

UCLA

UCLA Public Law & Legal Theory Series

Title

Disciplining Death: Assessing and Ameliorating Arbitrariness in Capital Charging

Permalink

<https://escholarship.org/uc/item/322464fv>

Journal

Arizona State Law Journal, 49(1)

Author

Thaxton, Sherod

Publication Date

2017



UCLA | SCHOOL OF LAW

PUBLIC LAW & LEGAL THEORY RESEARCH PAPER SERIES
WORKING PAPER NO. 16-08

LAW & ECONOMICS RESEARCH PAPER SERIES
WORKING PAPER NO. 16-03

DISCIPLINING DEATH: ASSESSING AND AMELIORATING ARBITRARINESS IN CAPITAL CHARGING

Sherod Thaxton

2016

In Volume 49 of the *Arizona State Law Journal* (forthcoming 2017)

DISCIPLINING DEATH: ASSESSING AND AMELIORATING ARBITRARINESS IN CAPITAL CHARGING

Sherod Thaxton¹

Justice Stephen Breyer recently made international headlines when he emphasized that reforms to the capital punishment process have apparently failed to ameliorate the rampant arbitrariness, capriciousness, and bias that led the U.S. Supreme Court to temporarily invalidate the death penalty over forty years ago. According to the Justice, the primary cause of this failure has been the Court's backpedaling on the very substantive and procedural protections it initially articulated as necessary for the constitutional administration of the death penalty. The Court's capital punishment jurisprudence initially underscored the importance of social scientific evidence in assessing the fairness of capital punishment systems, but now the Court routinely minimizes, or outright ignores, social science evidence on the operation of the death penalty. This has led to the growing disjunction between the Court's rhetoric and the reality of capital punishment. Justice Breyer underscored the Court's responsibility in holding death penalty systems accountable and called for full briefing on the basic question of the social realities of the administration of capital punishment.

Meaningful death penalty reform, if possible, requires a more prominent role for social science in death penalty decision-making. In this Article, I develop a doctrinally anchored statistical model that carefully disentangles and evaluates questions of arbitrariness, bias, and disproportionality in capital charging. I begin by discussing the Court's inconsistent efforts to rationalize

¹ Assistant Professor of Law, UCLA School of Law. I would like to thank Anthony Alfieri, Stuart Banner, Bennet Capers, Devon Carbado, Beth Colgan, Ingrid Eagly, Laura Gomez, Cheryl Harris, Daniel Ho, Jill Horwitz, Irene Joe, Russell Korobkin, Sung Hui Kim, Máximo Langer, Douglas Lichtman, Mathew McCubbins, Hiroshi Motomura, Jason Oh, Richard Re, Richard Sander, Joanna Schwartz, Seanna Shiffrin, Alex Wang, Adam Winkler, Noah Zatz, and the participants at Duke Law School's Culp Colloquium, UCLA's Criminal Justice Faculty Workshop, and UCLA's Faculty Colloquium for their valuable feedback. Naturally, all remaining errors are my own.

and regulate capital punishment systems. I then adopt a framework of statistical inference in an effort to provide greater definitional and analytical clarity. Finally, I describe a set of analytical tools uniquely suited for diagnosing capital charging errors that closely aligns with the Court's conceptualization of unacceptable arbitrariness. I illustrate the usefulness of the model on data involving actual death penalty-eligible defendants from Georgia.

My analysis reveals that death penalty charging practices are highly inconsistent, irrational, and disproportionate, both within and across jurisdictions in Georgia. The Article concludes by explaining how the empirical model might be used to improve accuracy and consistency in capital charging systems through empirically informed front-end charging screening.

INTRODUCTION	4
I. CONFRONTING CAPITAL ERROR.....	12
A. FURMAN V. GEORGIA AND ITS PROGENY.....	13
B. CONCEPTUALIZING AND OPERATIONALIZING CAPITAL ERROR	21
1. Reliability	22
2. Validity	23
3. Proportionality	26
C. SUMMARY	28
II. DIAGNOSING CAPITAL ERROR	29
A. MODELING CAPITAL CHARGING	29
B. APPLYING THE MODEL TO GEORGIA.....	36
1. Georgia Capital Charging Data	36
2. Arbitrariness as Unreliability.....	44
3. Arbitrariness as Invalidity.....	51
4. Arbitrariness as Disproportionality	57
5. Summary	61
C. MODEL PREDICTIONS FOR CASE-SPECIFIC OUTCOMES.....	62
D. GENERALIZABILITY AND EXPORTABILITY.....	66
1. Generalizability.....	66
2. Exportability	67
E. SUMMARY	69
III. POLICY IMPLICATIONS	70
A. DESIGNING FRONT-END PROPORTIONALITY REVIEW	73
B. SUMMARY	81
CONCLUSION	84
METHODOLOGICAL APPENDIX.....	85

INTRODUCTION

On June 28, 2015, by a vote of five-to-four, the U.S. Supreme Court rejected Oklahoma death row inmates' challenge to the constitutionality of the state's lethal injection protocol.² Richard Glossip, along with twenty other capital condemned inmates, argued that Oklahoma's method of execution created an unacceptable risk of severe pain, thereby violating the Eighth Amendment's prohibition against cruel and unusual punishment. Writing for the majority, Justice Alito noted that, because the death penalty is constitutionally permitted, some risk of pain is inherent in execution and the petitioners were unable to identify a reasonable alternative that would entail a significantly lower risk of pain.³

In his dissenting opinion, Justice Breyer (joined by Justice Ginsburg) emphasized that the Court should stop "try[ing] to patch up the death penalty's legal wounds one at a time" and consider a more fundamental question: whether the current death penalty system is violative of the U.S. Constitution?⁴ After juxtaposing the Court's capital punishment jurisprudence with the voluminous social science literature on capital charging-and-sentencing practices across the nation, Justice Breyer concluded the Court has developed a large body of procedural regulations to govern the administration of the death penalty while simultaneously doing little to ensure that the panoply of protections that exist on paper are provided, in reality, to capital defendants.⁵ Justices Breyer and Ginsburg are not alone in their assessment of the previous four decades of capital charging-and-sentencing practices. In fact, the persistent obstacles to the fair administration of capital punishment have caused several current and recently retired justices to openly question whether efforts to fix the system should be finally abandoned and the

² Glossip v. Gross, 135 S. Ct. 2726 (2015).

³ *Id.*

⁴ *Id.* (Breyer & Ginsburg, JJ., dissenting).

⁵ *Id.*; see also James S. Liebman, *Slow Dancing with Death: The Supreme Court and Capital Punishment, 1963-2006*, 107 COLUM. L. REV. 1, 6 (2007).

country should move towards complete abolition.⁶

Scholars have also described the death penalty system as being in a state of “perpetual malfunction.”⁷ There is strong evidence that death sentences are being imposed just as arbitrarily, discriminatorily, and excessively as they were prior to the Court expressly ruling that the death penalty must be administered fairly and evenhandedly, or not at all.⁸ Current capital charging-and-sentencing practices have resulted in a system marred by inexcusably high rates of reversals and retrials of capital verdicts, as well as extremely lengthy delays in executions.⁹ These problems have significantly undermined the credibility of the death penalty “whose chief function appears to be making mistakes, then taking years in a sometimes vain effort to correct them” rather than deterring potential killers and punishing those murders most deserving of the ultimate sanction.¹⁰

⁶ In recent years, Justices Anthony Kennedy, Sandra Day O'Connor, and John Paul Stevens have made statements either expressly condemning the practice of capital punishment, or raising serious concerns as to its fair administration. See Brian Bakst, *O'Connor Questions Death Penalty*, ASSOCIATED PRESS, Jul. 2, 2001; Brent E. Newton, *Justice Kennedy, the Purposes of Capital Punishment, and the Future of Lackey Claims*, 62 BUFF. L. REV. 979 (2014); James S. Liebman & Lawrence C. Marshall, *Less Is Better: Justice Stevens and the Narrowed Death Penalty*, 74 FORDHAM L. REV. 1607 (2005–2006). Justices Stevens, Lewis Powell, and Harry Blackmun all voted to uphold the constitutionality of capital punishment in *Gregg*, but publically criticized the death penalty after retiring from the Court. Andrew Cohen, *Why Don't Supreme Court Justices Ever Change Their Minds in Favor of the Death Penalty?*, THE ATLANTIC, Dec. 10, 2013.

⁷ James S. Liebman, *Opting for Real Death Penalty Reform*, 63 OHIO ST. L. 315, 321 (2002); see also Carol Steiker & Jordan Steiker, *Lessons for Law Reform from the American Experiment with Capital Punishment*, 87 S. CAL. L. REV. 733 (2014).

⁸ See, *infra*, Part I.

⁹ American Law Institute, *Report of the Council to the Membership of the American Law Institute On the Matter of the Death Penalty* (2009) 1 (withdrawing its endorsement of the death penalty framework it had developed and promoted for over four decades “in light of the current intractable institutional and structural obstacles to ensuring a minimally adequate system for administering capital punishment”).

¹⁰ Liebman, *supra* note 7, at 320. See also, *infra*, Part I.

Some scholars have argued that, ultimately, a nationwide prohibition against the death penalty may be the only reasonable response to the chronic problems that have plagued the practice.¹¹ But death penalty abolitionists are unlikely to “unplug the machine of death” anytime soon given its continued popularity among legislatures and the general public.¹² Unless the Supreme Court unexpectedly reverses direction and decides that the death penalty is unconstitutional, *per se*, both death penalty abolitionists and retentionists must continue to focus their attention on whether the death penalty is capable of being administered in a manner that comports with the legal standards announced by the Court.¹³ “Only if it is not, is abolition a constitutional imperative” under the Court’s current death penalty jurisprudence.¹⁴

The legitimacy of the death penalty system rests, primarily, on the ability of applicable policies and procedures to reduce foreseeable errors that undermine fairness.¹⁵ These constitutionally impermissible errors come in the form of

¹¹ See, e.g., Liebman, *supra* note 5.

¹² Although public support for the death penalty is at its lowest point in four decades, the majority of states still authorize the death penalty and the majority of Americans continue to endorse the practice. Frank Newport, *In U.S., Support for Death Penalty Falls to 39-Year Low*, GALLUP POLL NEWS SERVICE, Oct. 13, 2011; Rebecca Stewart, *Number Who Prefer Death Penalty on Decline*, CNN NEWS POLLS, Oct. 12, 2011. But see Kenneth E. Shirley & Andrew Gelman, *Hierarchical Models for Estimating State and Demographic Trends in US Death Penalty Public Opinion*, 178 J. ROYAL STAT. SOC’Y 1 (Series A, 2015) (noting significant racial/ethnic, gender, and geographic differences in support for capital punishment).

¹³ Justice Ginsburg was the only other member of the Court to join Justice Breyer’s dissent in *Glossip*. Justice Sotomayor also authored a dissenting opinion, joined by Justices Breyer, Ginsburg, and Kagan, but her opinion was not as far-reaching in its condemnation of the capital punishment; rather she expressed her view that Oklahoma’s execution protocol was unconstitutional.

¹⁴ Liebman, *supra* note 5, at 5; See also *Glossip v. Gross*, 135 S. Ct. 2726 (2015) (Breyer, J., dissenting) (stating his belief that it is highly likely that the administration of capital punishment violates the Eighth Amendment).

