</sup> and it owes its significance at *s*=3 to an especially large effect among (1,1,1)-PR cases.<sup>50</sup> Yet (1,1,1)-PR killings comprise only about five percent of the 606 murders studied. Even if race were irrelevant throughout the data set, apparently strong effects could arise in small subsets by chance alone.

But, while not without some force, such an argument might not be compelling here. The (1,1,1) murders committed by prior felons are not just a random subset of the universe of homicides. They bring forth anger and a great deal of fear and, because they are more "rational" than other homicides, they might more plausibly be deterred by capital punishment. Yet they are not so gruesome as to be hideous murders *per se*. Against such conflicting pressures, the race of the victim might attain greater importance than in more "clear-cut" situations.

There could be at least two different sources of such a racial pattern. Zeisel has suggested that, because blacks are far more hostile to the death penalty than whites, their insistence on its use in black victim cases might nowhere approach the comparable white attitude.<sup>51</sup> Especially in a "borderline" category like (1,1,1), such a difference might influence prosecutors pondering whether to seek the death penalty. And, as Gross and Mauro have noted, predominantly white prosecutors

---

<sup>49</sup> In the 12 (i, j, k) cells outside of Category 3, the white victim rate was higher in four, the black victim rate in three, and the rates equal in the remaining five. And in most of the cells in which a difference arose, there was either one death case or one life case, meaning that some disparity was inevitable.

<sup>50</sup> Here and subsequently, the notation (i, j, k)-PR refers to the (i, j, k) killings committed by defendants with prior records.

<sup>51</sup> Zeisel, *supra* note 10, at 467.

and juries might feel special sympathy for white victims resembling themselves.<sup>52</sup> The limited data we have studied cannot in themselves ratify any such theories.<sup>53</sup>

A summary of our findings might go as follows: Salient differences in the details of the killings of blacks and whites could, to a considerable extent, explain the higher rate of death sentences in the white-victim cases. But in a limited fraction of cases — exemplified by the robbery-killing of a merchant — the race of the victim might matter a great deal. Thus, while it could idealize the murder trials to call them color-blind, it might caricature them to speak of ubiquitous racism.

### VII. THE CATEGORY 3 DILEMMA

Despite the efforts in the last few sections, I have been unable to uncover even one subdivision of Category 3 in which a majority of trials ended in death sentences. While I will not discuss formally the legal ramifications of this outcome, we might do well to speculate briefly on how it arose.

Even when acting under statutory guidance, individuals will differ in their assessments of when death sentences are warranted. Research summarized by Kadane, for example, has suggested that only one percent of the adults in this country would always impose a death verdict in cases where the option exists.<sup>54</sup> Most people apparently have thresholds that separate the “deathworthy” situations from the others.

Judging from the trial outcomes, there is overwhelming sentiment against the death penalty for killings with *s*-values lower than three. And there is a clear consensus that death is appropriate in the classes (1,1,2) and (2,1,2). (The consensus is a bit shaky in the class (2,0,2), but the “halfway” death rate of 50 percent is passed.) In Category 3, which contains 20 percent of the killings and 30 percent of the death sentences, matters are far less settled. It could be that many of the personal thresholds that divide “life” from “death” cases fall within that category’s boundaries.

Yet, the outcomes of the Category 3 trials, viewed collectively, might

---

<sup>52</sup> Gross & Mauro, *supra* note 5, at 107.

<sup>53</sup> Zeisel, *supra* note 10, and Zimring, Eigen & O’Malley, *supra* note 21, also mention “social distance” theories under which the slayings of “high-status” people by those with lower status would bring forth the harshest punishment. But in class (1,1,1), the social status of the victims (taxi drivers, merchants, etc.) is neither exceptionally high nor correlated with race.

<sup>54</sup> Kadane, *Juries Hearing Death Penalty Cases: Statistical Analysis of a Legal Procedure*, 78 J. AM. STATISTICAL ASS’N 544 (1983).

define a public position on such killings far more clearly than do general statutes or vaguely worded opinion polls. And that position seems decisively to *reject* the use of the death penalty in the cases in Category 3. Those perpetrators of such murders sentenced to death can, with some justification, view themselves as unlucky: they received harsher sentences than the heavy majority of their "peers."

To be sure, the people might prefer a more complex sentencing strategy than an "all or nothing" approach. While they might recoil from putting to death every robbery-killer, they might nonetheless support the occasional execution within that group; such sporadic acts, it might be reasoned, might keep alive a flicker of deterrence. But such deliberate caprice in the sentencing policy would seem to offend present legal doctrine and is precisely what proportionality review is meant to prevent.

Suppose, for argument's sake, that all Category 3 death sentences in Georgia were vacated. If our classification scheme makes sense, this act *alone* might greatly reduce the arbitrary element in the Georgia death sentencing.<sup>55</sup> (The four aberrant death sentences in Categories 1 and 2 would also presumably be vacated.) Yet, Georgia would still have a death penalty rate of 12 percent for jury-trial murders; this would entail something like eight executions per year. Georgia might be able to satisfy the requirements for proportionality review, therefore, without coming anywhere close to abolishing capital punishment.

Despite Category 3, it does appear that Georgia's prosecutors, judges, and juries behave with a fairly high degree of consistency. The variance of outcomes in the trials of "similar" murders is probably far less than the dispersion of attitudes among the citizenry of Georgia. But I should say more about the norms around which this consistency takes place, and do so in the next part.

### VIII. SOME IMPLICATIONS OF THE SCORING RULES

The classification rules in Appendix A contain a large number of qualifying remarks, such as:

- (i) "If the only evidence for self-defense is the defendant's uncorroborated claim, assume its absence . . . ."
- (ii) "Neglect references to insanity if the defendant has no

---

<sup>55</sup> Again, readers should recall that I have not considered possible anomalies in the process under which murder charges are filed. Note, however, that if Category 3 death sentences were vacated, the racial disparities suggested in the last section would also diminish greatly.

apparent medical history.”

