

# Monitoring and Evaluating Contemporary Death Sentencing Systems: Lessons from Georgia\*

*David C. Baldus\*\**

*George Woodworth\*\*\**

*Charles A. Pulaski, Jr.\*\*\*\**

*In his article "Some Distribution Patterns for the Georgia Death Sentence" Arnold Barnett develops and applies to post-Furman data from Georgia an imaginative and useful method for monitoring and evaluating a death sentencing system. Professor Barnett developed his methodology as a consultant for the National Center for State Courts research*

---

\*Funding for this research was provided by the National Institute of Justice, The National Science Foundation, the Iowa Law School Foundation, the University of Iowa Graduate College, and Arizona State University College of Law.

We are grateful for the assistance provided by the staffs of the Georgia Department of Offender Rehabilitation and the Board and staff at the Georgia Department of Pardons and Paroles. We are particularly grateful for the assistance of Dr. Timothy Carr and Marsha Chalker Hopkins at the Department of Offender Rehabilitation and of Donnie Lee and Susan Aiken at the Board of Pardons and Paroles. We are also indebted to the staff of the Georgia Supreme Court, especially Joline B. Williams, the Clerk of the Court; the staff of the Georgia Department of Archives and History (which archives retired Georgia Supreme Court records); and the staff of the Department of Human Resources, especially William McWade, Michael R. Lavoie, and Gilmer Chatham of the Vital Records Unit. We also want to acknowledge the valuable contribution of Edward Gates and Kathy Christian, who were our principal data collectors in Georgia, and the research assistance of law students at the University of Iowa, particularly the contribution of Kurt Schroeder, Mary Thee, and Karen Romano. This study would not have been possible without the assistance and support of all these people.

\*\*Joseph B. Tye Professor of Law, University of Iowa College of Law. A.B. 1957, Dartmouth College; M.A. 1962, University of Pittsburgh; LL.B. 1964, LL.M. 1969, Yale University.

\*\*\*Associate Professor of Statistics, University of Iowa. B.A. 1962, Carleton College; Ph.D. 1966, University of Minnesota.

\*\*\*\*Professor of Law, Arizona State University College of Law. B.A. 1964, LL.B. 1967, Yale University.

*and demonstration project — Proportionally Review in Death Sentence Cases. His approach should be of great interest to state supreme courts that want to establish an empirically sound and systematic method for conducting comparative proportionality reviews in capital sentence cases without having to rely on computer technology or complicated statistics. Our analysis of the Georgia data convinces us that without some method of comparing the relative culpability or deathworthiness of all death eligible defendants, the ability of a state supreme court to identify excessiveness or discrimination in its capital sentencing process will be quite limited.*

*In our analysis of the Georgia data, we employed a computerized method of measuring case culpability which differs somewhat from that of Professor Barnett. In spite of the methodological differences between our two studies, however, the substantive results are comparable. Both studies conclude that over one-half of the death sentence cases do not appear to be excessive or disproportionate in a comparative sense. Each also concludes that a good proportion of the remaining death sentence cases in the data set may be excessive in a comparative sense.*

*Perhaps more significantly, both studies show a comparable race-of-victim effect (which disadvantages defendants whose victims are white) among the mid-range of cases where the facts do not clearly dictate either a life or death sentence. Finally, neither study shows a race-of-defendant effect on a statewide basis. However, when urban and rural cases are analyzed separately, our study shows a race-of-defendant effect in rural areas which puts black defendants with white victims at a slight disadvantage. In urban areas, our analysis reveals a race-of-defendant effect which puts white defendants at a distinct disadvantage.*

#### INTRODUCTION

Professor Arnold Barnett's article "Some Distribution Patterns for the Georgia Death Sentence"<sup>1</sup> applies an imaginative and useful method of comparing sentences in different cases to a data set from Georgia that we collected between 1979 and 1981. We collected the data in order to examine the impact of the capital sentencing reforms that Georgia adopted after *Furman v. Georgia*<sup>2</sup> and to assess the implications of any observed changes in the sentences imposed in light of *Furman* and *Gregg v. Georgia*.<sup>3</sup> That study is essentially complete, and

---

<sup>1</sup> Barnett, *Some Distribution Patterns for the Georgia Death Sentence*, 18 U.C. DAVIS L. REV. 1327 (1985).

<sup>2</sup> 408 U.S. 238 (1972).

<sup>3</sup> 428 U.S. 153 (1976).

we contemplate its publication within the next year.

We have accepted an offer to comment on Professor Barnett's article because his findings are so comparable to ours and because we believe the implications to be drawn from both his findings and ours are more significant than he acknowledges. We also believe that Professor Barnett underestimates the ability of multiple regression methods to measure relative case culpability in sentencing studies and that he overlooks some problems associated with the measure of relative culpability that he applies.

We believe that no perfect measure of relative case culpability exists, that different measures have different strengths and limitations, and that one should use as many different measures as time and resources allow. We make this last recommendation because consistency in culpability rankings produced by alternative measures can enhance one's confidence in the reliability of each.

Part I of this Article describes and evaluates the methods used by Professor Barnett and in our study to measure relative case culpability. Part II compares our respective findings concerning the degree of arbitrariness and discrimination observed in Georgia's post-*Furman* system of imposing capital sentences.

## I. MEASURING CASE CULPABILITY

### A. *The Problem*

A capital-sentencing system operates arbitrarily when, without apparent justification, it occasionally imposes a death sentence in cases which, because of their circumstances, usually result in only a prison sentence.<sup>4</sup> We call such death sentences comparatively excessive. Moreover, they are unconstitutional because the defendants can be distinguished in "no principled way" from the other defendants in the life sentence cases.<sup>5</sup> In addition, if one can explain such otherwise excessive death sentences by a factor, such as race, that is prohibited by the equal protection clause, a claim of discriminatory sentencing may also exist. In either case, however, the process of identifying arbitrary or discriminatory death sentences requires a comparison of the sentences imposed in different cases which, in terms of legitimate sentencing criteria, should all be treated the same.

For this reason, the key methodological challenge to any sentencing

---

<sup>4</sup> See Baldus, Pulaski, Woodworth & Kyle, *Identifying Comparatively Excessive Sentences of Death: A Quantitative Approach*, 33 STAN. L. REV. 1, 9-16 (1980).

<sup>5</sup> *Godfrey v. Georgia*, 446 U.S. 420, 433 (1980).

study of this sort is developing valid procedures for identifying those cases which should all be treated the same.<sup>6</sup> Ideally, of course, as Professor Barnett suggests, one should include in the pool of cases deemed "similar" only cases which are factually identical in all pertinent respects to the particular death sentence case or cases being evaluated. The problem with this approach, however, is that rarely, if ever, do a sufficient number of other cases on "all fours" exist to permit a meaningful analysis.<sup>7</sup> Thus, it becomes necessary to develop a system for classifying cases as "similar" in terms of more generally applicable criteria than strict factual identity. If valid, such criteria will permit one to identify cases that are "similar" in terms of relative culpability or deathworthiness despite factual differences.<sup>8</sup>

There are two basic approaches to such an undertaking — the *a priori* and the empiric. The *a priori* approach endeavors to classify cases as "similar" on the basis of criteria which, from a legal or moral perspective, should govern the appropriate sentence. In the capital sentencing context, therefore, one might classify as "similar" all cases involving the same statutorily designated aggravating circumstances.<sup>9</sup> For

---

<sup>6</sup> Appellate courts conducting so-called comparative proportionality reviews in death sentence cases confront the same challenge. See *Pulley v. Harris*, 104 S. Ct. 871 (1984). That review process requires them to determine whether any particular death sentence is "excessive or disproportionate" in comparison to other, similar cases. *Id.* at 876 n.7. Thus, deciding which other cases are similar for purposes of this comparison is a crucial step in the comparative review process. See Baldus, Pulaski & Woodworth, *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 J. CRIM. L. & CRIMINOLOGY 661, 664 (1983); Van Duizend, *Comparative Proportionality Review in Death Sentence Cases: What, How and Why*, 8 STATE CT. J., Summer 1984, at 9.

<sup>7</sup> In legal parlance, two or more cases on "all fours" with each other are identical with respect to all relevant factors.

<sup>8</sup> A ranking of each case in the jurisdiction in terms of its relative culpability or deathworthiness provides a culpability map which permits a reviewing court to conduct comparative proportionality reviews with relative facility. When a death sentence case comes before it for review, the court adds the review case to the culpability map according to its relative culpability level. This step will identify the other cases in the jurisdiction with the same general level of relative culpability and will provide the basis for the court's comparative review. To be sure, the process of ranking all the prior cases in the jurisdiction may involve a substantial undertaking, but, once that process is complete, the court can review subsequent death penalty cases with relative dispatch.