¹⁵ BRIAN FORST, *ERRORS OF JUSTICE: NATURE, SOURCES, AND REMEDIES* 7 (2004); HANNAH QUIRK ET AL., *REGULATION AND CRIMINAL JUSTICE: INNOVATIONS IN POLICY AND RESEARCH* 35 (2010).

inconsistent, irrational, discriminatory, and excessive charging-and-sentencing outcomes. Developing a system for the management of these errors requires methods that “open an inquiry into the effects of changes in standards, policies, and practices on the incidence of errors.”¹⁶ In this Article, I develop a doctrinally anchored empirical model that disentangles and evaluates questions that are central to the constitutionally permissible administration of capital punishment: how arbitrary, biased, and disproportionate is capital charging? Properly identifying, quantifying, and ultimately discouraging the inappropriate use of the death penalty at the charging stage is likely the most effective and efficient way to reduce the overall prevalence of constitutional error in capital charging-and-sentencing systems.¹⁷ My model is exportable, thus with minor modifications, it can be used to directly diagnose the level of arbitrariness at which a statute operates in any jurisdiction. Additionally, the proposed model evaluates the potential excessiveness of any individual capital charging decision. In concrete terms, the model is capable of predicting the likelihood that a defendant would face the death penalty based on prior capital charging decisions in the jurisdiction. This statistic provides meaningful and verifiable information to decision-makers when evaluating the appropriateness of pursuing the ultimate sanction against a particular defendant. I demonstrate the usefulness of the model by analyzing eight years of capital charging decisions from Georgia.

The article is organized into three parts. Part I discusses the U.S. Supreme Court’s inconsistent efforts to rationalize and regulate capital punishment systems. The Court’s vague and, often internally incoherent, jurisprudence has undermined efforts to identify constitutional errors and develop systems capable, at least in theory, of eliminating them. On the one hand, the Court has repeatedly emphasized that “death is different” and capital punishment systems require heightened consistency and accuracy, even within the narrow class of death-eligible defendants. On the other hand, the Court has failed to provide sufficiently precise or consistent workable

¹⁶ FORST, *supra* note 15, at 31–32.

¹⁷ *See, infra*, Parts II & III.

definitions of reliability, validity, and proportionality—all central concepts articulated in its doctrines. Legally meaningful definitions are not only crucial to properly evaluate capital punishment systems' conformity with the heightened standards announced by the Court, but also indispensable when considering adequate modifications to existing procedures and remedies for undesirable outcomes. In an attempt to provide greater definitional and analytical clarity, I adopt a framework of statistical inference that is particularly well-suited for identifying and quantifying the types of constitutional error that animated the Court's modern capital punishment jurisprudence.¹⁸

Building upon the analytical framework described in the previous section, Part II develops a doctrinally anchored empirical model capable of assessing the level of arbitrariness—i.e., unreliability, invalidity, and disproportionality—in capital charging that exists both within and across jurisdictions. The model has two key features that are essential to properly studying capital punishment practices. First, by focusing on a single decision point at the very outset of the capital punishment process—where prosecutorial discretion is nearly unfettered—one can obtain a much clearer picture of the dynamics driving variability in capital charging practices observed under a single capital statute.¹⁹ Second, consistent with the decentralized and

¹⁸ Other scholars have also advocated for the adoption of a framework for statistical inference for such a task. *See, e.g.*, Alberto Alesina & Eliana La Ferrara, *A Test of Racial Bias in Capital Sentencing*, 104 AM. ECON. REV. 3397 (2014) (adopting a statistical framework for errors of inference to determine racial bias in sentencing by examining appellate review of capital cases); QUIRK ET AL., *supra* note 15 (suggesting that criminal justice evaluation should be informed by principles of statistical design and inference); David C. Baldus et al., *Reflections on the "Inevitability" of Racial Discrimination in Capital Sentencing and the "Impossibility" of Its Prevention*, 51 WASH. & LEE L. REV. 359 (1994) (advocating a statistical-inference based approach to determining racial discrimination in death penalty sentencing).

¹⁹ *See* McCleskey v. Kemp, 481 U.S. 279 (1987) (statistical evidence of racial disparities derived from multiple decision-makers is insufficient to mount a constitutional challenge under the Court's current death penalty jurisprudence); *United States v. Bass*, 536 U.S. 862 (2002) (per curiam) (holding that nationwide statistics of racial disparities in capital charging

county-centric nature of death penalty charging authority, my proposed model mobilizes analytical tools uniquely developed to investigate hierarchically structured (“clustered”) data—that is, individual cases nested in counties or similar sub-state units. The Court has repeatedly emphasized the importance of explicitly examining both intra- and inter-jurisdiction processes when assessing the constitutionality of criminal punishments.²⁰ Identifying and quantifying variation in this manner—i.e., single decision point and contextual influences—more closely aligns with the Court’s conceptualization of unacceptable arbitrariness under its capital punishment jurisprudence. To the best of my knowledge, this Article is the first to apply these tools to capital charging decisions.

Beyond diagnosing an entire system for macro-level constitutional errors of arbitrariness, I describe how the model can be used when making case-by-case capital charging assessments to detect potential micro-level errors. These micro-level errors are excessive charging decisions in cases that are technically eligible for the death penalty. Such errors have been definitively linked to other types of legal error: prosecutors who pursue the death penalty in cases that are not highly aggravated are significantly more likely to have those very same cases overturned on appeal because of serious trial-level.²¹

The model is designed to be exportable, so its usefulness transcends its application in any particular jurisdiction. I illustrate the usefulness of the model on data involving actual death penalty-eligible defendants from Georgia, but there is good reason to believe the patterns I identify are present in other capital jurisdictions. The data reveal several constitutionally problematic features of Georgia’s capital charging practices. First, capital charging practices are highly

decisions are insufficient to support an equal protection violation claim).

²⁰ *Solem v. Helm*, 463 U.S. 277, 290–92 (1983); *Enmund v. Florida*, 458 U.S. 782, 794–96 (1982).

²¹ See, e.g., Andrew Gelman et al., *A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States*, 1 J. EMPIRICAL LEGAL STUD. 209 (2004).

inconsistent within jurisdictions (i.e., “between-case” heterogeneity) even for factually similar cases.²² Second, capital charging practices for alike cases are highly inconsistent across jurisdictions (“between-jurisdiction” heterogeneity) for factually similar cases.²³ Third, the jurisdiction where a case is prosecuted exerts a strong influence on whether a defendant is charged with the death penalty.²⁴ Locales characterized by large deviations in death noticing behavior, relative to the statewide baseline, for similarly situated defendants, may be interpreted as being unjustifiably idiosyncratic given existing constitutional constraints on the capital punishment process. These jurisdictional effects are indicators of “institutional performance” and permit the comparative ranking of the jurisdictions.²⁵

Fourth, case-level characteristics only explain a small percentage of the variation in capital charging decisions both within and across jurisdictions.²⁶ This lack of a strong correspondence between the legally legitimate case characteristics and charging outcomes is a marker of the irrationality of the charging process, which is an important component of arbitrary government action.²⁷ Fifth, the race of defendants and victims is strongly associated with the level of inconsistency and irrationality in capital charging practices.²⁸

²² See *infra* Part II.B.2

²³ *Id.*

²⁴ *Id.*

²⁵ 1 SOPHIA RABE-HESKETH & ANDERS SKRONDAL, MULTILEVEL AND LONGITUDINAL MODELING USING STATA: CONTINUOUS RESPONSES, 50 (3d ed. 2012) (describing contextual effects as “valued added” by the location, all else being equal); Craig Duncan et al., *Context, Composition and Heterogeneity: Using Multilevel Models in Health Research*, 46 SOCIAL SCIENCE & MEDICINE 97, 111 (1998) (discussing the use of predictions of cluster-specific effects to rank institutions).

²⁶ See *infra* Part II.B.3.

²⁷ *Burlington Truck Lines, Inc. v. United States*, 371 U.S. 156, 167–68 (1962) (defining arbitrary government action as the lack of a rational connection between the facts that should govern a decision and the choice being made); *accord* *Motor Vehicle Mfrs. Ass’n, Inc. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

²⁸ See *infra* Part II.B.4. *Furman v. Georgia*, 408 U.S. 238, 365 (1972)

Caucasian-defendant and Caucasian-victim cases are handled more consistently than in non-Caucasian-defendant and non-Caucasian-victim cases. Similarly, with respect to the rationality of charging decisions, case-level characteristics explain a larger percentage of variation in outcomes in Caucasian-defendant and Caucasian-victim cases than in non-Caucasian-defendant and non-Caucasian-victim cases. Lastly, the race of the defendant and the race of the victim appear to have a direct influence on capital charging decisions. Specifically, when looking at Georgia as whole, cases involving Caucasian defendants and victims are more likely to be noticed for the death penalty, all else equal; however, this effect of race is highly variable across jurisdictions. The magnitude of the difference of the effect of race, from the statewide average, is as high as a factor of four for race-of-defendant and three for the race-of-victim.²⁹

Part III sketches some ideas about how my model can be used to improve accuracy and consistency in capital charging systems. The development of a fully-specified policy proposal is beyond the scope of this project,³⁰ but this section does lay a foundation, with rigorous social scientific inquiry at its core,³¹ upon which meaningful death penalty reform can be erected. Specifically, I identify what I believe are some essential features of a feasible and effective reform policy. Generally speaking, any defensible reform must provide meaningful disincentives for poor prosecutorial charge screening and save both state and federal governments much of the expense of error correction occurring at the appellate review stages.

(White, J., concurring) (arguing that arbitrary capital sentencing standards are an “open invitation to discrimination”).

²⁹ See *infra* Part II.B.4.

³⁰ See Sherod Thaxton, *Death, Dollars, and Deference: Rethinking Prosecutorial Accountability in Capital Charging* (Law and Economics Research Paper Series, UCLA School of Law 2016) (describing a system of financial, administrative, and reputational disincentives for poor charge screening in potentially capital cases).

³¹ See CRAIG HANEY, *DEATH BY DESIGN: CAPITAL PUNISHMENT AS A SOCIAL PSYCHOLOGICAL SYSTEM* 214–15 (2005) (advocating for a more prominent role for social science in the Court’s death penalty jurisprudence).

I. CONFRONTING CAPITAL ERROR

Capitally condemned inmates have challenged the legality of the death penalty, as applied, under three provisions of the U.S. Constitution: the Eighth Amendment's Cruel and Unusual Punishment Clause, the Fourteenth Amendment's Due Process Clause, and the Fourteenth Amendment's Equal Protection Clause.³² Both cruel and unusual punishment and due process violation claims have focused on the alleged arbitrary administration of capital punishment; whereas equal protection challenges highlighted the racially discriminatory application of the death penalty.³³ The Cruel and Unusual Punishment Clause also formed the basis of challenges to the excessiveness of the death penalty as applied in particular cases (disproportionality). A routinely divided court accepted some of these claims and rejected others. The consequence of these cases has been the development of an increasingly complex, and often contradictory, capital punishment jurisprudence that has been derived, primarily, from the Court's interpretation of the Cruel and Unusual Punishment Clause.

Part I.A discusses the key aspects of the Court's jurisprudence as it pertains to the constitutionally permissible administration of the death penalty and state legislatures' responses to these rulings. The Court emphasized distributive justice as a key component to the fundamental fairness of capital punishment systems,³⁴ but this commitment has waned over the

³² Due Process Clause and Equal Protection Clause arguments for challenges to the application of the death penalty for cases originating in federal court are governed by the Fifth Amendment's Due Process Clause. *Bolling v. Sharpe*, 347 U.S. 497, 499 (1954) (explaining that the Fifth Amendment's Due Process Clause implicitly incorporates an equal protection guarantee).