- (iii) “The killing has an ‘accidental’ touch about it, because . . . (three specific conditions are set out).”
- (iv) “Give a rating of zero on ‘certainty’ if it seems clear that the defendant neither ordered the killing nor was the triggerman. (Note that this *differs* from the weaker statement that it is uncertain whether the defendant was the triggerman.)”

Although rather specific, such statements hint at broader tendencies. In a substantial fraction of homicide cases, the accused contends that he did not kill the victim, or that, if he did, the killing was an accident, an act of self-defense, or the consequence of temporary or longer-term mental illness. What (i)-(iii) imply is that juries tend to greet such claims with skepticism and to give them little weight unless strong evidence supports them.

The caveat in (iv) deserves elaboration. Suppose three accomplices commit a robbery-homicide, but it is unclear which one killed the victim. Juries appear to disregard this indeterminacy in setting punishment for any one of the defendants. Perhaps there is an inchoate fear that to do otherwise would allow the co-perpetrators, through a strategy of collusion, to reduce the total punishment that they receive. It is noteworthy in this connection that, in 1982, Texas executed a participant in a robbery-killing even though it was never established whether he or his partner fired the fatal shot.<sup>56</sup>

In some respects, the rating procedure is as revealing for what it leaves out as for what it includes. It makes no reference to the age of the defendant, to his being retarded or killing while intoxicated, or to his showing remorse after the act. Given that Georgia has no statutory mitigating factors, the disregard of such details might not be inherently improper. Still, their apparent insignificance in the punishment decision — coupled with the exacting definitions of self-defense, accident, and mental illness — suggest a pattern: given flexibility about how to interpret the word “mitigating,” prosecutors and juries incline towards narrow rather than expansive solutions.

Of course, the theoretical basis for death sentence decisions is the Georgia homicide statute. The empirical correlation is of moderate strength: the circumstances that increase the risk of a death sentence are consistent but not coincident with the statutory aggravating factors. Un-

---

<sup>56</sup> See N.Y. Times, Dec. 6, 1982, at A16, col. 4; N.Y. Times, Dec. 7, 1982, at A1, col. 1, A19, col. 2.

like the law itself the classification rules distinguish whether the felony accompanying a homicide is a robbery or a rape. And while the statute imparts special significance to the killing of a police officer, the rules treat as equally aggravating the slaying of any person in his professional capacity. The statutory reference to "great risk to a large number of persons" seems to have little practical importance; the phrase "wantonly vile," by contrast, seems exceedingly significant. Indeed, the rules could be construed as providing a detailed definition of the concept.

Certain provisions of the classification rules do not have their roots in the statutory aggravating factors. For killings in a private home, for example, the statute makes no reference to the number of victims. But so long as some statutory aggravating factor is present (for example, the element of robbery, or the defendant's prior record), the jury could cite it to justify a death verdict that it really believes is warranted for a broader set of reasons. Thus, differences between the rules and the statute could reflect not contradictions, but rather the exercise of a certain flexibility that is built into the law.

I should stress that the classification scheme is intended to model the general tendencies in the data. Thus, I am not suggesting, for example, that there was never a case in which a jury felt mercy towards an intoxicated defendant. But this caveat does not eliminate the need to verify the assumptions embedded in the scoring rules, a task I attempt in Appendix C.

#### IX. SOME THREATS TO THE ANALYSIS

Several potential problems endanger both our specific findings and any generalization that might be drawn from them. The hazards deserve explicit recognition and discussion.

Questions arise about the validity of my perceptions concerning Georgia death sentencing. For one thing, I only considered the minority of homicide cases in which the defendant was tried for murder. If, in the setting of charges, prosecutorial discretion is exercised in an arbitrary manner, Georgia's overall punishment structure might be far less comprehensible than the one depicted.<sup>57</sup> Should a large number of cases I would have classified (1,1,2)-PR for example, have elicited only manslaughter charges, the 94 percent death rate in that class could be

---

<sup>57</sup> This is a special case of the problem of "sample selection bias," which is discussed in Berk, *An Introduction to Sample Selection Bias in Sociological Data*, 48 AM. SOC. REV. 386 (1983).

highly misleading.

Even among the cases I did consider, the issue of accuracy is not absent. Professor Baldus and his associates did an exemplary job of gathering data, but narrative summaries and computer variables cannot possibly illuminate everything of interest. Brief case descriptions can only crudely portray the credibility of the various witnesses, the eloquence of the opposing attorneys, and the emphasis placed in court on the various elements of the case. Of course, perfection in this regard is unattainable, but one cannot deny that its absence could have adverse consequences.

Indeed, the case summaries could suffer from more than just random lapses in reporting. Ultimately, the information that they contain was provided by the Georgia justice authorities. It is possible, therefore, that I have something between a neutral description of the case and an attempt to justify its outcome.<sup>58</sup> A cynic might argue that I have said less about the realities of the Georgia justice system than about the success of the "revisionism" that was engaged in on its behalf.

These are serious matters, and no one can blithely discount the threats that they represent. But to acknowledge that point is not to concede that the problems are genuinely damaging. It is certainly conceivable that any worst-case scenario is an enormous magnification of the actual difficulty.

Consider, for example, the hypothetical danger to the statistic about the (1,1,2)-PR class. The data set contains hundreds of "lover's quarrel" and "barroom brawl" type slayings involving both black and white victims and defendants with and without prior records. For a (1,1,2)-PR case to have escaped my attention, therefore, it would have to have been downgraded past large numbers of transparently less aggravated killings. That such anomalies occur with regularity is not self-evident.