<sup>9</sup> Statutorily designated aggravating circumstances are those factors identified by the legislature as necessary prerequisites for the imposition of a death sentence. Unless one or more of these factors is present in a case, making it "death eligible," the sentencing judge or jury has no authority to consider imposing capital punishment. In addition, some state statutes also specify certain mitigating circumstances which the sentencing authority should consider when imposing sentence. These, too, could serve to sort the

example, under the Georgia statute, any defendant convicted of capital murder in connection with an armed robbery is eligible for the death penalty.<sup>10</sup> On that basis, one might consider all armed robbery murder cases to be of comparable culpability and, therefore, "similar."

Experience proves, however, that armed robbery murder cases from Georgia do not all result in the same sentence. This is also true for every other group of cases that share a common statutory aggravating circumstance.<sup>11</sup> Furthermore, this variation in sentencing results appears to reflect the presence or absence of other, nonstatutory aggravating and mitigating factors which seem to influence prosecutors and juries in such cases. In other words, the existence of a particular statutorily designated aggravating circumstance in a group of cases does not by itself serve effectively to classify them as "similar" for comparative purposes. Thus, one must refine the mechanism for selecting "similar" cases by taking into account additional, nonstatutory aggravating and mitigating circumstances that also seem relevant from one's *a priori* perspective.

The empirical approach also begins by presupposing that certain factual characteristics in the case being reviewed will serve to divide other cases into "similar" and dissimilar categories. In contrast to the *a priori* approach — which selects those factual characteristics on a normative basis — the empirical approach employs these factors which best explain the observed sentencing results. Obviously, some overlap between the criteria selected by each of these two methods should be expected and, indeed, may be substantial. The difference between the two methods is that the *a priorist* selects those factors that he believes should influence the sentencing decision, while the empiricist selects

---

cases into different subgroups of "similar" cases for comparative purposes on an *a priori* basis. Georgia law specifies ten statutory aggravating circumstances and no mitigating circumstances, GA. CODE ANN. § 27-2534.1(b) (1983).

<sup>10</sup> *Id.* § 27-2534.1(b)(2).

<sup>11</sup> The death sentencing rates for death eligible cases under each of the statutory aggravating circumstance were as follows:

those factors that actually appear to do so.<sup>12</sup>

---

DEATH SENTENCING RATES FOR DEATH ELIGIBLE DEFENDANTS  
UNDER EACH OF GEORGIA'S STATUTORY AGGRAVATING FACTORS:  
POST-FURMAN<sup>1</sup>

| A   | B                     |
|---|-----------------------|
| Georgia Statutory Aggravating Factors                                   | Death Sentencing Rate |
| 1. Prior capital record<br>(B-1)  | .34 (14/41)           |
| 2. Enumerated contemporaneous offense, including<br>armed robbery (B-2) | .37 (100/272)         |
| 3. Risk of death to 2 or more in public<br>(B-3)                        | .23 (22/97)           |
| 4. Money/value motive<br>(B-4)  | .37 (79/213)          |
| 5. Victim/judicial officer<br>(B-5)                                     | ____ <sup>a</sup>     |
| 6. Murder for hire<br>(B-6)   | .20 (4/20)            |
| 7. Murder vile, horrible or inhuman<br>(B-7)                            | .30 (94/309)          |
| 8. Victim/police or fire person<br>(B-8)                                | .35 (6/17)            |
| 9. Defendant prisoner or escapee<br>(B-9)                               | .56 (10/18)           |
| 10. Killing to avoid/stop arrest<br>(B-10)                              | .33 (41/125)          |

---

<sup>1</sup> This measure refers to the presence of a statutory aggravating factor in a case regardless of whether it was found by the jury or whether there was even a penalty trial in the case. Many cases have more than one statutory aggravating factor present.

<sup>a</sup> "\_\_\_\_" means no cases.

<sup>12</sup> When selecting other "similar" cases for the purpose of conducting comparative proportionality reviews in death penalty cases, state supreme courts appear to employ a combination of these two approaches. They frequently begin this selection process by classifying cases as presumptively "similar" in terms of the statutory aggravating circumstances. Then, based on an intuitive "feel" for prior sentencing results and *a priori* notions of what nonstatutory factors should affect the sentencing decision, they refine their choice of other "similar" cases in light of the nonstatutory aggravating and mitigating factors present in the case being reviewed.

Still another approach involves an assessment of the overall culpability of the case

### B. Two Empirical Measures of Relative Case Culpability

Although the empirical approach seeks to identify those characteristics or features of the cases that best explain all the sentences actually imposed, the techniques employed to carry out this function can vary substantially. One approach, similar to that employed by some state supreme courts when conducting comparative proportionality reviews, involves an examination of the facts of the death sentence case under review and a selection of certain key factors, to be used for selecting other "similar" cases for comparative purposes, on the basis of intuition and common sense.

This, essentially, is the technique that Professor Barnett employed to construct his system for classifying cases as "similar." However, Professor Barnett was extremely systematic in his procedure. Nor did he limit his selection criteria for identifying "similar" cases to their major factual characteristics. Rather he studied short factual summaries (100-300 words) of more than 500 cases in our data set and, using his judgment, identified those case characteristics that best explained the sentencing results in all those 500 cases.<sup>13</sup> The ultimate product of this effort was a meta-variable classification scheme that permitted him to categorize cases as "similar" or dissimilar in terms of the following three dimensions:

1. Certainty defendant was a deliberate killer — (0) = No, (1) = (Neither 0 nor 2), (2) = Yes,
2. Close relationship between defendant and victim — (0) = Yes, (1) = No,
3. Vileness of the killing — (0) = (Elements of Self Defense), 1 = (Neither 0 nor 2), (2) = (Vile Killing)

Professor Barnett classified the cases in terms of these three dimensions according to the presence or absence of thirty-six more specific variables.<sup>14</sup> For example, if a case involved "mutilation" or any one of thir-

---

under review and a selection of a pool of "similar" cases which are of comparable culpability without much regard to factual similarities. Presumably, courts that use this overall approach assess the culpability of each case considered for the pool of "similar" cases by weighing the aggravating and mitigating factors present in accordance with unarticulated criteria concerning the factors that should determine the appropriate sentence. See Baldus, Pulaski & Woodworth, *supra* note 6, at 675-78; Van Duizend, *supra* note 6, at 10-11.

<sup>13</sup> From a different perspective, one can use the combined "weights" of all the relevant variables present in a particular case to predict what the actual sentence imposed would be, based on the sentences imposed in all other cases.

<sup>14</sup> See Barnett, *supra* note 1, at 1338-42.

teen other nasty case characteristics, it received a score of two on the vileness dimension. This three-dimensional classification scheme produced a total of eighteen potential categories of "similar" cases, each of which could then be examined for evidence of excessiveness or discrimination.<sup>15</sup>

A different approach, which we have employed in our analyses of the Georgia data, involves the use of multiple regression analysis to identify statistically the factors that best explain who is sentenced to death.<sup>16</sup> To carry out this process, we first collected information in every case concerning a large number of variables that might have influenced the sentencing decisions.<sup>17</sup> We then computed for each variable a regression coefficient (or "weight") that reflected its individual power in explaining the sentencing outcomes.<sup>18</sup> Next, we calculated the relative culpabil-

---

<sup>15</sup> See *infra* Figure 1 for the distribution of cases among the 18 possible cells.

<sup>16</sup> Our final report also applies two *a priori* measures, one based on the number of statutory aggravating factors in the case and another based on an intuitive weighing of aggravating and mitigating circumstances.

<sup>17</sup> The file included data on over 200 background variables concerning the defendant, the victim and the crime. The data sources are described in Baldus, Pulaski & Woodworth, *supra* note 6, at 680-81. With a few exceptions, the sample includes all defendants convicted of murder in a jury trial. After a conviction a defendant could receive a death sentence only if (a) the prosecutor requested a penalty trial (which occurred in about one-third of the cases), and (b) the jury imposed the death sentence requested (which occurred in about 55% of the penalty trial cases). When the records examined provided no basis for concluding that a *legitimate* background factor was present in a case (e.g., there was no mention of torture in the case) we would treat the factor as not present. However, when the record was silent concerning an *illegitimate* factor (such as the victim's race) or a procedural feature of the case (such as whether there was a penalty trial), we treated that factor as "missing data." In 5 cases or 1% of the 607 cases we studied, the race of victim was unknown (because the state of Georgia had no record of the decedent's death certificate), and in 23 cases or 4% of the cases there was no indication that a penalty trial had occurred. In these cases we imputed the missing race-of-victim data on the basis of the defendant's own race. We imputed whether a penalty trial had occurred when that information was missing with a statistical model that best explained which cases advanced to a penalty trial.

The impact of the "missing data" on our overall analysis was slight. When we deleted from the analysis the five cases missing data on the race-of-victim variable, the coefficient estimated for the race-of-victim effect did not change. The missing race-of-victim data did not affect our analyses of excessiveness at all because neither the victim's race nor the conduct of a penalty trial was a relevant variable for that purpose. The missing data on whether there was a penalty trial affected only the separate analysis of prosecutorial decisions. When we deleted from the analysis the cases with missing data on that variable and the race of victim, the logistic regression coefficient estimated for the race-of-victim variable declined from 1.87 to 1.76.