³³ *Glossip v. Gross*, 135 S. Ct. 2726 (2015) (Breyer, J., dissenting) (noting that irrelevant factors, such as race, that determine who receives the death penalty is indicative of arbitrariness).

³⁴ Distributive justice is generally defined as the perceived fairness of the allocation of rewards and costs. Procedural justice, on the other hand, focuses on the fairness and transparency of the processes that resolves

years as social facts about the actual operation of capital punishment unequivocally describe a system plagued by unpredictability, irrationality, disproportionality, and multiple forms of discrimination. Procedural justice concerns now dominate the Court's jurisprudence, but the Court has failed to identify any evidence suggesting the procedures developed by legislatures to promote consistency and accuracy in capital charging and sentencing are capable of satisfying the requisite constitutional standards.³⁵

I argue that meaningful death penalty reform can only be possible if courts and legislatures reengage with the social realities capital punishment practices. Central to this reengagement is the articulation of clear and workable standards that permit careful assessments of capital charging-and-sentencing behavior. The Court has consistently refused to provide the necessary guidance, which has both allowed it to insulate itself from the difficult task of enforcing its own doctrines and exacerbated the gulf between the rhetoric and reality of the death penalty. In Part I.B, I turn to the framework of statistical inference for identifying and expressly measuring the types of errors that are of primary concern to the Court: errors of arbitrariness, bias, and disproportionality. The field of statistics has long been concerned with these types of errors, and offers useful guidelines on their detection and measurement. These guidelines provide the necessary foundation for a systematic inquiry into the social facts of capital punishment.

A. *Furman v. Georgia and its Progeny*

In the landmark case, *Furman v. Georgia*,³⁶ the Court held, by

disputes and allocates resources. OXFORD DICTIONARY OF SOCIOLOGY 329 (John Scott & Gordon Marshall eds., 3d ed. 2005).

³⁵ *Glossip v. Gross*, 135 S. Ct. 2726 (Breyer, J., dissenting) (arguing that the administration of capital punishment remains constitutionally infirm).

³⁶ 408 U.S. 238, 310 (1972). The Court agreed to hear four cases out of a pool of nearly two hundred pending capital cases: two non-homicidal rape cases from Georgia (Jackson) and Texas (Branch) and two murder cases from

a vote of five-to-four, that all existing capital punishment statutes were unconstitutional, as applied, because they failed to articulate to decision-makers any rational basis by which to distinguish those limited number of defendants sentenced to death from the thousands of other similarly situated defendants who were not subject to the death penalty. The Court was primarily troubled by three glaring problems with the existing practice of capital punishment: (1) the small number of death sentences handed out relative to potentially capital crimes (infrequency and arbitrariness); (2) the lack of statutory restrictions upon sentencing discretion of judges and jurors (standardlessness); and (3) sentencing disparities based on race/ethnicity and social class (bias/discrimination).³⁷ *Furman* lacked a true holding because all nine Justices wrote separate opinions;³⁸ nonetheless, Justices Brennan,³⁹ Douglas,⁴⁰ Marshall,⁴¹ Stewart,⁴² and White,⁴³ comprising the majority, all expressed serious concern over the irrational and inconsistent imposition of the death penalty. The justices were split over whether racism still infected the death penalty process. Although the justices all acknowledged that racism in

California (*Aikens*) and Georgia (*Furman*). The writ in *Aikens* was dismissed after the California Supreme Court ruled that the state's capital punishment violated the state's Constitution. *People v. Anderson*, 493 P.2d 880 (Cal. 1972).

³⁷ DAVID GARLAND, *PECULIAR INSTITUTION: AMERICA'S DEATH PENALTY IN AN AGE OF ABOLITION* (2010).

³⁸ *Id.* at 225–30. Justices Stewart and White's opinions provided the narrowest ground for agreement, and therefore were deemed controlling. Both Justices focused, primarily, on the arbitrary administration of the death penalty. GARLAND, *supra* note 37; Liebman & Marshall, *supra* note 6 (same).

³⁹ *Furman v. Georgia*, 408 U.S. at 293 (Brennan, J., concurring) (describing the death penalty system as "little more than a lottery system").

⁴⁰ *Id.* at 249 (Douglas, J., concurring) (explaining that equality in the administration of the death penalty is the key consideration for the Court).

⁴¹ *Id.* at 362–63 (Marshall, J., concurring) (noting that convicted murderers are seldom sentenced to death).

⁴² *Id.* at 309 (Stewart, J., concurring) (noting that "death sentences are cruel and unusual in the same way that being struck by lightning is cruel and unusual").

⁴³ *Id.* at 311 (White, J., concurring) (stating that there was no principled way to distinguish defendants who received the death penalty from those who did not).

the administration of the death penalty was evident in the past, they disagreed as to its continuing relevance. Chief Justice Burger and Justice Powell (both dissenting), as well as Justice Douglas (joining the majority), all hinted that an equal protection challenge might prevail if there was compelling evidence of racial bias. The immediate impact of the Court's ruling was the commutation of all defendants who were under the sentence of death and a *de facto* moratorium on executions.⁴⁴

The Court's decision to strike down prevailing death penalty statutes in *Furman* was mixture of procedure and substance.⁴⁵ A year prior to the *Furman* decision, in *McGautha v. California*,⁴⁶ the Court ruled that statutes guiding juror discretion in capital cases were not constitutionally required under the Due Process Clause. Justice Harlan, authoring the majority opinion for the Court, believed that it was impossible to develop a legal formula capable of distinguishing the worst-of-the-worst cases from the vast majority of murder cases that were not capitally prosecuted, so the unstructured sentencing authority of juries neither violated capital defendants' due process nor equal protection rights.⁴⁷ The *Furman* Court expressly stated that its decision did not overrule *McGautha* because that case had only considered due process and (arguably) equal protection objections.⁴⁸ *Furman* deemed the broad and unbridled

⁴⁴ *Furman v. Georgia*, 408 U.S. 238.

⁴⁵ Steiker & Steiker, *supra* note 7, at 762.

⁴⁶ 402 U.S. 183 (1971).

⁴⁷ *Id.* (Harlan, J.) (reasoning that it was both unwise and futile to attempt to determine, a priori, the factors that would warrant a death sentence).

⁴⁸ *Furman v. Georgia*, 408 U.S. at 310 (Stewart, J., concurring). Undoubtedly, due process concerns were at the core of *McGautha* and *Crampton*; nevertheless several Justices also emphasized that due process and equal protection considerations were closely linked: "A vague statute may be applied one way to one person and a different way to another. Aside from the fact that this in itself would constitute a denial of equal protection [...] the reasons underlying different applications to different individuals may in themselves be constitutionally impermissible." *McGautha v. California*, 402 U.S. at 259 fn. 9 (Brennan, J., dissenting) (joined by Douglas and Marshall, JJ.). *But see Furman v. Georgia*, 408 U.S. at 400 (Burger, C.J., dissenting) (arguing that *Furman* could only be reasonably read as overruling *McGautha*).

discretion afforded to capital juries violated the Eighth Amendment's prohibition against cruel and unusual punishment precisely because a permissible process could generate an impermissible result. In other words, the Eighth Amendment's focus was on actual punishments, and not merely the process by which the punishment was decided. The Court declined to offer guidance, however, as to what types of procedures, if any, would produce outcomes satisfying the Eighth Amendment or how those outcomes would be policed. States were left to devise their own statutes that would, ostensibly, pass constitutional muster.

Following *Furman*, many states immediately revamped their capital statutes and resumed sentencing defendants to death.⁴⁹ Uncertain as to what was now constitutionally permissible under the newly revised statutes, defendants sentenced to death challenged their sentences in state and federal courts. The Court agreed to hear a group of five cases that, roughly, represented the range of post-*Furman* capital statutes: three of the cases involved guided-discretion statutes (Florida, Georgia, and Texas), while two others completely eliminated sentencing discretion and required the death sentence for a very narrow class of defendants (Louisiana and North Carolina). The Court ultimately approved the modified guided-discretion death penalty statutes in *Gregg v. Georgia*,⁵⁰ *Jurek v. Texas*,⁵¹ and *Proffitt v. Florida*,⁵² and invalidated the mandatory death penalty statutes in *North Carolina v. Woodson*⁵³ and *Roberts v.*

⁴⁹ DAVID C. BALDUS ET AL., *EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS* (1990); HERBERT H. HAINES, *AGAINST CAPITAL PUNISHMENT: THE ANTI-DEATH PENALTY MOVEMENT IN AMERICA, 1972-1994* (1996). "The new sentencing schemes [adopted by many states after *Furman*] were immediately put to use. Only 42 people were sentenced to death in 1973, but there were 149 death sentences in 1974, probably more than any year since 1942. In 1975, 298 people were sentenced to death—far more than any previous year for which data exist." STUART BANNER, *THE DEATH PENALTY: AN AMERICAN HISTORY* 270 (2002).

⁵⁰ *Gregg v. Georgia*, 428 U.S. 153 (1976).

⁵¹ *Jurek v. Texas*, 428 U.S. 262 (1976).

⁵² *Proffitt v. Florida*, 428 U.S. 242 (1976).

⁵³ *Woodson v. North Carolina*, 428 U.S. 280 (1976).

Louisiana.⁵⁴ In each of these rulings, the Court reiterated that not only must the death penalty be reserved for the worst-of-the-worst offenses, but even among that limited group of persons, the death penalty is only permissible for the most culpable defendants. The Court was convinced that the guided-discretion statutes enacted after *Furman* would result in greater consistency and rationality/accuracy in the administration of the death penalty.

The statutes crafted by legislatures in Florida, Georgia, and Texas imposed different requirements on juries and reviewing courts. Under Florida's scheme, eight aggravating circumstances and seven mitigating circumstances were established to guide jury discretion. Juries were required to weigh aggravating and mitigation evidence and impose a death sentence if the latter did not sufficiently outweigh the former. The jury issued an "advisory" sentence by majority vote, and the judge was authorized to override the jury's sentencing recommendation. All death sentences were automatically reviewed by the Florida Supreme Court.⁵⁵ In Georgia, ten aggravating circumstances were developed, but no specific mitigating circumstances were specified.⁵⁶ Once the jury found at least one aggravating circumstance, it was required to weigh all of the aggravating and mitigating evidence when deciding whether to impose a death or life sentence. The jury was required to be unanimous and its sentencing recommendation was binding on the judge. Defendants sentenced to death received a non-waivable review by the Georgia Supreme Court. Texas's statute included nine "special circumstances" and defendants were subject to the death penalty only if the killing

⁵⁴ *Roberts v. Louisiana*, 428 U.S. 325 (1976). The *Gregg* court expressly recognized that its primary concern in *Furman* was the arbitrary and capricious manner in which defendants were being condemned to death. *Gregg v. Georgia*, 428 U.S. at 195 (noting that the Court's concern about arbitrariness in *Furman* could be adequately addressed by carefully drafted statutes that ensure the sentencing authority is provided relevant information and guidance). Several justices also emphasized the risk of the discriminatory imposition of the death penalty. *Id.* at 206.

⁵⁵ *Proffitt v. Florida*, 428 U.S. 242.