As for distortions within the summaries themselves, they could be of two kinds: (i) inaccurate statements (whether deliberate or accidental); or (ii) the absence of important information (whether intentional or otherwise). Of the possibilities just cited, all but "unconscious omission" raise direct questions about competence or integrity. Such involuntary bias might therefore appear the most plausible source of misrepresentation. It is worth noting that, in the literature, unconscious processes are given a large share of blame for observed inequities in capital sentencing.<sup>59</sup>

---

<sup>58</sup> Professor Samuel Gross made this point forcefully in a most insightful letter to the author. Letter from Samuel Gross to Arnold Barnett (Sept. 14, 1984).

<sup>59</sup> For example, in his letter to me, Professor Gross said that everyone he knew who

But how substantial in the present context is the risk of unconscious distortion? Suppose that, in support of a claim of self-defense, ballistics evidence is introduced that shows that the victim had fired his gun six times. Or suppose that, in a case involving several defendants, the accused made three contradictory statements to police about his personal involvement in the killing. Is it realistic to think that experienced Georgia personnel would suppress all *memory* of such details while writing summaries about the case?<sup>60</sup>

As these last paragraphs imply, I am skeptical that the Georgia records analyzed had been subject to severe distortion. But there is no simple way I can dispel the concerns of those predisposed to think otherwise.

Quite apart from such issues, it is unclear how far one can generalize either my particular results or the method by which they were obtained. It should be remembered that the scheme is exploratory in nature, emerging from Georgia data rather than some *a priori* theory. For this reason, the classification procedure itself — and not just the conclusions arising from its use — might not survive a transplant to another state. (That possibility could, of course, be the subject of further investigation.)

Indeed, there are reasons to suspect that Georgia is not fully comparable to (say) Pennsylvania, Kansas, California, or even Louisiana. Georgia's relatively high death-sentencing rate might alone suggest less flagrant inconsistency than in other states where capital verdicts are far rarer. And the asymmetric manner in which its statutes treat aggravating and mitigating circumstances (seeming to deemphasize the latter) could induce different sentencing behavior than other, more "balanced" laws. To be prudent, I should bound the region in which the findings apply with the Georgia state line.

And there is another limitation of this work that transcends such practical problems. I have tried to depict how Georgia's judges, juries, and prosecutors actually behave, which is quite separate from the issue of how they *should* be behaving. In the language of social science, this

---

had studied the issue "believe[s] that much — or most — of the racial discrimination in the imposition of capital punishment can be explained by processes that are on the whole unconscious." *Id.*

<sup>60</sup> To say this is not to deny that unconscious biases can exist. It is possible to imagine that a white juror, confronted with a set of aggravating and mitigating factors, might unintentionally give more weight to the latter when the victim is black rather than white. It seems less likely, however, that the juror would simply forget the aggravating circumstances, and it is a distortion of this kind that would imperil case summaries.

is a descriptive study and not a normative one. As such, it is not meant as some simple blueprint for proportionality review or anything else.

This Article *does* suggest that Georgia sentencing behavior is somewhat consistent. But even if consistent, the distinctions made by juries could well (in Kalven's and Zeisel's phrase) be demeaningly trivial compared with the differences in punishment that they entail. In any case, consistency need not assume some preeminent status in the distribution of punishment. Suppose (to put the matter graphically) that juror choice in homicide cases were restricted to death or probation.<sup>61</sup> There would probably be clear regularities in the sentencing outcomes, but they could never make defensible the monstrous choice imposed on the jurors.

This Article did not aspire to — *nor could it ever* — provide a justification for death verdicts. While the findings furnish empirical evidence concerning certain attributes of just sentencing, about numerous others they have nothing useful to say.

## X. COMPARISON WITH THE WORK OF BALDUS, PULASKI, AND WOODWORTH

As noted earlier, Baldus, Pulaski, and Woodworth (BPW) performed their own analysis of the Georgia sentencing data. Here we review their general approach, discuss some strengths and weaknesses of both their methodology and my own, and assess the consistency of the principal findings of the two studies. To perform these tasks in a somewhat orderly manner, I divided this part into four sections.

### A. *The BPW Approach*

There are two major thrusts to the BPW analysis. In a "first pass" at analyzing the Georgia cases, BPW divided the killings according to their numbers of aggravating circumstances. They computed death-sentencing rates as a function of that number and then, among the cases at a given level of aggravation, they broke down the rates by some auxiliary variable like race-of-victim.

But closer to the heart of the BPW efforts were a series of multiple regression analyses.<sup>62</sup> These regressions assigned a numerical score to

---

<sup>61</sup> This example was suggested by Professor Frank Zimring, to whom I am grateful for many sharp observations about the uses and limitations of this work.

<sup>62</sup> A good introduction to regression analysis appears in T. WONNACOTT & R. WONNACOTT, *INTRODUCTORY STATISTICS FOR BUSINESS AND ECONOMICS* chs. 12-13 (1972).



each case meant to relate systematically to the chance the defendant would receive a death sentence. One major scoring rule that emerged was based on about two dozen variables: starting at zero, one added or subtracted a series of numbers depending on which factors were relevant to the case. An insurance motive, for example, increased the score by .17, the kidnapping of the victim by .10, and a murder-by-drowning by .12. In contrast, the score dropped by .08 if the defendant was not the triggerman, and by .03 if the victim had "low status." Nineteen other circumstances would also affect the score, which could range from about +2 to -.3.

### B. Some Comments on the BPW Approach

The "number of factors" rule reflects the intuitively appealing notion that, as the number of aggravating circumstances goes up, so does the appropriateness of the most extreme penalty. But the statistic has some serious drawbacks. It makes no provision at all for any mitigating factors present, and tacitly assumes that all aggravating factors should be attributed equal weight. Moreover, there is nothing in the Georgia statute that suggests basing decisions on such a variable. Thus, whether as a guideline to sentencing behavior or as a descriptor of it, the mere number of aggravating circumstances seems of limited relevance.