<sup>18</sup> See *infra* note 23 for a listing of the variables and coefficients used.



ity or deathworthiness of each case by summing the "weights" of all the important explanatory variables present in each case.<sup>19</sup> We then ranked all the cases according to their relative culpability scores, thereby constructing an overall culpability index along which the cases were distributed. Finally, we defined as "similar" all cases with comparable overall culpability scores.<sup>20</sup>

One consequence of the way in which these relative culpability measures are computed is that the impact of an aggravating factor in a given case, which enhances the risk of a death sentence (for example, the victim was taken hostage), may be offset by a mitigating factor (for example, defendant was an underling). As a result, a case with both aggravating and mitigating factors may yield a culpability score comparable to that of a case with no special aggravating or mitigating features. In other words, the cumulative nature of this scoring process can frequently produce similar scores for cases which are factually distinct. But, in terms of their relative culpability or deathworthiness in the eyes of the prosecutors and juries who processed the universe of cases from which the culpability index is derived, such factually different cases do constitute "near neighbors."<sup>21</sup>

For our study of excessiveness and discrimination we sorted the cases in the data set according to their relative culpability scores and then divided them into groups with comparable scores.<sup>22</sup> We then calculated

---

<sup>19</sup> Since a logistic regression analysis was used to estimate the coefficients or "weights" for each variable used to create the index, the final culpability score for each case is the log of the odds of receiving a death sentence.

<sup>20</sup> For a description of comparable methods of scaling the seriousness of offenses in the non-capital sentencing context, see Nevares-Muniz, *The Eighth Amendment Revisited: A Model of Weighted Punishments*, 75 J. CRIM. L. & CRIMINOLOGY 272, 278-80 (1984).

<sup>21</sup> Professor Barnett questions the "conceptual similarity" of the cases deemed comparable on a regression based index. "One case," he states, "could involve a web of aggravating and mitigating factors, while the other is comparatively nondescript." Barnett, *supra* note 1, at 1360. We agree, but do not consider this feature of the index to be a problem. The approach reduces many variables to a single dimension — the index which provides the best estimates of who will be sentenced to death. The conceptual similarity of groups of near neighbors on this index appears quite clear to us — people judged to be of comparable culpability or deathworthiness by Georgia's prosecutors and juries. What could be more relevant? We also note that Professor Barnett's classification system produces groups of similar cases which are not factually comparable. The difference is simply that he classifies the cases in terms of three or four dimensions rather than one.

<sup>22</sup> These categories are in terms of equal intervals on the culpability index, except for the least and most aggravated groups of cases which are in levels 1 and 6, respectively.

the death sentencing rate within each subgroup, both as a whole, to measure excessiveness, and after further disaggregation according to the racial characteristics of the cases, to measure race effects.<sup>23</sup>

---

<sup>23</sup> We created separate indices for (a) the prosecutorial decision to advance a case to a penalty trial, (b) the jury life or death sentencing decision, and (c) the combined effects of both the prosecutorial decisions. In this note we only report the variables and coefficients for the combined effect of the prosecutorial and jury decisions (which together determine the likelihood of a death sentence given a murder conviction at trial.) For this purpose we created three indices (OVERALLA, OVERALLB, and OVERALLC), the variables and coefficients for which appear below.

| Variables  | A<br>OVERALLA                         |                              | B<br>OVERALLB                         |                 | C<br>OVERALLC                         |                 |
|--|---------------------------------------|------------------------------|---------------------------------------|-----------------|---------------------------------------|-----------------|
|  | Logistic Coefficient (Standard Error) | Odds Multiplier <sup>1</sup> | Logistic Coefficient (Standard Error) | Odds Multiplier | Logistic Coefficient (Standard Error) | Odds Multiplier |
|  |                                       | P Value                      |                                       | P Value         |                                       | P Value         |
| Black Defendant (BLACKD)   | -.57<br>(.50)                         | .25                          | .00                                   |                 |                                       |                 |
| White Victim (WHVICRC)   | 2.66<br>(.65)                         | .0001                        | .16                                   |                 |                                       |                 |
| One or More Mitigating Circumstances and a White Victim (MCIRAWHV)   | -1.65<br>(.54)                        | .002                         | .11                                   |                 |                                       |                 |
| Low Status Defendant (LSTATDEF)  | .85<br>(.48)                          | .08                          | .05                                   |                 |                                       |                 |
| High Status Defendant (HISTD)  | 1.88<br>(.95)                         | .05                          | .06                                   |                 |                                       |                 |
| Female Defendant (DFEM)  | -.59<br>(1.03)                        | .56                          | .00                                   |                 |                                       |                 |
| # of Statutory Aggravating Circumstances plus the number of B7 (vile murder) circumstances in the case (DELB7EX) | 1.22<br>(.20)                         | .00001                       | .25                                   | 1.05<br>(.16)   | 2.9                                   | .00001          |
|  |                                       |                              |                                       |                 | 1.22<br>(.20)                         | 3.4             |
|  |                                       |                              |                                       |                 |                                       | .0001           |

|  | A<br>OVERALLA   |             | B<br>OVERALLB |                | C<br>OVERALLC |                             |
|--|-----------------|-------------|---------------|----------------|---------------|-----------------------------|
| Female Victim<br>(FEMVIC)                          | 2.26<br>(.52)   | 9.5 .00001  | .17           | 1.87<br>(.47)  | 6.5 .0001     | 2.26<br>(.52) 9.5 .00001    |
| Def. Underling<br>in the Murder<br>(DUNDERLG)      | -6.88<br>(1.51) | .001 .00001 | .18           | -4.95<br>(1.3) | .007 .0001    | -6.88<br>(1.51) .001 .00001 |
| Victim Stranger<br>(VICSTRAN)                      | 1.00<br>(.51)   | 2.7 .051    | .06           | 1.34<br>(.47)  | 3.8 .005      | 1.00<br>(.51) 2.7 .051      |
| Multiple Stabs<br>(MULSTAB)                        | 1.71<br>(.71)   | 5.5 .01     | .09           | 1.67<br>(.59)  | 5.3 .005      | 1.71<br>(.71) 5.5 .01       |
| Defendant Killed<br>Two or More People<br>(TWOVIC) | 1.66<br>(.65)   | 5.3 .01     | .09           | 1.56<br>(.61)  | 4.8 .01       | 1.66<br>(.65) 5.3 .01       |
| Victim Hostage<br>(HOST)                           | 3.69<br>(2.03)  | 40 .07      | .05           | 3.20<br>(1.45) | 24.5 .027     | 3.69<br>(2.03) 40 .07       |
| Def. Cooperated<br>with Authorities<br>(STMIT9)    | 2.64<br>(.81)   | 14 .001     | .12           | 2.24<br>(.75)  | 9.4 .003      | — — —                       |
| Victim Police<br>or Fire person<br>(VICPFIR)       | 2.00<br>(.89)   | 7.4 .02     | .07           | .94<br>(.75)   | 2.6 .09       | 2.00<br>(.89) 7.4 .02       |
| Victim 12 yrs.<br>or Younger<br>(YNGVIC)           | 2.3<br>(.97)    | 10 .02      | .08           | 1.51<br>(.92)  | 4.5 .09       | 2.3<br>(.97) 10 .02         |
| Def. lay in<br>wait<br>(AMBUSH)                    | 2.16<br>(.63)   | 8.6 .001    | .13           | 1.78<br>(.55)  | 6.0 .001      | 2.16<br>(.63) 8.6 .001      |

|   | A<br>OVERALLA   |               | B<br>OVERALLB |                | C<br>OVERALLC |        |                |      |        |
|---|-----------------|---------------|---------------|----------------|---------------|--------|----------------|------|--------|
| Def. in Military<br>(MILDEFN)   | 2.43<br>(.86)   | 11.4<br>.005  | .10           | 1.86<br>(.80)  | 6.4           | .02    | 2.43<br>(.86)  | 11.4 | .005   |
| Race Hatred Motive<br>(RACE)  | 4.68<br>(1.97)  | 108<br>.02    | .08           | 4.12<br>(.54)  | 61.6          | .008   | 4.68<br>(1.97) | 108  | .02    |
| Victim Low Stat.<br>(VICLSTAT)  | -2.63<br>(.80)  | .07<br>.001   | -.12          | -2.49<br>(.78) | .08           | .001   | —              | —    | —      |
| A Bloody Murder<br>(BLOODY)   | 1.43<br>(.49)   | 4.18<br>.004  | .10           | 1.05<br>(.42)  | 2.9           | .01    | 1.43<br>(.49)  | 4.18 | .004   |
| Kidnap and<br>Multiple Shots<br>(KDNPAMSH)  | -2.60<br>(1.03) | .07<br>.012   | -.09          | -1.22<br>(.94) | .30           | .19    | —              | —    | —      |
| Neither Kidnap nor<br>Multiple Shots<br>(NKNPOMSH)  | -2.46<br>(.52)  | .09<br>.00001 | -.19          | -2.14<br>(.45) | .12           | .00001 | -2.46<br>(.52) | .09  | .00001 |
| Aggravated Motive<br>and 1 or more prior<br>felony convictions<br>(AGMOTAPF)  | 1.69<br>(.58)   | 5.4<br>.004   | .11           | 1.36<br>(.51)  | 3.9           | .008   | 1.69<br>(.58)  | 5.4  | .004   |
| Number of Convictions for<br>violent personal crimes (other than<br>murder, rape, armed robbery,<br>kidnap) and number of statutory<br>aggravating circumstances<br>interaction<br>(DLXXW15D) | 0.36<br>(.19)   | 1.4<br>.07    | .05           | .25<br>(.18)   | 1.3           | .17    | .36<br>(.19)   | 1.4  | .07    |
| Defendant Resisted<br>Arrest and 1 or more<br>prior felony convictions<br>(DEFRAAPF)  | 1.38<br>(.70)   | 4<br>.06      | .05           | 1.22<br>(.65)  | 3.9           | .06    | 1.38<br>(.70)  | 4    | .06    |

<sup>1</sup> The odds multiplier is the anti-log of the logistic partial regression coefficient.