⁵⁶ *Gregg v. Georgia*, 428 U.S. 153.

was unprovoked, deliberate, and the defendant was likely to commit violent acts in the future. The jury's vote for death needed to be unanimous and a death-sentenced defendant received an automatic (and non-waivable) review by the Texas Court of Criminal Appeals.⁵⁷

Despite their differences, the important commonalities of the statutes were (1) a list of factors that would, ostensibly, narrow the reach of the death penalty and identify the most culpable defendants to the sentencing authority⁵⁸ and (2) mandatory appellate review of death sentences by the jurisdiction's highest criminal court that would assess the appropriateness of every death sentence imposed. Whereas Georgia's statute clearly outlined the appellate court's task,⁵⁹ the capital statutes from Florida and Texas did not clearly specify what appellate review would entail.⁶⁰ Nonetheless, the Court noted that, in practice, the reviewing courts in those states were determining whether each defendant's death sentence was

⁵⁷ *Jurek v. Texas*, 428 U.S. 262 (1976).

⁵⁸ By extension, the factors would also constrain the charging authority because certain elements of the crime must be proven to the sentencing authority (e.g., killing of a police officer) in order for the death penalty to be an available sentencing option. The Court would repeatedly (re)emphasize that capital statutes must "genuinely narrow" the death-eligible class to encompass only defendants materially more depraved than the average murderer. *Zant v. Stephens*, 462 U.S. 862 (1983); *Atkins v. Virginia*, 536 U.S. 304 (2002).

⁵⁹ The Georgia Supreme Court was required to decide three things: (1) whether the death sentence was imposed under the influence of passion, prejudice or any other arbitrary factor; (2) whether the evidence supports the aggravating circumstance(s) found by the jury; and (3) whether the death sentence is excessive or disproportionate, relative to the penalty imposed in similar cases, considering the crime and the defendant. *Gregg v. Georgia*, 428 U.S. 153.

⁶⁰ All death sentences were automatically reviewed by the Florida Supreme Court. The statute did not specify what the mandatory review would entail, but the Court underscored that, in practice, inter-case review was conducted by the Florida Supreme Court. *Proffitt v. Florida*, 428 U.S. 242. Texas's revised statute did not require comparative proportionality review, although each death sentence was to be reviewed, at least, on its own merits to ensure that death sentences "will not be 'wantonly' or 'freakishly' imposed." *Jurek v. Texas*, 428 U.S. 262.

arbitrarily imposed, disproportionate, or the product of any impermissible consideration.⁶¹ The Court, once again, signaled that the consideration of both procedure (i.e., narrowing death-eligibility) and results (i.e., appellate review of capital sentences irrespective of the whether the process was followed) were indispensable components of a constitutionally permissible death penalty system. In other words, a “fair” death penalty system must satisfy procedural and distributive justice concerns.

Glaring omissions from both the revised statutes and the Court’s analysis of them, however, were workable definitions of arbitrariness, bias, and disproportionality. The Court and legislatures employed intuitive understandings of these concepts, but they failed to translate these general principles into terms that frontline legal actors—e.g., prosecutors, juries, and appellate courts—could actually put into operation. How were errors of arbitrariness, bias, and disproportionality to be measured in the capital sentencing context? What baselines should be used? What threshold showings must be made before these various claims of constitutional error were cognizable by the Court? These key unresolved questions jeopardized the heightened reliability required under the Court’s “death is different” approach to the Eighth Amendment.

Several of the Court’s subsequent rulings underscored its coarse and inelegant analysis in *Gregg*. In each of these cases, the Court appeared to retreat from its initial positions in *Furman* and *Gregg*, yet the Court neither expressly overruled those initial cases, redefined the core principles and standards articulated in those cases, nor provided meaningful clarity to legislatures and frontline actors in the capital charging-and-sentencing process to assist in devising systems capable of minimizing errors. Complicating matters further, after emphasizing the centrality of narrowly-crafted capital statutes to guide the sentencing authority and minimize arbitrary (and capricious) decision-making, the Court ruled that states were prohibited from limiting the type of mitigation (i.e., exculpatory)

⁶¹ *Gregg v. Georgia*, 428 U.S. 153.

evidence defendants could present at trial.⁶² A few years later, the Court held that the prosecutor could present, and the sentencing authority could consider, aggravating (i.e., inculpatory) evidence that was not specifically enumerated in the capital statute.⁶³ The Court waited seven years to clarify its Eighth Amendment proportionality analysis framework,⁶⁴ although it deemed proportionality determinations by state reviewing courts critical features of the post-*Furman* statutes.⁶⁵ It took the Court an additional four years to announce that statistical evidence of pronounced and persistent racial disparities in death penalty charging and sentencing patterns was inadequate to prove an unacceptable risk of arbitrariness or intentional racial discrimination in the administration of capital punishment in a state.⁶⁶ More than four decades after

⁶² *Lockett v. Ohio*, 438 U.S. 586 (1978).

⁶³ *Zant v. Stephens*, 462 U.S. 862; *Wainwright v. Goode*, 464 U.S. 78 (1983).

⁶⁴ *Solem v. Helm*, 463 U.S. 277, 290–92 (1983) (Powell, J.) (noting that reviewing courts were required to do three things when determining whether a punishment was disproportionate: [1] compare the nature and gravity of the offense and the harshness of the penalty; [2] compare the sentences imposed on other criminals in the same jurisdiction [intra-jurisdictional]; and [3] compare the sentences imposed for commission of the same crime in other jurisdictions [inter-jurisdictional]).

The Court acknowledged the difficulty reviewing courts face in attempting to draw distinctions between similar crimes (and criminals), but it did not believe this was an insurmountable obstacle because reviewing courts were competent to judge the gravity of the offense and the defendant's culpability on a relative scale. *Id.* at 294.

Gregg did announce an analytical framework for assessing whether a punishment was disproportionate, and therefore violative of the Eighth Amendment. *Gregg v. Georgia*, 428 U.S. 153. But the Court's focus was on the constitutionality of capital punishment, *per se*, and not the potential excessiveness of any individual death sentence.

⁶⁵ The Court previously ruled that punishments were "excessive," and therefore prohibited by the Eighth Amendment, if not graduated and proportioned to the offense. *Weems v. United States*, 217 U.S. 349 (1910). This definition failed to provide reviewing courts with any meaningful guidance in *Gregg*.

⁶⁶ *McCleskey v. Kemp*, 481 U.S. 279 (1987). A growing body of evidence documenting persistent gender bias in the administration of the death penalty has emerged since the Court's ruling in *McCleskey*. See, e.g., Steven F. Shatz & Naomi R. Shatz, *Chivalry is Not Dead: Murder, Gender, and the Death*

Furman, the doctrinal haziness remains and there is little indication that the Court will offer any elucidation.⁶⁷ The post-*Furman* Court appears content with looking where it believes the light is brightest and never journeying too far from its perceived areas of competence: statutory interpretation and procedural innovation. However, there may be a strong shadow where there is much light.⁶⁸ The Court's "continued treatment of the social facts and empirical data that document systemic failures in the administration of the death penalty as somehow irrelevant to constitutional decision-making seems increasingly indefensible."⁶⁹ If repairing the death penalty system is possible, it requires the Court's reengagement with the social facts of capital charging-and-sentencing practices. Indispensable to this undertaking are clear and consistent rules and standards governing the administration of capital punishment from both a process and outcome perspective. To date, the Court has refused to perform these essential functions.

B. *Conceptualizing and Operationalizing Capital Error*

The Court's consistent refusal to provide clear standards and guidance to lower courts, lawyers, and legislatures for assessing the forms of constitutional error described in *Furman* and *Gregg* has allowed it to avoid upholding the very principles and rules it established.⁷⁰ By sidestepping precise definitions of constitutional error, and by extension, the evidentiary thresholds that parties must satisfy to make a colorable claim to the Court, the justices were insulating themselves from the

Penalty, 27 BERKELEY J. GENDER L. & JUST. 64 (2012) (capital cases involving female murder defendants are treated more leniently and cases involving female victims are treated more harshly, all else being equal).

⁶⁷ In Part I.B, I provide improved clarity of the Court's doctrine utilizing the framework of statistical inference.

⁶⁸ Johann Wolfgang von Goethe, *Götz von Berlichingen*, Act I (1773).

⁶⁹ HANEY, *supra* note 31, at 216.

⁷⁰ See, e.g., Liebman, *supra* note 5 (describing the Court's avoidance of policing the standards it developed for the regulation of capital punishment systems).

“excruciatingly difficult responsibility for deciding who the State may and may not constitutionally kill.”⁷¹ The promise of *Furman* (and arguably *Gregg*) cannot be fulfilled without the Court, intelligibly, describing how systems may satisfy or fail the constitutional standards it developed. In the absence of these standards from the Court, scholars must look elsewhere for guidance. Even if the Court is skeptical of, or unreceptive to, analysts resorting to standards developed outside of its jurisprudence to identify and quantify the types of constitutional error announced in *Furman* and its progeny, the use of articulable and defensible standards may gain an audience and traction outside of the courtroom. This, in turn, could potentially exert pressure on the Court to adopt these general standards, or more precisely define its own. The widely accepted framework of statistical inference can provide the depoliticized clarity that is lacking from the Court’s current death penalty jurisprudence.⁷² “Statistics is the art of making numerical conjectures about puzzling questions.”⁷³ Statistical inference, among other things, provides a useful framework for identifying and quantifying errors produced by a system. The Court’s concern over errors of arbitrariness directly implicates the reliability, validity, and proportionality of capital decision-making. These three components of arbitrariness are discussed *in seriatim* below.

1. Reliability

The reliability of a system stems from the absence of random error, and is related to the consistency or stability of outcomes across raters (e.g., prosecutors).⁷⁴ Thus, when a repeated process gives highly similar results, the process is said to be

⁷¹ *Id.* at 5.

⁷² Oliver Wendell Holmes famously wrote, “For the rational study of the law the blackletter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.” Oliver Wendell Holmes, *The Path of Law*, 10 HARV. L. REV. 457, 469 (1897).

⁷³ DAVID A. FREEDMAN ET AL., STATISTICS, at xiii (2007).

⁷⁴ GARY KING ET AL., DESIGNING SOCIAL INQUIRY: SCIENTIFIC INFERENCE IN QUALITATIVE RESEARCH 25 (1994).

reliable.⁷⁵ While random errors are inevitable and some may even be socially optimal given the costs of reducing such error relative to the overall benefit from the reduction, a system purporting to provide heightened reliability—as required by the Court’s modern death penalty jurisprudence—must minimize the inconsistency of the outcomes that system generates. Random errors can often be reduced through systems that structure the exercise of discretion of criminal justice actors.⁷⁶

It may difficult, if not impossible, to specify *a priori* how much inconsistency is allowable before a system should be deemed unconstitutionally arbitrary; nevertheless, a highly inconsistent system can hardly be characterized as a “fair and evenhanded” one.⁷⁷ Relatedly, if the inconsistency of the behavior of legal actors is strongly associated with legally illegitimate factors, such as race/ethnicity and gender, then there should be heightened concern about the risk of arbitrariness. The guided-discretion statutes approved by the Court in *Gregg* were designed to rationalize the capital punishment process. Such rationalization required the significant reduction of the arbitrary application of the death penalty. Statutes that do not meaningfully minimize erratic charging-and-sentencing outcomes have failed to sufficiently constrain frontline legal actors—namely prosecutors, judges, and jurors—to satisfy *Furman*.