On the other hand, the scoring rules embody subtle distinctions, careful measurements, and considerable statistical power. Any qualms about it are traditional ones concerning the limits of regression analysis. Part III discussed some potential troubles (correlated variables and the assumption of independent effects); other difficulties could involve definitions of variables, certain linearity assumptions, and the tendency of ordinary-least squares methods to weigh some data points more than others.<sup>63</sup> The authors also note the danger of "overfitting" (that is, the inclusion through chance of noncausal variables in the regression equation).<sup>64</sup>

The study itself provides evidence that such problems might not be mere technicalities. The scoring formula cited arose from averaging three regression analyses that were calibrated from overlapping data.

---

<sup>63</sup> See D. BELSLEY, E. KUH & R. WELSCH, *REGRESSION DIAGNOSTICS: IDENTIFYING INFLUENTIAL DATA AND SOURCES OF COLLINEARITY* (1980). The authors also use logistic regression analysis to duplicate certain results; this technique produces scoring rules that, roughly speaking, are multiplicative rather than additive. Though perhaps more natural than linear regression for the estimation of probabilities, it is subject to variants of all the problems cited.

<sup>64</sup> Baldus, Pulaski & Woodworth, *supra* note 3, at 695 n.106.

Three regressions were necessary, the authors explained,

because the tendency of each analysis to produce a unique solution which omitted obviously important and relevant variables. For example, the following variables were omitted from one or more of the models: number of prior felony convictions, victim a hostage, number of convictions for violence, defendant not the triggerman, defendant created great risk in a public place, and insurance motive.<sup>65</sup>

Moreover, some variables appeared in initial regressions with coefficients whose signs seemed absurd.<sup>66</sup>

Quite beyond such matters, there is the issue of how to interpret the regression results. The scoring rule identifies some major influences on sentencing and suggests their relative importance, but it leaves far less transparent what death rates result from the interplay of various circumstances. For example, the authors define Index Category IV — murders involving death risks near the overall Georgia average — as all cases with final scores between .23 and .36. There are literally *thousands* of combinations of elements that would assign a case a score in this range. Thus, a simple characterization of the killings with “average” death risk would seem very hard to devise.

Indeed, two cases with the same score (for example, .25) need have no overlap *at all* in the contributions to their common rating. One case could involve a web of aggravating and mitigating factors, while the other is comparatively nondescript. Thus, statistical similarity need bear no clear relation to *conceptual* similarity. The scoring rule might do well at predicting jury behavior without making clear the interaction of forces that generates that behavior.

These statements do not imply that the scoring method is terribly flawed, nor do they indicate that this Article’s approach — which has problems of its own — is clearly superior. But they do suggest that another try at analyzing the Georgia data was not utterly redundant.

### C. *More on the Present Methodology*

This Article tries to bring the flavor of exploratory data analysis to a large-scale data base. With no prior structure having been imposed on the analysis, the case summaries were read individually. Ultimately, lots of single observations were merged into a succinct theory (though

---

<sup>65</sup> *Id.* at 689 n.98.

<sup>66</sup> *Id.* Because regression results are subject to an analogue of sampling error, these problems are not inherently devastating. But they might be warning signs that something is amiss and, if that is the case, averaging several regressions together is unlikely to eliminate the trouble.

the detailed implications of that theory, as Appendix A suggests, might not be especially succinct).

Beyond the hazards that were cited earlier, there are others that are intrinsic to this approach. It is farfetched if not preposterous to suggest that anyone comes to the Georgia cases with no prior conceptions whatsoever. And there is the danger of impressionism, of giving excessive weight to a few unusual cases, or of such pedestrian problems as lapses in memory. There is even the unsettling possibility that the order in which the cases were read influenced the conclusions reached about them.

Accompanying such dangers, however, is the opportunity for a more textured, more nuanced view of sentencing behavior. There are enough cases that there is a good prospect of grasping underlying patterns and, to some extent, of testing hypotheses about them (see Appendix C). And one retains an alertness to unanticipated "signals" from the data, an alertness that can suffer when the records are processed *en masse* with some predetermined method. For these reasons, certain weaknesses of both the classical studies and small-scale exploratory work might be diminished in the present effort.

To an extent, the outcome of this exercise provides an advertisement for it. The partition of cases that arose are conceptually simple: a particular class like (1,1,2) or (1,0,0) is fairly easy to describe, and the description is sufficiently "tight" that the class's member cases differ relatively little. Thus, the calculated death risks by class (sometimes amplified by prior record, region, and race-of-victim) suggest with some clarity how juries respond to each juxtaposition of primary factors.

There is the further point that conceptual simplicity was not achieved at the expense of explanatory power. A small set of dimensions identifies some killings as having a high likelihood of a death sentence, and others for which that outcome is almost inconceivable. And, identified by these same dimensions as falling between these two "poles," one finds the important minority of Georgia cases in which the facts lead to death sentences only one time in four. Hence, the "margin" for death decisions — as determined by opposite sentencing patterns on opposite sides of it — turns out to be the place where sentencing outcomes are, in fact, least consistent.

#### D. *The Results Compared*

To what extent do the two papers differ in their primary findings? The quick answer would be "not very much." While the detailed results are not identical, there is broad concurrence on the factors that,

viewed individually, the juries find aggravating and mitigating. One could easily support the generalization that, if a killing falls into high death-risk category in one of the papers, it will most probably do so in the other.

If there are any major conclusions that seem discrepant, they concern the subject of race. BPW suggest that the race of the victim is a strong and systematic influence; as previously noted, they go so far as to speak of a two-track system for processing cases. The present Article, while not suggesting that race is immaterial, nonetheless accords it a smaller explanatory role. Many early readers of this Article were perplexed that, on an issue of such critical importance, the two papers seemed to reach such conspicuously divergent outcomes.