## *B. Strengths and Weaknesses of the Alternative Culpability Measures*

### 1. Strengths

A test of the strength of an empirical measure of relative case culpability is how well it explains the actual results. An examination of Figure 1 and Table 1 below indicates that both measures distinguish quite well between the cases in which death sentencing rates are quite high and those cases in which there is no risk or only a negligible risk of a death sentence.

A second test of the strength of a measure is the validity of the culpability scores it produces. We validated the culpability scores produced with our regression based indices by reading narrative summaries of the cases deemed to be "similar" and by assessing intuitively the plausibility of their relative culpability rankings. We also found considerable consistency between the case culpability classifications produced by our index and by Professor Barnett's system.<sup>24</sup>

A third strength of both Professor Barnett's classification system and our regression-based indices is that they permit one to use both cross-tabular and regression methods of statistical analysis. Professor

---

We estimated the coefficients used for the excessiveness index (OVERALLC) by means of a logistic regression (OVERALLA) which included all legitimate variables with a statistically significant relationship (beyond the .10 level) to the sentence imposed, as well as variables for the race, sex, and socio-economic status of the defendant and the race of the victim (items 1-5). We produced OVERALLC by purging from model OVERALLA, the sex, race, and other suspect variables plus two other variables the statistical impact of which on the sentencing decision was counter-intuitive. See *infra* note 33 for a discussion of these two counter-intuitive variables. The purging of ethically questionable variables is suggested in RESEARCH ON SENTENCING: THE SEARCH FOR REFORM 147 (A. Blumstein, J. Cohen, S. Martin & M. Tonry eds. 1983). We also used model OVERALLA to estimate partial regression coefficients for the racial variables while controlling for statistically significant ( $p < .10$ ) legitimate variables and the illegitimate and suspect variables in the model. To measure race effects with a crosstabular analysis which did not adjust for other illegitimate and suspect variables, we created index OVERALLB and a scale based on it. Race of victim results for the prosecutorial and jury decisions are shown in *infra* note 46.

<sup>24</sup> See *infra* note 39 for the results of our comparison of the culpability classifications produced by the two systems. Also Professor Barnett conducted a series of experiments to determine the "reliability" of his system. He found that with only a limited amount of training, student coders were able to replicate his results with a high degree of consistency. See Barnett, *supra* note 1, at 1371.

Barnett's analysis is limited to cross-tabular methods, but one can easily adapt his coding scheme to multiple regression analysis.<sup>25</sup> The ability to use regression techniques is important because it permits one to analyze small samples. If one wants to look into the nooks and crannies of a data set (for example, measuring race effects in given geographic localities), declining sample sizes quickly limit the statistical power of cross-tabular methods compared to regression analysis.

On the other hand, a great strength of Professor Barnett's classification system is that it does not employ complicated statistical methods or require the use of a computer. Professor Barnett has demonstrated that, even with a very large data set, a perceptive analyst can develop a highly explanatory model on the basis of intuition and common sense. Indeed, Professor Barnett identified several variables superior in this respect to those which we developed for our coding instrument before data collection began. For example, his variable for the relationship between the defendant and victim was considerably more explanatory than our analogous variable, and his variable for the "deliberateness" of the killing filled a significant gap in our data set.<sup>26</sup>

This aspect of Professor Barnett's classification system deserves special emphasis. State supreme courts interested in conducting a systematic proportionality review in death sentence cases will find Professor Barnett's approach to be a useful starting point for selecting "similar" cases on the basis of relative case culpability. We applaud this development because we are convinced that a system of comparative proportionality review can only be effective if it permits the reviewing court to compare the relative culpability of all the death eligible cases within its jurisdiction. Only with this overview can the reviewing court effectively limit the death penalty to the most extreme cases. Professor Barnett's methodology will permit courts to gain this empirically based overview

---

<sup>25</sup> For example, in the regression analysis reported in *infra* note 39, we used the following variables coded from the three Barnett dimensions:

MIDINTNT: Code 1 if Barnett Deliberateness Dimension = 1 else Code 0.

HIINTENT: Code 1 if Barnett Deliberateness Dimension = 2 else Code 0.

VCSTRANG: Code 1 if Barnett Victim Dimension = 1 else Code 0.

MIDVILE: Code 1 if Barnett Vileness Dimension = 1 else Code 0.

HIVILE: Code 1 if Barnett Vileness Dimension = 2 else Code 0.

PRIORREC: Code 1 if Barnett Prior Record Dimension = 1 else Code 0.

<sup>26</sup> When included in the index we used to measure excessiveness (OVERALLC in *supra* note 23) the logistic regression coefficient estimated for Professor Barnett's defendant victim relationship dimension was 4.4 ( $p=.001$ ), while the two variables created for his deliberateness dimension yielded coefficients of 2.17 ( $p=.001$ ) and 4.2 ( $p=.003$ ). His measure of vileness, however, was no better than our comparable measures and added no explanatory power to our index.

without recourse to computer technology or to complex statistics.

Having said this, however, we must also stress that, compared to intuitively driven systems like Professor Barnett's, the great advantage of a formal, statistically based regression approach is its capacity systematically to screen large numbers of variables and to identify those which provide additional explanatory power despite the presence in the analysis of other variables with similar or even greater explanatory power. As a consequence, a regression-based measure of relative culpability is less likely to overlook relevant case characteristics than a measure based simply on intuition and common sense.<sup>27</sup>

## 2. Weaknesses

All statistical methods have limitations which may affect the validity of the results they produce. However, when the object of the procedure is simply to measure the relative culpability of a group of cases clustered together on a culpability index, validation of the results is a straightforward matter. Whatever statistical method is used to measure relative case culpability, the bottom line result is a ranking of the cases from most to least deathworthy. A qualified evaluator can test the validity of that ranking by reading factual summaries of the cases and then determining whether the statistical measure employed properly ranked them in terms of overall culpability. This process also permits the correction of misranked cases. Thus, the final, validated determinations of relative case culpability are the product of both an objective, statistical analysis and the validator's intuitive judgment.<sup>28</sup> Consequently, concerns about the ranking process that focus on assumptions underlying the use of regression analysis become inconsequential.

---

<sup>27</sup> Professor Barnett's analysis omitted the following legitimate variables that we found to be statistically important ( $p=.10$ ) in explaining the death sentencing result after controlling for his three underlying dimensions and his prior record variable:

1. Number of statutory aggravating factors, plus the number of multiple B7 (Vile murder) circumstances in the case (DELB7EX) ( $b = .82$ , SE .16)

2. Female victim (FEMVIC) ( $b = 1.1$ , SE 4.5)

3. Victim 12 or younger (YNGVIC) ( $b = 2.1$ , SE 1.03)

4. Bloody murder (BLOODY) ( $b = 1.3$ , SE .45)

5. Multiple stab wounds (MULSTAB) ( $b = 1.6$ , SE .68)

6. Two or more victims (TWOVIC) ( $b = 1.5$ , SE .63)

7. Three terms representing combinations of two legitimate variables, *see* in Model OVERALLA, *supra* note 23, variables AGMOTAPF, DEFRAAPF, NKNPOMSH.

<sup>28</sup> Van Duizend, *supra* note 6, at 11, suggests that for comparative proportionality review "a set of criteria should be developed through a combination of intuitive and empirical approaches with each approach used to check and inform the other."



There is a threat, however, to the validity of both our culpability indices and Professor Barnett's classification system: the risk of omitting relevant variables. In any study of the sort involved here, it is possible for researchers to overlook relevant variables for which data are available. For example, Professor Barnett's classification system omits variables that have important explanatory power.<sup>29</sup> Similarly, our indices do not reflect variables which Professor Barnett subsequently developed and coded after he reviewed narrative summaries of the cases.<sup>30</sup>

The validity of a regression-based culpability index may also be threatened by the inclusion of inappropriate variables.<sup>31</sup> None of the variables in our data set is inappropriate per se. We selected each one

---

<sup>29</sup> See *supra* note 27.