2. Validity

The validity of a system refers to its truthfulness or accuracy.⁷⁸ That is, the system actually produces what it is

⁷⁵ FORST, *supra* note 15, at 33; DICTIONARY OF STATISTICS & METHODOLOGY: A NONTECHNICAL GUIDE FOR THE SOCIAL SCIENCES 274 (W. Paul Vogt ed., 2005).

⁷⁶ FORST, *supra* note 15.

⁷⁷ *Furman v. Georgia*, 408 U.S. 238, 312 (1972) (the death penalty must be meted out in a consistent, fair, and even-handed manner).

⁷⁸ There are many sub-species of “validity.” DICTIONARY OF STATISTICS & METHODOLOGY, *supra* note 75. In the present context, validity refers to “internal validity.” FORST, *supra* note 15, at 33. See also THOMAS D. COOK &

purported and designed to produce.⁷⁹ Validity stems from both the absence of systematic error and justifiability of the interpretations about the data.⁸⁰ A system typically generates invalid results when it is incorrectly calibrated or actors implementing the system use different procedures and considerations.⁸¹ Assessments of culpability should be aligned with the standards set forth in the applicable capital statutes.⁸² Validity, then, refers to the degree to which theory and evidence support the interpretation of legal decision-makers.⁸³ The validity of an assessment is also closely related to its fairness. In the educational testing literature, assessments are deemed fair when they assess what is “taught.”⁸⁴ In the capital charging context, the death penalty statute provides prosecutors the standards/metrics upon which culpability assessments must be made in order to comport with the constitutional requirements announced in *Furman* and its progeny. So, in concrete terms, if the factors specified in a capital statute fail to predict actual outcomes within an acceptable range, then there is compelling evidence that the system permits an impermissible degree of arbitrariness and is, therefore, constitutionally dubious. Of course, the definition of an “acceptable range” will be subject to debate, but relative assessments can be made by gauging the predictability of similarly structured systems with aligned objectives. Low explanatory power of statutorily relevant factors provides evidence of invalid charging decisions.⁸⁵

DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD SETTINGS* (1979) (defining and identifying threats to internal validity).

⁷⁹ KING ET AL., *supra* note 74, at 25.

⁸⁰ The lack of reliability qualifies as a threat to the validity of a system, whereas the absence of validity is not necessarily a threat to the reliability (i.e., consistency) of a system. FORST, *supra* note 15, at 17; *see also* W. Steve Lang & Judy R. Wilkerson, Accuracy vs. Validity, Consistency vs. Reliability, and Fairness vs. Absence of Bias, presented at Annual Meeting of the American Association of Colleges for Teacher Education 7–8 (New Orleans, La. Feb. 2008) (“[V]alidity means that assessors are making justifiable interpretations about their data and good decisions.”).

⁸¹ FORST, *supra* note 15, at 33–34.

⁸² Lang & Wilkerson, *supra* note 80.

⁸³ *Id.*

⁸⁴ *Id.*

⁸⁵ *Id.* (valid assessments must comport with predefined standards);

Unmeasured or improper factors that influence the functioning of a system also undermine its accuracy.⁸⁶ When the relative explanatory power, and therefore (ir)rationality, of capital charging decisions is associated with legally illegitimate factors, such as race/ethnicity, gender, and socioeconomic status, then there is reason to believe that those decisions are unconstitutionally arbitrary because prosecutors, judges, and juries are prohibited from considering those factors. A valid assessment is free of racial/ethnic bias and stereotypes so, by definition, decisions based in whole or in part on these factors are legally invalid.⁸⁷ The same would hold true for cases differing only with respect to the gender of the defendant or victim.⁸⁸

Admittedly, the strength of the association between legally impermissible factors and charging-and-sentencing outcomes necessary to qualify as constitutional error is far from obvious. Similar to the aforementioned assessment of arbitrariness, it will be difficult to specify, *a priori*, the magnitude of the association necessary to qualify as constitutionally impermissible error. Relative assessments, again, may be required. A rational death penalty system, however, clearly identifies the factors that decision-makers must not consider when making culpability determinations with respect to charging and sentencing. Conscious discrimination on the part of decision-makers need not be demonstrated in order to

Richard A. Berk et al., *Statistical Difficulties in Determining the Role of Race in Capital Cases: A Reanalysis of Data from the State of Maryland*, 21 J. QUANTITATIVE CRIMINOLOGY 365, 387 (2005) (the low explanatory power of statistical models of capital charging-and-sentencing decisions is attributable to an arbitrary and irrational process).

⁸⁶ FORST, *supra* note 15, at 37.

⁸⁷ Lang & Wilkerson, *supra* note 80 (racial and ethnic bias undermine the validity of an assessment). *But see* Russell D. Covey, *The Unbearable Lightness of Batson: Mixed Motives and Discrimination in Jury Selection*, 66 MD. L. REV. 279 (2006) (describing some courts' endorsement of a "mixed-motives" analysis in jury selection that permits racially-motivated reasoning).

⁸⁸ *E.g.*, *Furman v. Georgia*, 408 U.S. 238 (1972) (noting that unjustified gender disparities in the administration of capital punishment would be unconstitutional).

substantiate a claim of a legally arbitrary system. Not only is such evidence nearly impossible to obtain because actors have very little incentive to admit this type of wrong-doing,⁸⁹ but evidence of unconscious/implicit bias in capital punishment decision-making is equally discriminatory and violates the Court's heightened reliability standard.⁹⁰ While it is true that the U.S. Constitution has never been interpreted to require identical punishments for similar situated defendants,⁹¹ evidence of a robust pattern of the influence of impermissible factors on these charging and sentencing outcomes is highly probative of the level bias permitted by the system.⁹² A justice policy, *de jure* or *de facto*, that biases outcomes towards over- or under-sanctioning produces systematic error. These "less honest errors appear all too common" and typically result from a lack of professionalism coupled with flawed systems of accountability.⁹³

3. Proportionality

Concerns about proportionality have both macro-level and micro-level dimensions. At the macro-level, a punishment can be excessive and therefore prohibited by the Eighth Amendment, if not graduated and proportioned to the offense.⁹⁴

⁸⁹ STEVE WEINBURG, *HARMFUL ERROR: INVESTIGATING AMERICA'S LOCAL PROSECUTORS* (2003) (noting that there is little incentive for prosecutors to admit wrongdoing in murder cases because misconduct is treated with great leniency).

⁹⁰ Sheri L. Johnson, *Unconscious Racism and the Criminal Law*, 73 CORNELL L. REV. 1016 (1988); Scott Howe, *The Futile Quest for Racial Neutrality in Capital Selection and the Eighth Amendment Argument for Abolition Based on Unconscious Racial Discrimination*, 45 WM. & MARY L. REV. 2083 (2004).

⁹¹ See generally *McCleskey v. Kemp*, 481 U.S. 279 (1987); Ernest van den Haag, *The Ultimate Punishment: A Defense*, 99 HARV. L. REV. 1662 (1986).

⁹² E.g., Jon R. Sorensen & Donald H. Wallace, *Prosecutorial Discretion in Seeking Death: An Analysis of Racial Disparity in the Pretrial Stages of Case Processing in a Midwestern County*, 16 JUST. Q. 559 (1999) (drawing inferences of intentional discrimination from statistical data of capital charging decisions by a single prosecutor).

⁹³ FORST, *supra* note 15, at 17.

⁹⁴ *Weems v. United States*, 217 U.S. 349 (1910).

The inquiry does not focus on the individual defendant, rather the gravity of the alleged offense and the harshness of the penalty is at issue.⁹⁵ The Court has narrowed the reach of the death penalty over the past four decades, but has maintained that the death penalty is not, *per se*, excessive for the crime of murder when committed by an adult who does not suffer from extreme intellectual disability.⁹⁶ At the micro-level, a proportionality analysis for an individual defendant requires a comparison to sentences imposed on other criminals in the same jurisdiction (intra-jurisdictional) and sentences imposed for commission of the same crime in other jurisdictions (inter-jurisdictional).⁹⁷ An individual criminal charge or punishment may be excessive because it is used very infrequently against similarly situated defendants (i.e., arbitrary). Similarly, if a charge or punishment is used very frequently against defendants who share a trait that is prohibited from legal consideration, but not against other similarly situated defendants who do not share that trait, that charge or punishment may be excessive.

Disproportionality is a matter of degree and no clear

⁹⁵ The Court's analytical framework for assessing whether a punishment was disproportionate, *per se*, and therefore violative of the Eighth Amendment was announced in *Gregg*. The three-step inquiry involved: (1) assessing whether the punishment employs cruel methods or involved needless and wanton infliction of pain, (2) considering society's evolving standards of decency as reflected by legislative judgments and jury verdicts, and (3) determining whether the punishment offends "human dignity." *Gregg v. Georgia*, 428 U.S. 153 (1976). The Court would later note that "State legislatures and sentencing juries do not wholly determine this controversy [over what constitutes excessive punishment under the Eighth Amendment], for the Constitution contemplates that in the end our own judgment will be brought to bear on the question." *Coker v. Georgia*, 433 U.S. 584 (1977). *See also, supra*, note 64.

⁹⁶ *Roper v. Simmons*, 543 U.S. 551 (2005) (outlawing capital punishment for juvenile offenders, but upholding its constitutionality for adult murderers). *See also Coker v. Georgia*, 433 U.S. 584 (prohibiting the death penalty for non-homicidal rape); *Eberheart v. Georgia*, 433 U.S. 917 (1977) (prohibiting the death penalty for non-homicidal kidnapping); *Atkins v. Virginia*, 536 U.S. 304 (2002) (prohibiting the death penalty for mentally disabled defendants).

⁹⁷ *See, supra*, note 64. *See also Solem v. Helm*, 463 U.S. 277 (1983).

guidelines exist for determining what is excessive in any particular situation. As explained, *supra*, the Court has expressly endorsed capital punishment statutes that required reviewing courts to make proportionality determinations for each defendant receiving a death sentence. Irrespective of the specific designs of those proportionality review systems, meaningful proportionality review requires the development of a data-driven metric that permits the analyst to clearly situate each defendant vis-à-vis other defendants when making these determinations.⁹⁸

C. Summary

Reliability, validity, and proportionality are empirical questions, not legal ones.⁹⁹ They all relate to the patterns of measurement error. By clearly defining how different patterns of error correspond to these distinct, but often related concepts, it is possible to assess the performance of a system. Attorneys in *Furman* advanced factual claims about the death penalty that were subject to social scientific inquiry and, as a consequence, *Furman* contained the most extensive discussion of social science research in any decision before or since.¹⁰⁰ In response to these facts, the *Furman* court announced that systems must be modified to substantially reduce these patterns—the procedures that states’ developed were to be a means to an end: fundamental fairness. The careful collection and analysis of detailed data on death penalty charging-and-sentencing practices still provides the best opportunity to directly explore

⁹⁸ I will discuss proportionality review systems in greater detail in Part III.

⁹⁹ Robert J. Smith, *The Geography of the Death Penalty and Its Ramifications*, 92 B.U. L. REV. 227, 255 (2012); HANEY, *supra* note 31, at 3–25 (explaining that the style of reasoning in *Furman* suggested that the real facts and actual operation of the death penalty would be at the forefront of any future litigation and judicial decisions that pertained to its constitutionality).