A close look, however, makes clear that the differences are more apparent than genuine. In the BPW paper, the race-of-victim effect is strongest when the cases are disaggregated by the number of aggravated factors; that partitioning rule, as I argued above, is especially vulnerable in its underlying premises. In the regression results on post-*Furman* data, by contrast, the role of race drops substantially; the average racial disparity is estimated at three percentage points and the effect is not significant at the 5 percent or even the 10 percent level.<sup>67</sup> That is actually a *weaker* finding than the one arising in this Article, where the average discrepancy is approximated as five points.<sup>68</sup>

Although they suggest that race has a major effect on sentencing, BPW make no specific assessment of the kinds of cases in which its influence is the greatest. This Article, by contrast, calls attention to such particular classes as (1,1,1)-PR. Perhaps the reason for the difference is that, because the BPW partitioning mechanisms divide the cases into conceptually diffuse categories, they are not especially strong in pinpointing the origin of any given phenomenon. If so, that circumstance supports a major rationale for this Article: the view that, by forging close contact with the individual case records, the exploratory approach might enhance the precision of their analysis.

Whatever the case, any balanced comparative analysis would emphasize similarities rather than differences. The fact is that two distinct

---

<sup>67</sup> See Baldus, Pulaski & Woodworth, *supra* note 3, at 708 Table 6.

<sup>68</sup> Even when prior records were considered, I found a 22-point disparity in Category 3, which contains 22% of the cases. No notable differences showed up at the other *s-values*, see *supra* note 49. The overall average effect, therefore, was about 22% x 22, approximately 5 percentage points.

The BPW estimate of a three point difference appears at the bottom of the table found at Baldus, Pulaski & Woodworth, *supra* note 3, at 710 (table concerns the only post-*Furman* regression reported on in any detail).

approaches with very different limitations achieved a noteworthy degree of consensus about the Georgia sentencing data. And because they did so, each one serves to increase the credibility of the other.

#### CONCLUSION

The analysis described in this Article suggests that, broadly speaking, Georgia's murders fall into three groups. In the first and largest, death sentences are so rare as to resemble the "strikes of lightning" that led to *Furman v. Georgia*. In another, comprising about one-seventh of the cases, it is those *spared* a sentence of death whose treatment appears unusual. But about one-fifth of the cases fall into an intermediate range, in which death verdicts are neither especially rare nor especially common.

It appears that the general indicators of a defendant's death risk are the certainty that he killed and did so deliberately, the victim's status with respect to the accused, and the extent to which the slaying was heinous in its details. In murders whose facts make a death penalty a serious possibility, the defendant's prior record seems important, as does sometimes the race-of-the-victim. Such variables appear to explain a good deal about sentencing, but there remain some situations in which life and death sentences coexist side by side. From the standpoint of proportionality review, these situations are particularly troubling.

An empirical exercise like this one cannot yield incontrovertible truths. Rather, it presents a theory on how Georgia selects the homicide convicts it condemns to death. The soundness of any such theory depends on its innate plausibility, on the appropriateness of the data set that yielded it, and on its explanatory power within that data set.

Readers will have to assess for themselves how much credence to accord the theory presented in this Article. And if persuaded of its viability, they must face a deeper question: is the Article more comforting or is it more disturbing in what it suggests about post-*Furman* death sentencing?

## APPENDIX A: CLASSIFICATION RULES

## I. THE CERTAINTY THE DEFENDANT IS A DELIBERATE KILLER

Score the case either 0, 1, 2 on this dimension, applying the following criteria:

(i) The case is rated 0 if *any* of the following circumstances pertain:

- (1) The narrative indicates the evidence in the case seemed weak (e.g., “case based solely on circumstantial evidence”).
- (2) The narrative mentions evidence that worked *against* the view that the defendant was guilty (e.g., tests for residue on the defendant’s hand from firing a gun were negative).
- (3) It seems clear that the defendant neither ordered the killing nor was the triggerman. (Note that (3) *differs* from the weaker statement that it is uncertain whether the defendant was the triggerman.)
- (4) The killing has an “accidental” touch about it, because
  - (a) a fairly long period (perhaps a week or more) elapsed between the incident and the victim’s death, or
  - (b) the death was caused by a shot fired somewhat randomly (e.g., through a door), or
  - (c) the death was caused by a beating similar to previous beatings of the victim by the defendant.
- (5) There is reason to doubt that the defendant’s actions *in themselves* would have caused the victim’s death (e.g., (i) the defendant beat the victim, but it was a co-perpetrator’s stabbing that killed him, or (ii) the defendant’s beating of the victim induced a heart seizure).
- (6) The defendant was one of several participants in a conspiracy to kill, but took no part in the actual killing.
- (7) The narrative mentions that the defendant was previously treated for mental problems (e.g., institutionalized). Neglect references to insanity if the defendant has no apparent medical history.

(ii) The case is rated 2 if any of the following elements were present:

- (1) The killing was a murder-for-hire, and the defendant was either the sole instigator or the executioner.
- (2) The defendant plotted to kill the victim (e.g., a wife and her lover arrange to murder her husband). If, however, the defendant was one of several plotters, and clearly not the actual killer, assume (2) is not satisfied.
- (3) The narrative mentions that the defendant was officially implicated in other killings.
- (4) The narrative mentions that the defendant had tried previously to kill the victim.
- (5) The defendant announced *in advance* to a third party an intention to kill the victim. (Neglect this condition in a lover’s triangle or lover’s quarrel case, or when the third party was a co-perpetrator.)

- (iii) If the killing warrants neither a 0 nor a 2, give the case a rating of 1. If the killing satisfies conditions for *both* 0 and 2, also rate it 1. Most "common" slayings, such as killings during armed robberies or during barroom fights, would warrant this intermediate classification. Indeed, a 2 reflects unusually clear evidence of premeditation, while a 0 reflects unusually large doubt that the defendant knowingly acted to cause the victim's demise.

## II. THE STATUS OF THE VICTIM

On this dimension, the score is either 0 or 1.