<sup>30</sup> Another potential problem concerns variables which may have influenced the sentencing decision, but were omitted from the analysis because no data on them could be found in the available records. Such potentially influential omitted variables would include the defendant's demeanor, the credibility of witnesses, and the competency of the opposing attorneys. We did attempt to control for the strength of the evidence by limiting our universe of cases to those in which the evidence resulted in a murder conviction before a jury. We believe that for the uncoded, potentially relevant variables, such as the defendant's demeanor, the risk of bias is quite small given the large number of variables for which we do have information. Nor do we consider it likely that any omitted factors would be significantly correlated with both the defendant's race or sex and with the sentencing result, the condition under which the omission of these variables would bias the results of the racial analysis. We consider it more likely that the distribution of these omitted variables is random with respect to the racial characteristics of the cases. Most importantly, even though these omitted variables may have significantly influenced the outcome in individual cases, we consider it very improbable that they could systematically bias the culpability ranking of the cases used to measure excessiveness.

Two other sources of bias raised by Professor Barnett deserve mention. Data for one-half of the cases were obtained from Georgia Supreme Court records, primarily trial transcripts and briefs. For these cases we would not expect bias in the reported data. We coded the other half of the cases from the files of the Georgia Board of Pardons and Paroles. Because the Parole Board investigators obtained the data in these files after sentencing from local court records, police reports, and conversations with police and prosecutors, the information presented might tend to shade the facts in a way that supports the sentence imposed. If such a bias does exist, however, it would tend to minimize observable race effects in the results rather than to exaggerate them. Consequently, this potential form of bias is not a real concern.

Another dimension of the omitted variable problem, known as "sample selection bias," may be particularly important in a winnowing process of the type involved in Georgia's capital sentencing system. See RESEARCH ON SENTENCING: THE SEARCH FOR REFORM 93-110 (A. Blumstein, J. Cohen, S. Martin & M. Tonry eds. 1983) for a non-mathematical discussion of the issue. We will present an analysis of this problem in our final report.

<sup>31</sup> See Barnett, *supra* note 1, at 1337-38.

because it might plausibly influence a capital sentencing decision. However, for certain variables, the "sign" of the regression coefficient was contrary to what we expected (for example, a mitigating variable yielded a positive sign, giving it an aggravating "weight"). This suggested the possibility that including such variables in the model was inappropriate.<sup>32</sup> Alternatively, the sign of such a coefficient might reflect perversity or arbitrariness in the operation of the system. Our final model included two variables whose coefficients ran in an unexpected direction. On the basis of further analysis, we attributed one of these contrary coefficients to perversity in the system and one to possible misspecification.<sup>33</sup>

Professor Barnett's article also discusses certain threats to the validity of regression analyses that are relevant to those portions of our study which employ partial regression coefficients to estimate the effect of racial factors. One such threat involves the assumptions underlying ordinary least squares regression analysis and the consequences of failing to include interaction terms in such an analysis. We fully share Professor Barnett's concerns in this respect and have attempted to deal with them in our final analysis.<sup>34</sup>

---

<sup>32</sup> In methodological parlance, the perverse sign is evidence of possible "misspecification" of the model.

<sup>33</sup> In model OVERALLA, *see supra* note 23, the coefficients for one mitigating factor (defendant cooperated with the authorities (STMIT9)) and for one aggravating factor (kidnapping and multiple shots (KNAPAMSH)) yielded counter-intuitive signs. The sign for the latter variable we cannot explain. The positive association observed between the mitigating factor "defendant cooperation" and imposition of the death penalty, however, may be explained in that cooperative defendants tend to describe fully how they killed the victim. This detailed description may substantially strengthen the case for a death sentence. Whatever the cause of the unexpected signs, however, the rationale for excluding these variables from the model is the same. In a rational system it would be unethical for these factors to influence the results in the direction noted. Moreover, such a result would be inconsistent with the teaching of *Godfrey v. Georgia*, 446 U.S. 420, 433 n.16 (1980), that similar cases should be identified in terms of case characteristics that are rationally related to the purpose of the death sentencing statute. This approach is also consistent with the recent recommendations of the National Research Council's sentencing panel that the influence of unethical variables should be purged from indices used to measure sentencing disparity. *See supra* note 23.

<sup>34</sup> We selected the interactions included in the model with a residual correlation analysis that identified those interactions which would remain in the model at a level of statistical significance of .10 or better after adjustment for all main effect variables. We screened 434 interactions terms between two legitimate variables and 60 interactions terms between a legitimate variable and the race-of-victim or race-of-defendant variables. In comparing these results with the results of a preliminary race-of-victim analysis reported in *Baldus, Pulaski & Woodworth, supra* note 6, at 706-10, we note that

There is one other methodological issue that merits discussion, in part because it is relevant both to Professor Barnett's analysis and to our study. The issue is whether one should adjust a relative culpability index used to measure sentencing disparity for the impact of illegitimate factors such as race. All agree that, when examining disparities in sentences, one should measure relative culpability solely on the basis of legitimate or "rational" variables. Both Professor Barnett's model and our indices conform to this convention. However, a panel of the National Research Counsel (NRC) recently recommended that one should

---

the introduction of the interaction terms, which were not included in the earlier analysis, enhanced both the magnitude and the statistical significance of the race-of-victim effects. The preliminary results of our race-of-victim analysis reported in the *Journal of Criminal Law and Criminology* were intended only to show the relationship between racial discrimination and the principal subject of the article, comparative excessiveness in capital sentencing.

The criticism that additive linear regression does not capture the system's response to particular combinations of circumstances has some merit; however, as noted above we checked for nonadditive responses to combinations of factors by residual correlation analysis and incorporated them. We also tested quadratic terms for the culpability index which did not prove to be significant. These efforts may partly explain why the culpability rankings produced by our aggregation based indices appear plausible, thereby increasing our confidence in the validity of the entire model. We note, however, that whether we employed an ordinary least squares analysis or a logistic regression analysis, the results were essentially the same.

Multicollinearity diagnostics were generated by the REG procedure of the SAS statistical software package. SAS USERS GUIDE: STATISTICS (1982); see D. BELSLEY, E. KUH & R. WELSH, REGRESSION DIAGNOSTICS (1980). We regressed logistic residuals,  $(y-p)/pq$ , against each independent variable where  $p$  is the predicted probability that  $y=1$ , given the values of the independent variables. Collinearity diagnostics produced by this program show the proportion of variance of each regression coefficient loaded on each principal component of the variance-covariance matrix of the independent variables. Collinear sets of variables are identifiable by high loadings on the same minor principal component. One rule of thumb was that two variables loaded on a principal component with a condition index score greater than 30 indicated a problem. See D. BELSLEY, E. KUH & R. WELSH, *supra*, at 105. Each of the collinearity problems thus identified involved two legitimate variables with one of them displaying a perverse sign (such as a positive regression coefficient for a mitigating factor), and in each case we concluded that the perverse sign was due to collinearity and removed from the model the least substantively meaningful variable in the collinear set.

The analysis did not reveal any problems of multicollinearity between a legitimate background variable and the race or suspect variables. More importantly, our primary use of regression was to construct a case matching index and culpability map which do not involve the interpretation of given regression coefficients; for this purpose multicollinearity is an issue only to the extent that it may affect the relative culpability rankings of the cases, an outcome which we verified by a qualitative evaluation of the ranking produced by the index. See *supra* note 24 and accompanying text.

adjust for the effects of suspect variables before calculating the weights to be attributed to legitimate factors.<sup>35</sup> The purpose of this recommendation is to ensure that the weights attributed to legitimate background variables do not incorporate any of the effects of racial or other suspect status variables. Professor Barnett's classification system involved no such adjustment. However, as we will show below, the results produced with his system and our index, which did make the adjustments recommended by the NRC, are essentially the same.<sup>36</sup>

## II. SUBSTANTIVE RESULTS: EVIDENCE OF ARBITRARINESS AND DISCRIMINATION IN OUR TWO STUDIES

### A. Arbitrariness

A principal objective of a selective, non-arbitrary death sentencing system is to limit death sentences to the most extreme cases. Such a system will yield a very high proportion of death sentences among "similar" cases.<sup>37</sup> Figure 1 and Table 1 indicate the extent to which Georgia's post-*Furman* system has achieved this goal. Figure 1 presents the distribution of cases among the eighteen categories gener-

---

<sup>35</sup> See *supra* note 23.

<sup>36</sup> See *infra* Figure 1; Table 1; and Table 2.

<sup>37</sup> See Baldus, Pulaski, Woodworth & Kyle, *supra* note 4, at 64-68; see also Bentele, *The Death Penalty in Georgia: Still Arbitrary*, 62 WASH. U.L.Q. 573 (1985).

ated by Professor Barnett's classification system.