¹⁰⁰ HANEY, *supra* note 31, at 10. *Furman* contained over 60 footnotes citing published social science research on the realities of the death penalty process. Justices writing for both the majority and the dissent grappled with social science evidence.

the level of arbitrariness at which a statute operates.¹⁰¹ It can also provide useful insights into the excessiveness of individual charging or sentencing decisions. Part II describes and implements an exportable framework for accomplishing these tasks.

II. DIAGNOSING CAPITAL ERROR

This section presents a statistical model capable of assessing the level of arbitrariness present in a capital charging system and implements the model using charging data from Georgia. Part II.A provides a very general overview of the model and explains why it is particularly well-suited for analyzing capital charging data. A more technical discussion of the model is provided in the Methodological Appendix. Part II.B illustrates the usefulness of the model in identifying macro-level (i.e., systemic) errors through an analysis of eight years of death penalty charging decisions in Georgia. Part II.C demonstrates the usefulness of the model for assessing micro-level errors—i.e., the (in)appropriateness of seeking the death sentence against a defendant in a particular case. Part II.D discusses the generalizability of the empirical findings as well as the exportability of the statistical model.

A. *Modeling Capital Charging*

Prosecutors wield nearly unbridled discretion when making charging decisions.¹⁰² The capital statutes approved by the Court in *Gregg* did very little to regulate prosecutorial behavior,

¹⁰¹ Disaggregating death penalty practices to the county-level (or some other sub-State level unit) permits the identification of a state-level baseline (i.e., stateside average) based on the same governing capital statute. Kyle Graham, *Overcharging*, 11 OHIO ST. J. CRIM. L. 7, 710–11 (2014) (noting the difficulty of defining a widely accepted baseline by which to compare prosecutors' charging decisions).

¹⁰² STEPHANOS BIBAS, *THE MACHINERY OF CRIMINAL JUSTICE* (2012) (describing the unchecked powers of prosecutors); James S. Liebman, *The Overproduction of Death*, 100 COLUM. L. REV. 2030 (2000) (same).

although the defendants in *Gregg* specifically identified the absence of appropriate checks on charging and clemency decisions as a fatal flaw of those capital statutes. Not only does inadequate charge screening increase the risk of arbitrariness, bias, and excessiveness, it can also have strong ripple effects throughout the remainder of the death penalty trial and appellate process: acquittal and reversal rates are greatest in cases that are not highly aggravated or involve defendants with strong mitigation evidence.¹⁰³ The practical consequences of this cycle are millions spent on correcting trial level error and tremendous delay and uncertainty of executions.¹⁰⁴ Modeling capital charging with a specific focus on arbitrariness not only provides invaluable insights into the “front-end” of the death penalty process, but it is also relevant to understanding downstream legal error.¹⁰⁵

¹⁰³ Nationally, one-third of capital trials result in a death sentence, one-third of those *initial* death sentences are upheld appeal, and one-tenth of inmates sentenced to death are executed. Richard C. Dieter, *Smart on Crime Reconsidering the Death Penalty in a Time of Economic Crisis* (Death Penalty Information Center 2009) (reporting conviction and execution statistics); James S. Liebman et al., *Capital Attrition: Error Rates in Capital Cases, 1973-1995*, 78 TEX. L. REV. 1839 (2000) (reporting that 68% of death sentences initially imposed at trial were overturned on appeal). *See also* Part I.

¹⁰⁴ Richard C. Dieter, *Millions Misspent: What Politicians Don't Say about the High Costs of the Death Penalty* (Death Penalty Information Center 1994) (discussing the great financial costs of error correction in capital cases); Arthur L. Alarcon & Paula M. Mitchell, *Executing the Will of the Voters: A Roadmap to Mend or End the California Legislature's Multi-Billion-Dollar Death Penalty Debacle*, 44 LOY. L.A. L. REV. S41 (2010-2011) (describing federal expenditures on capital cases arising in state court).

¹⁰⁵ The vast majority of studies examining the capital punishment process focus on the discretionary choices of actors in the criminal justice system *after* a death penalty notice has been filed (and frequently *after* a death sentence has been imposed). This limited focus largely stems from the fact that most analysts lack adequate information on the population of defendants who might be subject to capital punishment, so researchers concentrate on those cases once they formally enter the system. Focusing on capital charging decisions is especially important because many prosecutors charge defendants capitally in order to induce a plea agreement with the defendant and avoid the time and expense associated with trial. *See, e.g.*, Sherod Thaxton, *Leveraging Death*, 103 J. CRIM. L. & CRIMINOLOGY 475 (2013); Susan Ehrhard, *Plea Bargaining and the Death Penalty: An Exploratory Study*, 29 JUST. SYS. J. 313 (2008); Susan Ehrhard-Dietzel, *The Use of Life and Death*

Consistent with prior research, I model capital charging decisions as a function of legally permissible and impermissible case-level factors.¹⁰⁶ The empirical relationships between these factors and capital charging decisions is described through the use of multivariate regression. Regression analysis is the primary technique employed by quantitatively-oriented lawyers and social/behavioral scientists for the identification of general patterns of arbitrariness and bias in capital decision-making.¹⁰⁷ Indeed, regression is most powerful (and least controversial) when used as a descriptive technique to identify conditional averages and variances based on these observed patterns.¹⁰⁸ When factors specified in a capital statute are highly predictive of actual outcomes, one can infer that the process which generated the outcome is rational and, therefore, non-arbitrary. Relatedly, when like cases are treated similarly, then a process is considered consistent.¹⁰⁹ Prosecutors are prohibited from basing their decisions to seek the death penalty on defendants' or victims' race/ethnicity,¹¹⁰ so if these factors exert an influence on capital charging, all else being equal, one can infer the process that generated the outcomes is legally arbitrary. It is unnecessary to assume that the racial/ethnic

as Tools in Plea Bargaining, 37 CRIM. JUST. REV. 89 (2012).

¹⁰⁶ See, e.g., John J. Donohue, *An Empirical Evaluation of the Connecticut Death Penalty System Since 1973: Are There Unlawful Racial, Gender, and Geographic Disparities?*, 11 J. EMPIRICAL LEGAL STUD. 637 (2014).

¹⁰⁷ E.g., BALDUS ET AL., *supra* note 49; Donohue, *supra* note 106, at 650.

¹⁰⁸ RICHARD A. BERK, REGRESSION ANALYSIS: A CONSTRUCTIVE CRITIQUE 13 (Advanced Quantitative Techniques in the Social Sciences, volume 11, 2003).

¹⁰⁹ A model of prosecutorial charging behavior should approximate the actual data-generating process, so it is necessary to include legally impermissible factors when it is hypothesized that such factors are predictive of outcomes. To do otherwise would imply that the model precludes these variables from having any predictive power, and such an assumption would result in an improperly specified model. This also holds true when modeling discretionary choices at other stages of the criminal justice process. See, e.g., John Wooldredge, *Distinguishing Race Effects on Pre-Trial Release and Sentencing Decisions*, 29 JUST. Q. 41 (2012) (modeling pretrial detention behavior and including race/ethnicity as an explanatory variable).

¹¹⁰ See *Furman v. Georgia*, 408 U.S. 238 (1972); *McCleskey v. Kemp*, 481 U.S. 279 (1987).

disparities are motivated by intentional bias. Compelling evidence of the effects unconscious/implicit bias in capital decision-making may still support a constitutional challenge under the Eighth Amendment.¹¹¹

The empirical strategy adopted in this paper uses a variant of regression analysis to develop a model for analyzing the arbitrariness in prosecutors' charging decisions.¹¹² The model, commonly referred to as a "multilevel model" (MLM) (or "variance components model"), has been under-utilized in death penalty research, but has important advantages over other types of regression when analyzing differences in case outcomes across jurisdictions.¹¹³ I will demonstrate certain of these advantages in the application segment of the analysis, *infra*, Part II.B.¹¹⁴ This Article is the first application of MLMs to

¹¹¹ See, *supra*, note 90 and accompanying text. Although it remains unclear what constitutes "compelling" evidence, the New Jersey Supreme Court stated that it would be receptive to statistical proof of significant racial disparities in the administration of capital punishment and believed the evidence presented in *McCleskey* was very persuasive. *State v. Loftin*, 724 A.2d 129 (N.J. 1999); cf. *United States v. Bass*, 266 F.3d 532 (6th Cir. 2001) (holding that national statistics of racially disparate death penalty charging patterns were sufficient to satisfy the threshold showing of a selective prosecution claim to merit discovery of the prosecution's files pertaining to charging decisions).

¹¹² A detailed discussion of these models is beyond the scope of this project, but very accessible descriptions of these models are readily available. See ANDREW GELMAN & JENNIFER HILL, *DATA ANALYSIS USING REGRESSION AND MULTILEVEL/HIERARCHICAL MODELS* (2007); JOSÉ C. PINHEIRO & DOUGLAS M. BATES, *MIXED-EFFECTS MODELS IN S AND S-PLUS* (2000); STEPHEN W. RAUDENBUSH & ANTHONY S. BRYK, *HIERARCHICAL LINEAR MODELS: APPLICATIONS AND DATA ANALYSIS METHODS* (Advanced Quantitative Techniques in the Social Sciences, volume 1, 2d ed. 2002).

¹¹³ MLMs have a long history in agriculture and educational statistics, where a nested data structure is natural in these settings (e.g., animals on plots, students in classrooms, etc.), and have become more popular in the social sciences in recent years. GELMAN & HILL, *supra* note 112, at 276. Recently, MLMs have been used to study state and demographic trends in public opinion about the death penalty. Shirley & Gelman, *supra* note 12, at 17 (discovering that support levels in northern and southern states have moved in opposite directions over the past 50 years).

¹¹⁴ In brief, these models offer significant improvements with respect to data reduction, prediction, and causal inference. GELMAN & HILL, *supra* note

capital charging decisions.¹¹⁵ A more technical discussion of the model is provided in the Methodological Appendix.¹¹⁶

The MLMs I employ allow a more accurate examination and quantification of the sources of variability in capital charging practices. Uniformity in prosecutorial charging decisions is not required under the Court's death penalty jurisprudence, but substantial variation in charging behavior both within and across jurisdictions that is not attributable to legally legitimate case characteristics may be indicative of arbitrariness or bias (or both) in a system.¹¹⁷ Empirical studies of capital charging-

112, at 246; Marco R. Steenbergen & Bradford S. Jones, *Modeling Multilevel Data Structures*, 46 AM. J. POL. SCI. 218, 219 (2002).