Give a score of 0 if:

- (1) The victim was a relative of the defendant (even his or her child).
- (2) The victim was a friend of the defendant. (Interpret the word "friend" loosely; if, for example, two people of similar age are riding together voluntarily in a car, consider them friends. However, the mere fact that two people know each other is not sufficient. Neighbors of vastly different ages, or the bank teller and the depositor, are not assumed friends barring other evidence of social ties.)
- (3) The victim was an enemy of the defendant, though not the defendant's employer. (Interpret the word "enemy" loosely; if, for instance, the victim and defendant vied for the affections of the same woman, if the victim had harassed one of the defendant's loved ones, if there was a feud of some sort that turned violent, assume enmity existed. If, however, the victim could be viewed as the defendant's employer — whether as (say) his supervisor in a factory or the person who hired him to perform some chores — do not give a score of 0 under (3).)
- (4) The victim, although a stranger to the defendant, acted in a highly provocative manner just prior to the killing (e.g., racial taunts).
- (5) The victim was engaged in an illegal or often-disapproved activity at the time of the killing (e.g., a drug dealer, a prostitute or prostitute's customer, owner of a homosexual bathhouse, etc.).

If the case does not warrant the rating 0, give it the score 1. 1 is the appropriate rating for most stranger-to-stranger killings and those in which the defendant only knew the victim in the latter's official capacity (e.g., as employer, or attendant in a local gas station). If there are several victims, give the case a 0 if *any* of those slain qualify for it.

## III. THE HEINOUSNESS OF THE MURDER

There are two aspects to this dimension: the question whether self-defense motivated the killing and how "gruesome" it was.

Self-defense is an element in the case under any of the following circumstances:

- (1) The victim had *at hand* a deadly weapon at the time of the killing. (Merely having a gun in the store or house does not satisfy (1).)
  - (2) The victim was killed with his own weapon. (This is taken to imply (1) is satisfied even if the narrative does not explicitly say so.)
- NOTE: If the victim was a police officer, do not invoke self-defense did (1) or (2) unless the officer fired shots before the defendant did.
- (3) The victim had threatened to kill the defendant or one of the defendant's loved ones.
  - (4) The victim had attacked the defendant at the time of the killing.

If none of the above conditions existed, self-defense was not a mitigating circumstance in the homicide.

NOTE: If the only evidence for self-defense is the defendant's uncorroborated claim, assume its absence even if any of (1)-(4) is alleged.

A homicide is classified as vile if one of the following circumstances is present:

- (1) It was accompanied by rape, or sexual abuse, either against the victim or someone in the company of the victim.
- (2) There were at least two homicide victims.
- (3) The deceased was a kidnapping victim at the time he was slain.
- (4) Psychological torture preceded the killing (e.g., Russian roulette, a sustained period of terror).
- (5) The victim was shot several times in the head at close range.
- (6) The killing was execution-style (i.e., victim forced to kneel or squat, then shot in head).
- (7) The death was caused by strangulation, or arson.
- (8) The death was caused by a drowning in which physical force kept the victim below water.
- (9) The killing involved ten (10) or more shots or stab wounds, except when the murder weapon was a penknife or other small cutting instrument.
- (10) The physical details of the killing are unusually repulsive (e.g., the victim drowned in his own blood).
- (11) The body was mutilated, or otherwise grossly disfigured (except in an attempt to conceal the homicide).
- (12) The killing was performed with a bizarre weapon (e.g., a hacksaw, a claw hammer, an icepick).
- (13) The defendant apparently derived pleasure from the very act of killing. (This is distinct from his believing the victim deserved to die, and taking pleasure on that account.)
- (14) The crime was specifically described in the narrative as extremely bloody.

Absent *all* these circumstances, the homicide is categorized as not vile. Despite the length of the list above, most "simple" shootings, stabbings, and beatings would not be classified as vile under these rules.



## APPENDIX B: FOUR ILLUSTRATIVE CLASSIFICATIONS

*Example 1*

The defendant, a 27-year-old female, fatally shot her husband. The defendant was angry that the victim was messing around with another woman. The defendant asked a friend for a ride for the purpose of "locating and cutting up" the other woman. The friend refused and took her home. After her return home, the victim was leaving the house when a shot was heard. Witnesses saw the victim fall to the ground and the defendant emerged from the house carrying a gun. Witnesses testify that she said, "I told him if I ever caught him messing around with another woman, I'd kill him." Defendant told police, "I don't know why I shot him, I just did." Victim had a history of beating his wife.

TABLE 10

Ratings:

"Deliberateness"		<u>1</u>	
"Status of Victim"		<u>0</u>	
"Heinousness":	Self-defense	<u>NO</u>	} Thus heinousness rating was <u>1</u> .
	Vileness	<u>NO</u>	

*Explanation of Ratings:*

*Deliberateness:* No basis for either a zero or a two, a circumstance that dictates a score of one.

*Status of Victim:* The victim was the defendant's spouse.

*Heinousness:* There is no serious evidence that self-defense was a factor; while the victim "had a history of beating his wife," he was leaving the house at the instant she shot him. None of the criteria for "vileness" are satisfied in this case.

A killing like this one, incidentally, almost never leads to the death penalty (nor did it in this particular case).

*Example 2*

Defendant and co-perpetrator (both teenage males) lived close to victim (a 55-year-old female) and had conspired to rob her for some time. On the night of the murder, the co-perpetrator entered the house first and then forced the victim to let the defendant in. There was evidence that they raped the victim and that, to stop her from screaming, the co-perpetrator repeatedly stabbed her with the defendant's knife. The defendant made three different statements to the police following the crime, showing varying degrees of involvement in the murder.

Defendant was found guilty of rape and murder.