FIGURE 1

DEATH-SENTENCING RATES AMONG SUBGROUPS  
OF CASES DEFINED BY THREE BARNETT DIMENSIONS:  
POST-FURMAN GEORGIA<sup>1</sup>

|                     |                       |                         |                        |                        |                      |
|---------------------|-----------------------|-------------------------|------------------------|------------------------|----------------------|
|                     | .0 (0/1)<br>(0,1,0)   | .04 (1/23)<br>(0,0,2)   | .09 (1/11)<br>(2,0,1)  |                        |                      |
| .0 (0/8)<br>(0,0,0) | .02 (1/43)<br>(1,0,0) | .0 (0/10)<br>(1,1,0)    | .25 (15/60)<br>(1,1,1) | .81 (59/73)<br>(1,1,2) | .88 (7/8)<br>(2,1,2) |
|                     | .0 (0/55)<br>(0,0,1)  | .02 (1/48)<br>(0,1,1)   | .29 (6/21)<br>(0,1,2)  | .56 (10/18)<br>(2,0,2) |                      |
|                     |                       | .005 (1/184)<br>(1,0,1) | .26 (11/42)<br>(1,0,2) |                        |                      |
|                     |                       | .0 (0/1)<br>(2,0,0)     |                        |                        |                      |

*Sum of Scores for the Three Barnett Dimensions*

0                      1                      2                      3                      4                      5

Least Aggravated  $\longrightarrow$  Most Aggravated

<sup>1</sup> The three numbers in parenthesis at the foot of each cell indicate the coding for Barnett's three dimensions ((a) certainty that killing was deliberate (0 = low, 1 = medium, 2 = high), (b) relationship between defendant and victim (0 = family/friend, 1 = stranger), and (c) vileness of the killing (0 = element of self defense, 1 = neither self defense nor vile, 2 = vile killing)). The death sentencing rate and the actual numbers (death cases/all cases) are in the center of each cell. There were no cases in cells (2,1,0) and (2,1,1).

Each cell indicates the number of cases, the number and proportion of death sentence cases, and the characteristics of the cases in terms of the three relevant dimensions. For example, the cell to the extreme right includes the nine cases with the most aggravated set of characteristics, (2,1,2); 88 percent (7/8) of these cases resulted in death sentences.

Table 1 presents a distribution of the cases among six culpability levels based on the multiple regression index we developed to measure comparative excessiveness.<sup>38</sup>

<sup>38</sup> The index underlying Table 1 is based on Model OVERALLC, described in *supra* note 23. Another measure of the arbitrariness of a capital sentencing system is the influence of ethically questionable factors in the system. Model OVERALLA, also described in *supra* note 23, suggests that race of victim and the socio-economic status of both the defendant and victim may influence the process.

TABLE 1

DEATH SENTENCING RATES CONTROLLING FOR CASE  
CULPABILITY LEVEL: POST-FURMAN GEORGIA

| A<br><i>Case Culpability Level<br/>from 1 (low) to 6 (high)<sup>1</sup></i> | B<br><i>Death Sentence Rate</i> |
|---|---------------------------------|
| 1   | .02<br>(6/399)                  |
| 2   | .14<br>(9/65)                   |
| 3   | .38<br>(18/47)                  |
| 4   | .67<br>(22/33)                  |
| 5   | .85<br>(23/27)                  |
| 6   | 1.0<br>(35/35)                  |
| All Cases   | .19 (113/606)                   |

<sup>1</sup> This index is based on the following 18 factors: (a) # of Statutory Aggravating Circumstances in the case; (b) Female Victim; (c) Def. Underling in the murder; (d) Victim Stranger; (e) Multiple Stabs; (f) Defendant Killed Two or More People; (g) Victim Hostage; (h) Victim Police or Fire person; (i) Victim 12 yrs. or Younger; (j) Def. lay in wait; (k) Def. in Military; (l) Race Hatred Motive; (m) Victim Low Stat.; (n) Bloody Murder; (o) Neither Kidnap nor Multiple Shots; (p) Aggravated Motive and 1 or more prior felony convictions; (q) Number of Convictions for violent personal crimes (other than murder, rape, armed robbery, or kidnapping) and number of statutory aggravating circumstances interaction; (r) Defendant Resisted Arrest and 1 or more prior felony convictions.

For each level, Table 1 indicates the number of cases deemed to be "similar" and the death sentencing rate among them. For example, at level 6, the most aggravated level, the table indicates that all thirty-five cases resulted in death sentences.

The results shown in Figure 1 and Table 1 provide a good basis for assessing the level of excessiveness in Georgia's post-Furman capital sentencing system. In terms of the frequency with which death sentences are imposed among similar cases, the cases fall into the following main categories:

1. Very low (a death sentencing rate of less than .02)
2. Low (a rate of from .02 to .34)
3. Medium (a rate of from .35 to .79)
4. High (a rate of .80 or more)

Table 2 indicates how Professor Barnett's analysis and our study distributed the cases in terms of these four categories. For example, columns B and D of row 1 give the actual death sentencing rates for the cases identified as least aggravated by the two studies — .004 in the Barnett analysis and .015 in our analysis.

Table 2  
 Death Sentencing Rates and Distribution of Death Sentenced Defendants Controlling for the Likelihood that Defendants of Similar Culpability will Receive a Death Sentence

|  | BARNETT ANALYSIS   |                                |  | BALDUS, WOODWORTH, PULASKI ANALYSIS |  |  |
|--|--|--------------------------------|--|-------------------------------------|--|--|
|  | A<br>Death Sentencing<br>Rates Among<br>Similar Defendants | B<br>Death Sentencing<br>Rates | C<br>Proportion of Total<br>Death Cases at<br>each Level | D<br>Death Sentencing<br>Rate       | E<br>Proportion of Total<br>Death Cases at each<br>Level |  |
| 1) Very Low<br>(less than .02)   | .004<br>(1/259)  | .01                            | .015<br>(6/399)  | .05                                 |  |  |
| 2) Low<br>(.02 to .34)   | .15<br>(36/248)  | .32                            | .14<br>(9/65)  | .08                                 |  |  |
| 3) Moderate<br>(.35 to .79)  | .56<br>(10/18)   | .09                            | .50<br>(40/80)   | .35                                 |  |  |
| 4) High<br>(.80 to 1.0)  | .81<br>(66/81)   | .58                            | .94<br>(58/62)   | .51                                 |  |  |
|  | 113/606  | 1.0                            | 113/606  |                                     |  |  |
| II. NUMBER AND PROPORTION OF DEATH SENTENCE CASES FOR<br>WHICH THE DEATH SENTENCE FREQUENCY AMONG SIMILAR CASES WAS: |  |                                |  |                                     |  |  |
| A. .34 or less   | .33 (37/113)   |                                |  | .13 (15/113)                        |  |  |
| B. Between .35 and .79   | .09 (10/113)   |                                |  | .35 (40/113)                        |  |  |
| C. .80 or more (apparently evenhanded)   | .58 (66/113)   |                                |  | .51 (58/113)                        |  |  |



The findings from the two studies show considerable similarity. In each, slightly over half of the death sentences appear to be evenhanded, .58 in the Barnett analysis and .51 in ours.<sup>39</sup> His analysis shows, however, a higher proportion of apparently excessive sentences (.33) than does ours (.13). This difference is not surprising given the tendency of multiple regression analyses to produce a unique “overfitted” solution, which likely overstates the consistency of the system being analyzed.

## B. Racial Discrimination

### 1. Race-of-Victim Effects

Figure 2 presents a race-of-victim analysis of Georgia’s capital sentencing system after controlling for Professor Barnett’s case dimensions.<sup>40</sup> It shows the death sentencing rates for white and black victim

---

<sup>39</sup> The classification suggested presumes a death sentence is “presumptively evenhanded” if .80 or more similarly situated defendants receive death sentences, but is “presumptively excessive” if fewer than .35 similarly situated defendants receive a death sentence. See Baldus, Pulaski & Woodworth, *supra* note 6, at 695-98, for a discussion of the legal basis for this classification system.

We found a substantial level of consistency between the two systems in their classification of cases as “extreme” — i.e., those with a death sentence rate of .80 or more. Sixty-eight percent of the 59 death sentence cases in Professor Barnett’s extreme category in Table 2 were similarly classified in our system; 27% of his extreme cases were classified as moderate in our system. Of the 54 cases Professor Barnett classified as less than extreme, our system was in agreement 67% of the time. We also used a regression analysis to rank the cases from most to least culpable with the three Barnett variables and his prior record variable. The correlation coefficient between the index scores produced with this index and our OVERALLC index was .65.

The analyses shown in Figure 1 and tables 1 and 2 do not follow the admonition of Van Duizend in his report on the National Center for State Courts project on proportionality review that “cases in which the conviction or sentence is reversed should be dropped from the pool, regardless of the grounds for reversal.” Van Duizend, *supra* note 6, at 11 (emphasis omitted). In 25% of the death cases in this study, the Georgia Supreme Court either vacated the death sentence or reversed the underlying conviction. When those cases are excluded from the analysis the degree of excessiveness in the system changes only slightly. The proportion of presumptively excessive death sentences drops from .29 to .28 and the proportion of presumptively evenhanded death sentences rises from .51 to .58. See also Baldus, Pulaski & Woodworth, *supra* note 6, at 715.