¹¹⁵ A limited number of studies have examined capital charging decisions. Earlier research was criticized for failing to properly account for legally relevant differences across the cases. *See, e.g.*, Gary Kleck, *Racial Discrimination in Criminal Sentencing: A Critical Evaluation of the Evidence with Additional Evidence on the Death Penalty*, 46 AM. SOCIOLOGICAL REV. 783 (1981); Gary Kleck, *Equal Justice and the Death Penalty: A Legal and Empirical Analysis (Book Review)*, 20 CONTEMP. SOCIOLOGY 598 (1991). More recently methodologically rigorous studies revealed that prosecutors were four times more likely to seek the death penalty in white-victim cases compared to black-victim cases in Colorado, three times more likely in Georgia, twice as likely in Maryland, and three times more likely in South Carolina. *See* BALDUS ET AL., *supra* note 49; Stephanie Hindson et al., *Race, Gender, Region and Death Sentencing in Colorado, 1980-1999*, 77 U. COLO. L. REV. 549 (2006); Raymond Paternoster & Robert Brame, *Reassessing Race Disparities in Maryland Capital Cases*, 46 CRIMINOLOGY 971 (2008); Michael J. Songer & Issac Unah, *The Effect of Race, Gender, and Location on Prosecutorial Decisions to Seek the Death Penalty in South Carolina*, 58 S.C. L. REV. 161 (2006); Isaac Unah, *Choosing Those Who Will Die: The Effect of Race, Gender, and Law in Prosecutorial Decision to Seek the Death Penalty in Durham County, North Carolina*, 15 MICH. J. RACE & LAW 135 (2009). Although these studies have contributed, considerably, to our understanding of arbitrariness and bias in capital charging, the research designs fail to fully exploit the county-centric nature of capital decision-making in order to obtain a deeper understanding of the inconsistency in death charging dynamics.

¹¹⁶ *See generally* Andrew Bell & Kelvyn Jones, *Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data*, 3 POL. SCI. RESEARCH & METHODS 133, 139 (2015) (listing the key shortcomings of non-MLMs when applied to data that have a nested structure--e.g., cases nested within counties).

¹¹⁷ *See, e.g.*, *Bush v. Gore*, 531 U.S. 98, 109 (2000) (per curiam) (explaining that the absence of specific standards for the equal application of

and-sentencing decisions routinely discover that the death penalty is a “minority practice” that exerts a huge “majority burden.”¹¹⁸ Capital charging and sentencing activity remains concentrated among a small subset of counties.¹¹⁹ The regional clustering of capital charging decisions is important because “it permits a tailored and rigorous analysis of gauging the continued constitutionality of capital punishment.”¹²⁰ Developing legally cognizable claims about the arbitrariness (and biasedness) of the death penalty can be buttressed by disaggregating statewide data of capital charging practices to sub-state units (e.g., counties), and thereby permitting the careful examination of within- and between-county variability, in the use of capital punishment for (legally) similarly situated defendants.¹²¹ County-disaggregated analyses is relevant at

the law may be unconstitutional under certain circumstances when a government body is empowered to assure uniformity).

¹¹⁸ James S. Liebman & Peter Clarke, *Minority Practice, Majority's Burden: The Death Penalty Today*, 9 OHIO ST. J. CRIM. L. 255 (2011).

¹¹⁹ There are currently 3,143 counties in the United States, but only 15% of these counties account for all executions since the Court lifted its *de facto* moratorium on executions in 1976. Richard C. Dieter, *The Two Percent Death Penalty: How a Minority of Counties Produce Most Death Cases at Enormous Costs to All* (Death Penalty Information Center 2013); Smith, *supra* note 99. From 2004-2009, only 10% of counties returned a single death verdict, *Id.* and approximately 1% of counties consistently produced one death verdict per year over the six year period. *Id.*; See also Dieter, *supra* (noting that only 2% of counties account for the vast majority of death sentences since capital punishment was reinstated in 1976); Richard Willing & Gary Fields, *Geography of the Death Penalty*, USA TODAY, Dec. 20, 1999, at A1 (fifteen counties account for nearly a third of all prisoners sentenced to death, but only one-ninth of the population of the states with capital punishment).

From 1995-2000, 42% of federal death penalty prosecutions submitted for authorization came from just 5 of the 94 federal districts. U.S. Department of Justice, *The Federal Death Penalty System: A Statistical Survey, 1988-2000* (Government Printing Office 2000).

¹²⁰ Smith, *supra* note 99, at 229.

¹²¹ *Id.* at 247; Carol S. Steiker & Jordan M. Steiker, *A Tale of Two Nations: Implementation of the Death Penalty in Executing versus Symbolic States in the United States*, 84 TEX. L. REV. 1869 (2005).

Focusing to a single decision-maker (or smaller group of decision-makers at a particular stage of the capital charging-and-sentencing process) partly addresses the Court's concern about inferences of discrimination drawn from the product of the discretionary choices of multiple decision-

both trial and appellate levels when raising challenges focusing on intra-county, intra-state, and inter-state arbitrariness and bias,¹²² as well as assessing disproportionate (i.e., “excessive”) punishments at the case-level.¹²³ Whereas traditional studies focus on the average effect of case-level factors on outcomes, and treat variation in the impact of these factors across jurisdictions as a “nuisance” that must be taken into account but not closely examined, this article highlights the importance of focusing on this type of variability in the capital charging process.¹²⁴ The utility of the model for analyzing actual capital

makers. *McCleskey v. Kemp*, 481 U.S. 279 (1987) (criticizing petitioner’s use of data from multiple decision-makers in the capital charging-and-sentencing process to draw inferences of racial discrimination); Sorensen & Wallace, *supra* note 92 (finding evidence of intentional racial discrimination in death charging by a prosecutor).

¹²² See, e.g., Raymond Paternoster et al., *Justice by Geography and Race: The Administration of the Death Penalty in Maryland, 1978-1999*, 4 U. MD. L.J. RACE RELIGION GENDER & CLASS 1 (2004) (discussing variation in death charging-and-sentencing across counties for similarly situated defendants); Ashley Rupp, *Death Penalty Prosecutorial Charging Decisions and County Budgetary Restrictions: Is the Death Penalty Arbitrarily Applied Based on County Funding?*, 71 FORDHAM L. REV. 2735 (2003) (same).

Kentucky became the first state to enact a “Racial Justice Act.” The law, initially, permitted defendants to present statistical or other evidence suggesting that their race, the race of their victim(s), or both, played a significant part in prosecutor’s decision to seek the death sentence in their particular case. If a defendant prevailed on her claim, the prosecutor was required withdraw the capital charge. Kentucky Racial Justice Act, Ky. Rev. Stat. Ann. § 532.300 (Ky. 1998). See also David C. Baldus et al., *False Attacks on the Racial Justice Acts*, DES MOINES REGISTER, Jun. 7, 1994, at A9 (describing unsuccessful efforts to enact Racial Justice Acts at the federal and state levels); Michael Mears, *Georgia Needs a Racial Justice Act*, DAILY REPORT, Sept. 25, 1998 (same).

¹²³ See, *supra*, Part I. See, also, *Enmund v. Florida*, 458 U.S. 782 (1982) (describing the framework reviewing courts must adopt in assessing the constitutionality of a punishment under the Eighth Amendment); *Solem v. Helm*, 463 U.S. 277 (1983) (same).

¹²⁴ Raymond J. Carroll, *Variances Are Not Always Nuisance Parameters*, 59 BIOMETRICS 211 (2003) (emphasizing the importance of modeling the variance structure in statistical analysis); Donald Hedeker et al., *Modeling Between- and Within-Subject Variance in Ecological Momentary Assessment (EMA) Data Using Mixed-Effects Location Scale Models*, 31 STAT. MED. 3328, 3328 (2012) (arguing the importance of investigating why subjects differ in variability rather than just their average level).

charging decisions is demonstrated in the next section.

B. Applying the Model to Georgia

1. Georgia Capital Charging Data

I use the aforementioned model to diagnose Georgia's capital charging decisions. Specifically, I use the model to analyze eight years of death penalty charging from Georgia (1993–2000) in an effort to assess the potential arbitrariness in death penalty charging decisions. This timeframe was selected for three important reasons. First, Georgia enacted its life without the possibility of parole (LWOP) statute in 1993.¹²⁵ The legislation was specifically designed as a sentencing alternative in capital murder trials, therefore potentially having a substantial impact on prosecutorial, judicial, and jury discretion. Second, Georgia created an agency tasked with collecting data on capital charging-and-sentencing decisions throughout the state in October 1992, so comprehensive data is only available after that point.¹²⁶ Finally, the year 2000 was selected as the cut-off in order to allow sufficient time for all of the cases to advance from the charging phase through the (initial) penalty phase.

I collected and complied these data from five different sources: the Georgia Bureau of Investigation (GBI),¹²⁷ the

¹²⁵ 1993 Ga. Laws 569, § 4; Ga. Code Ann. § 17-10-30.1 (1993).

¹²⁶ MICHAEL MEARS, *THE DEATH PENALTY IN GEORGIA: A MODERN HISTORY, 1970-2000* (1999).

¹²⁷ The GBI collects data on all homicides known to the police in the state, disaggregated by year, month, and county of occurrence. Information in the GBI data include, among other things, the age, race/ethnicity, and gender of the victim(s) and alleged offender(s), the circumstances under which the homicide took place (e.g., robbery, burglary, etc.), the relationship between the victim and the offender, and the weapon used in the homicide. The major shortcomings of these data are the high rate of missing information and the limited information about each homicide.

Georgia Department of Corrections (GDC),¹²⁸ the Office of the Georgia Capital Defender (GCD),¹²⁹ the Clerk's Office of the Georgia Supreme Court (CO),¹³⁰ and the Atlanta Journal-Constitution (AJC).¹³¹ These various sources were consulted in order to obtain as much relevant case-level data on death-eligible defendants. From these sources, I coded facts about the defendant, codefendant(s), victim(s), judge, prosecutor, defense counsel, and the crime. The data consist of the entire population of homicide cases, and not a mere sample, so statistical inference based on sample statistics (e.g. *p*-values, significance tests, confidence intervals, etc.) does not apply in the convention sense; instead, attention is given to the direction and magnitude of the statistical parameters and quantities of interest derived from these parameters.¹³²

¹²⁸ The GDC compiles data on every defendant convicted of a criminal homicide, which includes both murder and manslaughter. Included in the GDC data is information on offender demographic characteristics, criminal history, employment status at time of the offense, alcohol/drug use and abuse history, and family background. The GDC does not keep information on victims.

¹²⁹ The GCD is a trial resource center for capital defense attorneys in Georgia. The attorneys from GCD serve as lead/co-counsel or consultants on the vast majority of capital cases throughout the state. The office keeps basic information on every capital case in which it is involved, including but not limited to demographic characteristics of defendant(s), victim(s), attorneys, and judges; defendant criminal and mental health history, crime-specific information; and method of disposition.

¹³⁰ Under Georgia law, all notices of intent to seek the death penalty must be filed with the CO. The clerk maintains a list of all death notices filed, recording the name of the defendant, the date the notice was filed, the county in which the notice was filed, and the name(s) of the prosecutor filing the notice. The Clerk's Office keeps case files from the local county courts for all cases resulting in a death sentence. These files are used by the Georgia Supreme Court when reviewing death sentences on direct appeal.

¹³¹ The AJC is Georgia's flagship daily newspaper. The newspaper collected information about death-eligible cases throughout the state as part of an investigative article on Georgia's capital punishment system. These data were made publicly available. Bill Rankin et al., *High Court Botched Death Reviews*, ATLANTA J. & CONST., Sept. 26, 2007, at A1.