TABLE 11

Ratings:

"Deliberateness"	<u>1</u>		
"Status of Victim"	<u>1</u>		
"Heinousness":	Self-defense	<u>NO</u>	} The heinousness rating was <u>2</u> .
	Vileness	<u>YES</u>	

*Explanation of Ratings:*

*Deliberateness:* No clear evidence of a prior intent to kill (as opposed to rob). While "there was evidence" that the co-perpetrator stabbed the victim, it was not overwhelming: the knife was owned by the defendant and his statements to the police about his guilt were contradictory. As noted earlier, uncertainty whether the defendant was the killer is a weaker condition than clear indications that he was not; it is the latter circumstance that justifies a "0" on deliberateness.

*Status of Victim:* Clearly a "1."

*Heinousness:* No self-defense; the rape in itself makes the killing vile.

This defendant was sentenced to death.

*Example 3*

Defendant was a 23-year-old military man. Victim was a male. Apparently, defendant got drunk with victim's nephew before the offense, and the nephew told defendant that the victim had a large amount of money in his house. It is not clear if this man specifically encouraged defendant. Defendant, while still intoxicated, went into the house and was intent on robbing the victim. Defendant claims that victim came out of his bedroom and fired a shotgun in the direction of the defendant. Defendant returned the fire with a pistol and killed the victim. Defendant then fled with some money he found. When defendant was questioned by police, he made a full confession, gave them the money and led them to where he had hidden the victim's shotgun.

TABLE 12

Ratings:

"Deliberateness"	<u>1</u>	
"Status of Victim"	<u>1</u>	
"Heinousness":	Self-defense	<u>NO</u>
	Vileness	<u>NO</u>

} Thus heinousness rating was 1.

*Explanation of Ratings:*

*Deliberateness:* The killing satisfies none of the conditions that would warrant a "0" or a "2." (Premeditation to rob is not equivalent to premeditation to kill.)

*Status of Victim:* While the defendant knew the victim's nephew, he apparently did not know the victim.

*Heinousness:* The ratings are appropriate because: (i) under the stated criteria, the killing is not especially vile, and (ii) there is no compelling evidence of self-defense, only the defendant's uncorroborated story. (Had a bullet from the victim's gun been found in the wall, the self-defense claim would be more credible and the rating on that subject changed to yes.)

The fact that the defendant left the house with the victim's shotgun does not prove the victim actually tried to use it.

This defendant was sentenced to death. The author suspects that, had the self-defense claim been stronger, the outcome of the trial would have been a prison term.

*Example 4*

Defendant (a 21-year-old male) was the next door neighbor of the victim (a 49-year-old male). The defendant's girl friend went to the victim's house and argued with his wife over a cake plate. Defendant went over and argued with victim and his wife also. Defendant claimed that the victim opened up a knife and threatened him. Defendant went home and later the victim went to his porch with the knife. Defendant shot three times. The third shot was fatal.

TABLE 13

Ratings:

"Deliberateness"	<u>1</u>	
"Status of Victim"	<u>0</u>	
"Heinousness":	Self-defense	<u>YES</u>
	Vileness	<u>NO</u>

} Thus heinousness rating was 0.

*Explanations of Ratings:*

*Deliberateness:* No reason for a zero or a two; thus a one.

*Status of Victim:* The victim and defendant clearly were acquainted; the notion of "dispute turned lethal" fits the spirit of II-(3) in Appendix A.

*Heinousness:* The killing was not vile. But self-defense apparently is an element, given that the victim arrived at the defendant's house with knife in hand.

The defendant in this case was sent to prison.

**APPENDIX C: SOME TESTS OF THE CLASSIFICATION MODEL**

I have presented the full text of the classification rules (Appendix A) and some data about their explanatory power. But there are still some questions including:

- (1) Is the scheme "objective" in the sense that two different people, applying the scoring rules in a given case, would probably reach the same ratings?
- (2) Several potential important factors are not considered in the rating procedure (e.g., age of victim). Is there empirical evidence supportive of such exclusions, and other kindred simplifications?
- (3) We have evidence that the scheme is fairly powerful within the set of cases that led to its formulation. But how well would it *predict* the sentencing decisions in other Georgia cases from the same period?

Below I describe some tests I performed related to each of these questions.

**A. Objectivity**

"Objectivity" was explored in an experiment conducted by Mary Elsner of the National Center for State Courts. She asked student volunteers from William and Mary Law School to read Appendices A and B, and then to consider the narrative summaries of a series of Georgia murder cases. Based on these narratives (from which all information about the verdict was deleted), the students were asked to classify the cases under the scoring rules.

The experiment took place in two parts. The first phase was a "trial run" in which the ratings of the students and the author were contrasted, the aim being to root out cryptic features of the classification scheme. Several minor changes were made, but they all involved elabo-

rating the rules rather than altering them.<sup>69</sup>

In the test phase, nine new students were given twenty-five Georgia cases to classify, none of which had been used in the "clarification phase" of the experiment.<sup>70</sup> For every case, the students provided ratings about deliberateness, victim-status, vileness, and self-defense. The total number of ratings that emerged was therefore  $25 \times 4 \times 9 = 900$ .

Over the 900 scores, the students and the author agreed 87 percent of the time. The 13 percent disagreement rate could have three possible origins: (a) the scheme might contain flaws that provoke occasional discrepancy; (b) some case narratives are genuinely ambiguous on key points; (c) some student ratings are clearly inconsistent with the classification rules. All of these problems probably were present to some extent.

When obvious student errors are excluded (e.g., the case was classified "not vile" although there were two murder victims), the agreement rate rises above 90 percent. And when I set aside rating "disputes" tied to subtle semantic distinctions (e.g., does an "ambush" in a robbery killing imply prior intent to kill, as opposed to prior intent to surprise and rob?), the rate of agreement reaches 93 percent.

Of course, the scheme is not perfectly objective: could there be any universal definition of (say) psychological torture? But the degree of consensus that *did* arise in the experiment strikes me as surprising and gratifying.