<sup>40</sup> See *supra* note 13 and accompanying text for a description of his methodology. Professor Barnett also used a variable for prior record in estimating the race-of-victim effects. We did not include that variable in Figure 2 but did include a prior record variable in the index underlying Table 3 in regression results presented below. See *supra* note 23. When we further adjusted for the Barnett prior record variable, the overall disparity shown in Figure 2 dropped to 6.5 points significant at the .01 level.

FIGURE 2: POST-FURMAN GEORGIA RACE OF VICTIM DISPARITIES,  
CONTROLLING FOR THE THREE BARNETT DIMENSIONS<sup>1</sup>

|   |                                   |                                     |                                    |
|---|-----------------------------------|-------------------------------------|------------------------------------|
| 0,0,0<br>(.00)                          | 0,0,2<br>(.04)                    | 2,0,1<br>(.09)                      | 2,1,2<br>(.88)                     |
| W.V. .0 (0/1)<br>B.V. .--- <sup>a</sup> | W.V. .07 (1/14)<br>B.V. .0 (0/9)  | W.V. .14 (1/7)<br>B.V. .0 (0/4)     | W.V. .86 (6/7)<br>B.V. 1.0 (1/1)   |
| 0 pts.                                  | 7 pts.                            | 14 pts.                             | -14 pts.                           |
| 0,0,0<br>(.00)                          | 1,1,0<br>(.00)                    | 1,1,1<br>(.25)                      | 1,1,2<br>(.81)                     |
| W.V. .0 (0/3)<br>B.V. .0 (0/5)          | W.V. .0 (0/6)<br>B.V. .0 (0/4)    | W.V. .34 (14/41)<br>B.V. .05 (1/19) | W.V. .83 (53/64)<br>B.V. .67 (6/9) |
| 0 pts.                                  | 0 pts.                            | 29 pts.                             | 16 pts.                            |
| 0,0,1<br>(.00)                          | 0,1,1<br>(.02)                    | 0,1,2<br>(.29)                      | 2,0,2<br>(.56)                     |
| W.V. .0 (0/30)<br>B.V. .0 (0/25)        | W.V. .03 (1/32)<br>B.V. .0 (0/16) | W.V. .38 (6/16)<br>B.V. .0 (0/5)    | W.V. .55 (6/11)<br>B.V. .57 (4/7)  |
| 0 pts.                                  | 3 pts.                            | 38 pts.                             | -2 pts.                            |
| 0,0,0<br>(.00)                          | 1,0,1<br>(.01)                    | 1,0,2<br>(.26)                      |                                    |
| W.V. .0 (0/1)<br>B.V. .--- <sup>a</sup> | W.V. .01 (1/89)<br>B.V. .0 (0/95) | W.V. .31 (8/26)<br>B.V. .19 (3/16)  |                                    |
| 0 pts.                                  | 1 pt.                             | 12 pts.                             |                                    |

<sup>1</sup> Distribution of cases is the same as is figure 1. Each cell indicates the coding on the three Barnett dimensions, the overall death sentence rate (in parenthesis), the death sentencing rates for white victim (W.V.) and black victim (B.V.) cases respectively and the arithmetic difference between the two rates in percentage points.  
<sup>a</sup> --- indicates no cases.

cases in each of the eighteen cells defined by those three dimensions. Figure 2 also indicates for each cell the overall death sentencing rate and the disparity between the death sentencing rates for black and white victim cases. Figure 2 shows a race-of-victim effect in four of the six cells in which the death sentencing rate exceeds .10, with quite substantial effects in the mid-range of cases in which the overall death sentencing rate for all cases ranged from .10 to .80. The overall race-of-victim effect after controlling for the three Barnett dimensions is eight percentage points significant at the .006 level.<sup>41</sup>

Table 3 shows the race-of-victim effects estimated with the index we designed specifically to measure race effects in the system.

TABLE 3

Race of Victim Effects in Death Sentencing Rates, Controlling for Case Culpability: Post-*Furman* Georgia

| A<br>Level of Culpability<br>from least (1) to<br>most (6) <sup>a</sup> | B<br>All<br>Cases | C<br><i>Death Sentence Rate</i> |              | D<br>Black<br>Victim<br>Cases | E<br>Difference <sup>1</sup><br>(Col. C-<br>Col. D) | F<br>Ratio <sup>1</sup><br>(Col. C/<br>Col. D) |
|---|-------------------|---------------------------------|--------------|-------------------------------|---|--|
|   |                   | White<br>Victim<br>Cases        |              |                               |   |  |
| 1   | .02               | .02 (5/207)                     | .005 (1/192) | 2                             | 4.6   |  |
| 2   | .14               | .19 (8/43)                      | .05 (1/22)   | 14                            | 4.1   |  |
| 3   | .38               | .44 (15/34)                     | .23 (3/13)   | 21                            | 1.9   |  |
| 4   | .65               | .71 (20/28)                     | .40 (2/5)    | 31                            | 1.8   |  |
| 5   | .85               | 1.0 (18/18)                     | .56 (5/9)    | 44                            | 1.8   |  |
| 6   | 1.0               | 1.0 (31/31)                     | 1.0 (4/4)    | 0                             | 1   |  |
| All Cases   | .19               | .27 (97/361)                    | .07 (16/245) | 20                            | 3.9   |  |

<sup>a</sup> See Table 1 for a list of the variables included in this index.

<sup>1</sup> The overall race of victim disparity after controlling for the levels of case culpability indicated in Column A is 8.5 percentage points, significant at the .001 level (Mantel-Haenszel  $Z = 3.36$ ).

It also shows that the race-of-victim disparities appear primarily in the mid-range of cases, in which the overall death sentencing rates are less than .80. This analysis shows an overall race-of-victim effect of 8.5 percentage points significant at the .001 level<sup>42</sup> after controlling for the

<sup>41</sup> (Mantel-Haenszel  $Z = 2.74$ ). See G. SNEDECOR & W. COCHRAN, STATISTICAL METHODS (6th ed. 1967). The race of victim disparity in the mid-range is 21 percentage points significant at the .02 level. (Mantel-Haenszel  $Z = 2.28$ ). The overall race of victim disparity after further adjustment for the Barnett prior record variable is 6.5 points significant at the .01 level (Mantel-Haenszel  $Z = 2.56$ ).

<sup>42</sup> (Mantel-Haenszel  $Z = 3.3$ ). The index underlying Table 3 is model OVER-

culpability level of the cases.<sup>43</sup>

Professor Barnett raises the possibility that the concentration of race-of-victim effects among the moderately aggravated cases may be a statistical artifact or chance result rather than evidence of true race-of-victim discrimination. He further suggests, however, that such an interpretation is implausible. We agree with him for four reasons. First, the overall disparity in Figure 2 is statistically significant at the .006 level, suggesting that chance is an unlikely explanation. The overall disparity in Table 3 is also statistically significant.<sup>44</sup>

Second, the appearance of racial discrimination in the mid-aggravation range of cases is consistent with prior research and with the so-called "liberation hypothesis."<sup>45</sup> This hypothesis suggests that the effect of racial or other illegitimate factors will appear most frequently in moderately aggravated cases the facts of which do not clearly dictate a particular sentence. In other words, the moderately aggravated nature of such cases liberates the decisionmaker to respond to impermissible considerations. Certainly the concentration of the observed race-of-victim effects in the mid-range of cases from our Georgia data set is consistent with this hypothesis.

Third, our separate analyses of prosecutorial decisions to seek and of jury decisions to impose the death penalty each show a race-of-victim effect concentrated in the mid-range of cases.<sup>46</sup>

---

ALLB, *supra* note 23. The race of victim effect in the mid range rows 2, 3, & 4 is 22 percentage points, significant at the .02 level (Mantel-Haenszel  $Z = 2.37$ ).

<sup>43</sup> Overall, there are 33 more death sentences in white victim cases than one would expect if death sentences were imposed in white victim cases at the same rate they were in black victim cases.

<sup>44</sup> See *supra* Table 3 note 1.

<sup>45</sup> H. KALVEN, JR. & H. ZEISEL, *THE AMERICAN JURY* 164-67 (1966).

<sup>46</sup> The following tabulation shows the intensification of race-of-victim effects in the mid-range in the jury and prosecutorial decisions.

| A<br>Culpability<br>Level<br>(1 low to<br>6 high) <sup>1</sup> | Prosecutorial Decision to Request a Death Sentence |                              |                              |  | Jury Decision to Impose a Death Sentence |                      |                              |                              |  |                                   |
|--|--|------------------------------|------------------------------|--|--|----------------------|------------------------------|------------------------------|--|-----------------------------------|
|  | B<br>Overall<br>Rate                               | C<br>White<br>Victim<br>Rate | D<br>Black<br>Victim<br>Rate | E<br>Difference<br>(Col. C -<br>Col. D) <sup>2</sup> | F<br>Ratio<br>(Col. C/<br>Col. D)        | G<br>Overall<br>Rate | H<br>White<br>Victim<br>Rate | I<br>Black<br>Victim<br>Rate | J<br>Difference<br>(Col. H -<br>Col. I) <sup>3</sup> | K<br>Ratio<br>(Col. H/<br>Col. I) |
| 1  | .03  | .05(6/124)                   | .01(2/144)                   | 4  | 5  | .05                  | .04(2/49)                    | .07(1/14)                    | -3   | .57                               |
| 2  | .14  | .21(10/48)                   | .05(2/39)                    | 16   | 4.2                                      | .33                  | .37(11/30)                   | .0(0/3)                      | 37   | —                                 |
| 3  | .49  | .56(40/72)                   | .34(11/32)                   | 22   | 1.6                                      | .71                  | .74(17/23)                   | .60(3/5)                     | 14   | 1.2                               |
| 4  | .88  | .95(38/40)                   | .71(12/17)                   | 24   | 1.3                                      | .96                  | 1.0(19/19)                   | .83(5/6)                     | 17   | 1.2                               |
| 5  | .91  | .94(31/33)                   | .80(8/10)                    | 14   | 1.2                                      | .89                  | 1.0(13/13)                   | .60(3/5)                     | 40   | 1.7                               |
| 6  | 1.0  | 1.0(32/32)                   | 1.0(2/2)                     | 0  | 1  | 1.0                  | 1.0(35/35)                   | 1.0(4/4)                     | 0  | 1                                 |

1. The two indices underlying the two scales are analogous to OVERALLB, see *supra* note 23. The prosecutorial analysis does not include cases in which a second or third penalty trial was held after a reversal on appeal; the jury analysis includes those cases.
2. The overall average disparity is 11.3 percentage points significant at the .001 level (Mantel-Haenszel Z = 3.83).
3. The overall average disparity is 14.5 percentage points significant at the .055 level (Mantel-Haenszel Z = 1.92).

The presence of this phenomenon in two distinct stages of Georgia's capital sentencing system also makes chance an implausible explanation. Finally, we note that in a parallel study which tracks Georgia's death eligible cases from the point of indictment to sentencing, we find race of victim disparities, both overall and in the mid-range, that are comparable to the disparities reported in this Article.<sup>47</sup>

## 2. Race of Defendant Effects

Professor Barnett's analysis of our Georgia data showed no evidence of discrimination against black defendants because of their race, and, when we examined that data on a statewide basis, we made the same finding.<sup>48</sup> These findings stand in stark contrast to the strong anti-black defendant bias we observed in our pre-*Furman* data, especially in cases from urban areas.<sup>49</sup> However, when we disaggregated the post-*Furman* data to examine separately urban and rural areas, the picture changed. We found that in rural areas black defendants with white victims still received somewhat more severe treatment, but the differences were not statistically significant.<sup>50</sup>

---

In a parallel analyses of prosecutorial and jury decisions which control for the three Barnett dimensions, the race of victim disparities were 15 ( $p < .001$ ) and 14 ( $p = .11$ ) percentage points, respectively.

<sup>47</sup> The results of both our studies are referred to in *McCleskey v. Kemp*, 753 F.2d 877 (11th Cir. 1985). The results reported in this Article are from what the court refers to as the *Procedural Reform Study*; our parallel study is referred to by the court as the *Charging and Sentencing Study*.

<sup>48</sup> The race of defendant coefficient in our principal model (OVERALLA) was  $-.57$  with a standard error of  $.50$  significant at the  $.25$  level.

<sup>49</sup> The analysis of pre-*Furman* data showed a strong statewide race-of-defendant effect. Statewide, the odds of receiving a death sentence were three times higher for black than for white defendants after adjusting for 12 statistically significant ( $p < .10$ ) legitimate background factors.

<sup>50</sup> The death sentencing rates controlling for defendant/victim racial combination culpability level, and the place of the murder were as follows:

| Culpability Level (from 1, least to 6, most) | Urban Georgia          |                        |                        |                        | Rural Georgia          |                        |                        |                        |
|--|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|  | Black                  |                        | White                  |                        | Black                  |                        | White                  |                        |
|  | Defendant/White Victim | Defendant/Black Victim | Defendant/White Victim | Defendant/Black Victim | Defendant/White Victim | Defendant/Black Victim | Defendant/White Victim | Defendant/Black Victim |
| 1  | .0 (0/33)              | .0 (0/95)              | .04 (2/48)             | .0 (0/4)               | .07 (2/28)             | .0 (0/97)              | .0 (0/8)               | .0 (0/8)               |
| 2  | .11 (1/9)              | .08 (1/12)             | .38 (5/13)             | .0 (0/1)               | .0 (0/8)               | .0 (0/14)              | .13 (1/8)              | .— <sup>a</sup>        |
| 3  | .20 (1/5)              | .18 (2/11)             | .71 (5/7)              | .0 (0/2)               | .38 (3/8)              | .27 (3/11)             | .0 (0/1)               | .—                     |
| 4  | .50 (2/4)              | 1.0 (1/1)              | 1.0 (3/3)              | .—                     | 1.0 (10/10)            | .81 (9/11)             | .50 (2/4)              | .—                     |
| 5  | 1.0 (3/3)              | .0 (0/2)               | 1.0 (4/4)              | .—                     | 1.0 (4/4)              | 1.0 (8/8)              | .83 (5/6)              | 1.0 (1/1)              |
| 6  | 1.0 (3/3)              | 1.0 (3/3)              | 1.0 (4/4)              | 1.0 (3/3)              | 1.0 (16/16)            | 1.0 (9/9)              | .—                     | .—                     |

<sup>1</sup> The scale is based on index OVERALLB, see *supra* note 23, which explains the slight difference in the distributions of cases when compared with the distribution in Table 1.

<sup>a</sup> .— means no cases.

By contrast, in urban areas, prosecutors and juries apparently continue to discriminate on the basis of the defendant's race, but this post-*Furman* urban discrimination was against *white* defendants.<sup>51</sup> These somewhat surprising findings indicate why we observed no race-of-defendant effects on a statewide basis — the discrimination against black defendants in rural areas and the discrimination against white defendants in urban areas cancelled each other out.

### CONCLUSION

Professor Barnett's classification system represents an imaginative and useful application of a conceptually sound method for rating and ranking cases. Certainly, his approach deserves consideration by state supreme courts interested in using an empirical and systematic, yet non-computerized, method of monitoring all the sentences imposed in capital cases. Our analyses of the Georgia data have convinced us that, unless a reviewing court is able to compare the relative culpability of all death eligible defendants, despite factual differences in their cases, its ability to identify excessive or discriminatory death sentences will be very much reduced.

Our analysis of the Georgia death sentencing data also suggests, however, the limits of approaches such as Professor Barnett's, which rely on intuition and common sense. It seems clear that formal, multivariate, statistical analysis can provide an additional insight into the factors that influence prosecutors and juries in capital cases, something with which reviewing courts are properly concerned. Nor do we believe the risks involved in using multiple regression analysis to measure relative case culpability are as great as Professor Barnett fears, either as a general rule or in connection with our specific study of the Georgia data. In particular, we believe that one can adequately compensate for any potential deficiencies in the use of regression based systems to measure relative case culpability by comparing the culpability rankings they produce with an intuitive assessment of relative deathworthiness of the cases involved.

However, in spite of the differences between Professor Barnett's

---

<sup>51</sup> The coefficient for an interaction term between urban place and race of defendant, added to Model OVERALLA in *supra* note 23 was -3.72, significant at the .0006 level. The coefficient obtained when we added the interaction term to a regression analysis employing variables for Professor Barnett's three dimensions and his prior record variable was -.82 significant at the .25 level. We deemed the following counties urban: Cobb, DeKalb, Fulton, Richmond, Chatam, Bibb, Muscogee, Dougherty, Clarke, Floyd, and Lowndes.



methodology and that used in our study, our respective findings are similar. Both studies suggest that over one-half of the death sentences imposed in the cases in the data set were not arbitrary or excessive. Both studies also suggest that a good proportion of the remaining death sentence cases may be comparatively excessive. Both studies found a race-of-victim effect among the mid-range of cases in which the combination of aggravating and mitigating factors did not dictate either a life or death sentence. Our analysis also shows a weak race-of-defendant effect for rural areas, which puts black defendants with white victims at a slight disadvantage. In urban areas, by contrast, our analysis reveals a race-of-defendant effect that works to the disadvantage of white defendants.

Both Professor Barnett's analysis and our study suggest that Georgia could avoid any risk of imposing excessive or discriminatory death sentences by restricting the death penalty to only the most aggravated cases, that is, those for which the death-sentencing rate in "similar" cases exceeds .85. Perhaps most importantly, the consistency of our respective findings suggests that reliable methods for comparing cases do exist and that state supreme courts can use those methods to monitor systematically dispositions in capital cases and thereby to eradicate both arbitrary and discriminatory death sentences.