### B. Simplifying Assumptions

Other tests that were performed pertained to the internal structure of the classification procedure. Unlike Gross and Mauro, for example, I had no variable related to the sex of the victim per se. How can such an omission, and various others like it that are implicit in the scoring rules, be defended?

The hypotheses in question are of the general form: characteristic *X* is irrelevant to the sentencing decision. To test any one of them, we might divide the cases in a given class into those having characteristic *X* and those that do not. If the death rates in these two subdivisions do not differ significantly, the axiom about *X* is sustained (at least in that class). The word "sustained" does not imply that the hypothesis has

---

<sup>69</sup> The version of the procedure in Appendix A is the final version.

<sup>70</sup> The cases were randomly chosen as follows: 100 cases had been selected by chance by the National Center for State Courts to illustrate the use of a certain computer program. The Center had numbered the cases, and I simply chose every fourth one for the sample.

been decisively verified; rather we have avoided the self-contradictory situation in which, having proceeded on the assumption that *X* is not important, we wind up producing clear evidence that it is.

We performed several such hypothesis tests, three of which are presented below. The results were consistent with the assumption under scrutiny although, as the numbers below will suggest, we sometimes had to proceed with rather small data samples.

*Hypothesis 1:* In death-sentencing decisions, the age and sex of the victim do not matter.

This contention goes against the notion that juries are especially harsh on those who kill women, the very old, or the very young. Killing such "helpless victims," it is sometimes suggested, is more abhorrent than slaying those who might somehow have "fought back."

Certainly the victim's age and sex do not matter when  $s \leq 2$ , for virtually no killings in that range lead to death sentences. After dichotomizing age into the two ranges 15-60 and "0-14 plus 61 and up," the following contingency table for Category 3 and class (1,1,2) appear:<sup>71</sup>

TABLE 14  
DEATH SENTENCING RATES BY AGE AND SEX OF  
VICTIM

	<u>CATEGORY 3 KILLINGS</u>		<u>CLASS (1,1,2)</u>	
	<u>MALE</u>	<u>FEMALE</u>	<u>MALE</u>	<u>FEMALE</u>
Age 15-60:	.21(80)	.29(34)	.71(28)	.85(20)
Other Age:	.30(10)	.30(10)	.90(10)	.80(10)

This table does not depict statistically significant variation around the relevant death risk (25 percent in Category 3; 81 percent in Class (1,1,2)).<sup>72</sup> To be sure, working-aged male victims evoked the lowest death rates in each of the two situations. That outcome hints at the possibility of a slight age-sex effect; however, given the small samples

<sup>71</sup> Killings with three or more victims were excluded from this table, except when all those slain fell in the same category (e.g., elderly women). In such cases, it might be hard for jurors to think about individual victims. In double-victim slayings, the case was classified as female and/or outer-aged if either of the victims met this description. Here both victims can be borne in mind and if, for example, the killing of an elderly person is viewed as especially reprehensible, the resulting hostility is unlikely to be mitigated by the killing of yet another person.

<sup>72</sup> Even when one considers the prior record of the defendant and the setting of the trial (urban/rural), this outcome is unchanged. (By "not statistically significant," I mean that usual sampling error could account for the observed differences.)

available, nothing approaching a decisive pattern can be established.

*Hypothesis 2:* The death risk associated with killing an employee in her official capacity is the same regardless of her occupation.

This hypothesis asserts that, despite statutory distinctions, killing a police officer is not more likely to elicit a death sentence than slaying a cab driver, gas station attendant, or liquor store cashier. Perhaps the "contemporaneous offense" aspect of the latter murders (which usually involve robbery) is given about equal weight to the attack on society itself when a law enforcement officer is slain.

An appropriate place to test Hypothesis 2 is the (1,1,1) class, which includes most employee killings that are not vile. Within that class, there were nine killings of peace officers (police officers, state troopers, detectives, security guards); four of them led to death sentences. There were twenty-seven killings of other kinds of workers; these brought forth six death verdicts for a rate of 23 percent. And the death rate for the other twenty-four (1,1,1) killings is 21 percent. From the overall average of 25 percent one would have expected two or three death sentences in the peace officer cases rather than four. But the excess is well within the range of chance variability and does not in itself suggest a compelling need for occupational distinctions.

*Hypothesis 3:* In killings clearly plotted in advance, the death risk is the same for the instigator and the executioner.

The classes (2,0,1) and (2,0,2) lend themselves to assessing this hypothesis. In (2,0,1), there were two cases in which the defendant instigated the murder but did not take part in it; neither ended in a death verdict. In (2,0,2), three cases involved instigators with no physical connection to the slaying; one of these defendants was sentenced to death.

Given the aggregate death rates in (2,0,1) and (2,0,2), one would have "expected" 1.9 death sentences among the five instigators studied. They collectively received only one but, as above, the deficit lacks significance given the sparsity of data. A more complex model that distinguishes instigators, executioners, and those active in both phases of the killing might be useful, but the data neither necessitate it nor suggest how it could be developed.

In all, the data did not force us to abandon *any* of the simplifying assumptions used in constructing the classification scheme.

### C. Predictive Power

After devising the classification rules from study of well over 400 cases, I considered a "holdout sample" of about 100 more. These new

cases were classified under the scoring rules (which they played no role in developing), and their dispositions were noted.

The sentencing patterns in the new data were similar to those in the original set. Of the nine additional (1,1,2) cases, eight led to death verdicts, a fraction that closely follows the proportion in the earlier cases. Of the sixty-two new cases with s-values between 0 and 2, only one yielded a death sentence. And, at a rate nearly indistinguishable from the earlier figure, 23 percent of the Category 3 killings brought death sentences to their perpetrators.

Being so similar in character to the original ones, the new cases were combined with them for the Table 1 calculations: But the test involving them supported the view that the classification rules reflect general sentencing practices, and not just the idiosyncracies of a limited number of cases.