¹³² See generally Darrell Steffensmeier & Stephen Demuth, *Ethnicity and Judges' Sentencing Decisions: Hispanic-Black-White Comparisons*, 39 CRIMINOLOGY 145, 160 (2001) (explaining that sample-based significance tests are inappropriate when analyzing the entire population of cases);

I created a master list of all potentially capital cases from which prosecutors could identify and select defendants for the death penalty. Georgia's death penalty statute lists eleven elements making a crime eligible for the death penalty,¹³³ and the list of death-eligible defendants is comprised of all persons 17 years of age or older who were convicted of murder and at least one of the eleven special elements was present.¹³⁴ These data allow for the determination of which defendants were factually eligible for the death penalty and which defendants received a formal death notice.¹³⁵

During the period under investigation (1993–2000), there were 1,238 cases resulting in a murder conviction that were eligible for the death penalty under Georgia's capital statute.¹³⁶ Prosecutors filed a notice of intent to seek the death penalty in 400 cases and 54 defendants ultimately received the death penalty. Of the 395 capitally charged cases in which the method of disposition is known, 59 percent were ultimately resolved by plea and 41 percent were resolved by trial. Cases that were technically death-eligible under the Georgia statute but were not charged with a capital crime were disposed by plea 39 percent of the time and by trial 61 percent of the time.

Georgia's Administrative Office of the Courts (AOC) organizes the state's 159 counties into 49 superior court judicial

accord Adrian E. Raftery, *Bayesian Model Selection in Social Research*, in 25 *SOCIOLOGICAL METHODOLOGY* 111 (Peter V. Marsden ed., 1995).

¹³³ See *supra* Part I.A; GA. CODE ANN. § 17-10-30 (West 2003).

¹³⁴ In *Roper v. Simmons*, 543 U.S. 551 (2005), the Court ruled that the death penalty was unconstitutional for defendants who were juveniles at the time they committed their crimes. Prior to *Roper*, Georgia permitted the death penalty for defendants ages 17 and older. The data examined in this paper focus on the pre-*Roper* period. Slightly under 3% of offenders in the data were technically death eligible but under 17 years of age.

¹³⁵ See Table 1 for descriptive statistics of the variables employed in this study.

¹³⁶ The data include 28 instances where a defendant was charged with the death penalty but the defendant was either acquitted of the capital crime or convicted of a non-homicide offense. The current analysis is limited to cases resulting in a murder conviction as a proxy for strength of evidence.

circuits.¹³⁷ Cases from the various counties were grouped at the judicial circuit level because, in Georgia, there is one district attorney per judicial circuit. While large counties comprise a single judicial circuit, many smaller counties are grouped together to form a judicial circuit. As a result, a single prosecutor may be responsible for charging and plea bargain decisions for several counties in her judicial circuit (see Table 2). Also, if a judicial circuit consists of multiple counties, trial judges rotate throughout these counties. Treating counties that share a single judicial circuit as if they were independent ignores the similarities they share in the administration of capital punishment resulting from shared decision-makers.¹³⁸ Figure 1 displays a map of Georgia, divided into counties and judicial circuits.¹³⁹

Relevant Variables. Prior research suggests the primary legally legitimate determinants of capital charging decisions are the: (1) death eligibility factors enumerated in the capital statute, (2) criminal charges concurrent with the homicide, (3) defendant's prior criminal history, and (4) relationship between

¹³⁷ Judicial Council of Georgia, *Your Guide to the Georgia Courts* (Administrative Office of the Courts 2003). The number of judicial circuits was 46 during the time period of the study.

¹³⁸ Issac Unah & John C. Boger, Race, Politics, and the Process of Capital Punishment in the South, presented at American Political Science Association 1, 32 (Boston, Mass. Sept. 2002). Death penalty cases are extremely rare events, so aggregating county-level data to the judicial circuit level offers the additional benefit of more cases per contextual unit. Better statistical estimates of the relationships occurring at both the case- and contextual-level can be obtained without altering the dependence structure of the cases due to their clustering. RAUDENBUSH & BRYK, *supra* note 112.

¹³⁹ The map also shows that Georgia's 49 judicial circuits are organized into Judicial Administrative Districts. These districts were created to provide regional court administration to the superior courts of Georgia. The districts were created along Georgia Congressional District lines and each district is served by an administrative judge and district court administrator selected by the superior court judges and senior judge in each particular district. The primary function of the administrative judge is to assist chief judges in preparing, presenting, and managing local court budgets. These judges, however, do not exert any influence on the capital charging-and-sentencing process in the judicial circuits that comprise the judicial district.

the defendant and victim(s).¹⁴⁰ Professor John Donohue's recent examination of Connecticut's death penalty system, which formed the centerpiece of litigation over its constitutionality, included nearly an identical set of variables that are included in my analysis.¹⁴¹ Methodologically rigorous examinations of discretionary choices of prosecutors, judges, and juries have included as many as 230 explanatory variables and as few as a dozen.¹⁴² The explanatory power of these models do not significantly vary—whether including a few hundred variables, or just a few handfuls. Moreover, the actual parameter estimates (i.e., regression coefficients) for many of these factors are statistically indistinguishable across the study designs, although the complexity of the models varied considerably.¹⁴³

It also must be emphasized that many of the factors impacting *capital sentencing* are unknown to prosecutors or defense attorneys at the time of the *capital charging* decision, and specifics about aggravation and mitigation evidence come to light in preparation for trial. As a result, an analytical model predicting capital sentencing decisions would necessarily include a more comprehensive set of variables stemming from the fact that prosecutors and defense counsel have access to a much wider range of information at the stage of the adjudicatory process.¹⁴⁴ It appears that a limited number of factors are

¹⁴⁰ David C. Baldus & George Woodworth, *Race Discrimination in the Administration of the Death Penalty: An Overview of the Empirical Evidence with Special Emphasis on the Post-1990 Research*, 39 CRIM. L. BULL. 194 (2003) [hereinafter Baldus & Woodworth, *Race Discrimination in the Administration of the Death Penalty*] (reviewing empirical literature on capital charging-and-sentencing decisions); David C. Baldus & George Woodworth, *Race Discrimination in America's Capital Punishment System Since Furman v. Georgia (1972): The Evidence of Race Disparities and the Record of Our Courts and Legislatures in Addressing this Issue* (American Bar Association 1997) [hereinafter Baldus & Woodworth, *Race Discrimination in America's Capital Punishment System Since Furman v. Georgia (1972)*] (same).

¹⁴¹ Donohue, *supra* note 106, at 646.

¹⁴² Baldus & Woodworth, *Race Discrimination in the Administration of the Death Penalty*, *supra* note 140.

¹⁴³ *Id.*

¹⁴⁴ See, e.g., Raymond Paternoster & Robert Brame, *An Empirical Analysis of Maryland's Death Sentencing System with Respect to the Influence of Race*

predictive of capital charging. An important limitation on model complexity of nearly every death penalty study is the small number of relevant cases. Even with study designs that examine the entire population of death eligible defendants under a period of study, there are limits on the amount of relevant data that can be examined. For example, recent studies of capital punishment systems in Colorado ($N=539$),¹⁴⁵ Connecticut ($N=205$),¹⁴⁶ Maryland ($N=1,041$),¹⁴⁷ North Carolina ($N=151$),¹⁴⁸ Texas ($N=504$),¹⁴⁹ and Washington ($N=266$)¹⁵⁰ all examined datasets with a smaller number of death eligible cases than the current study.¹⁵¹

These data limitations constrain the number of variables that can be included in any model that is applied to actual data. Analysts have typically developed streamlined models that include a few dozen explanatory variables, at most. I employ a model specification that includes 40 case-level variables indexing the heinousness of the crime and the culpability of the

and *Legal Jurisdiction* (University of Maryland, College Park 2003) (using different subsets of variables to model the death noticing and death sentencing decisions).

¹⁴⁵ Justin Marceau et al., *Death Eligibility in Colorado: Many Are Called, Few Are Chosen*, 84 U. COLO. L. REV. 1069, 1106 (2013).

¹⁴⁶ Donohue, *supra* note 106.

¹⁴⁷ Paternoster & Brame, *supra* note 115.

¹⁴⁸ Unah, *supra* note 115.

¹⁴⁹ Scott Phillips, *Continued Racial Disparities in the Capital of Capital Punishment: The Rosenthal Era*, 50 HOUS. L. REV. 156 (2012–2013).

¹⁵⁰ Heather Evans & Katherine Beckett, *Race, Death, and Justice: Capital Sentencing in Washington State, 1981-2014*, 6 COLUMBIA JOURNAL OF RACE AND LAW 77 (2016).

¹⁵¹ A recent study of South Carolina death charging decisions examined 2,227 non-negligent homicide cases occurring between 1993 and 1997, and identified 130 cases where a death penalty notice was actually filed. Songer & Unah, *supra* note 115, at 176. Although researchers investigated a significantly large population of homicides, they did not differentiate murder cases from non-negligent (i.e., voluntary) manslaughter cases. While it may be the case that non-negligent manslaughter cases may contain special circumstances enumerated in the South Carolina death penalty statute, these cases cannot be deemed “death eligible” because only homicides committed with either expressed or implied malice are potentially subject to the death penalty.

defendant (see Table 1). These variables are organized into three categories: *crime-related factors* (statutorily defined aggravating factors, type of murder weapon, motive for killing, strength of evidence,¹⁵² and jurisdiction where killing occurred), *defendant-related factors* (number of defendants, defendant's race/ethnicity, sex, age, level of education, employment status, marital status, number of children, military service, history of drug use, psychiatric status, IQ Score, troubled family history, prior felony conviction, county of residence, and trigger-person status), and *victim-related factors* (number of victims, victims' race/ethnicity, sex, age, and prior relationship with defendant). Important inculpatory variables include the total number of statutorily defined aggravating circumstances present in the case, defendant's contemporary convictions and prior criminal history, money- or sex-related motive, the number of victims, the relationship between the defendant and the victim(s) and the age of victim. Potentially mitigating factors include the defendant's age, marital status, educational background, and employment history, troubled family history, military service, history of drug and alcohol use/abuse, psychiatric status, IQ, and religious affiliation).¹⁵³ The race/ethnicity of the defendant and victim(s), which are included in the model, are clearly legally impermissible factors.¹⁵⁴

It cannot be emphasized enough that the Georgia dataset includes much more information than the 40 variables included in the model specification. Moreover, the model actually

¹⁵² Consistent with prior research, we limit our analysis to cases that ultimately resulted in a conviction for murder as a proxy for the strength of evidence in the case. BALDUS ET AL., *supra* note 49, at 40–42, 477.

¹⁵³ Admittedly, many of these factors can cut in either direction, so they may be deemed as aggravating depending on the situation, but the direction of the effect in any individual case is immaterial for the purposes of the analysis. The overall effect is estimated from the data and, therefore, reflects the manner in which prosecutors, on average, treat these factors.

¹⁵⁴ The fact that a homicide involved a defendant and victim(s) from different racial/ethnic groups does not imply that the homicide was racially motivated. Twenty-eight percent of the death eligible murders in the data were “interracial.” An examination of a random sample of interracial homicide cases did not suggest that there was a pattern of racially-motivated killings.

includes more information than the 40 variables imply because I employ a conservative counting method in order to reduce the number of parameters that must be estimated in the model.¹⁵⁵ For example, in terms of inculpatory/aggravation evidence, I have information on the presence or absence of the eleven statutorily defined special circumstances enumerated in Georgia's capital statute, but rather than count them separately, I combined them into a single variable that indexes the total number of statutory aggravating circumstances present in the case. Similarly, with respect to exculpatory/mitigation evidence, for example, I have information on the presence or absence of five types of "troubled family background" factors. I combine these factors into a single variable, capturing the total number of problematic family features occurring in a defendant's background. This is an extremely important point to highlight because studies purporting to include 100-plus or 200-plus variables in their analysis are employing a different counting method,¹⁵⁶ but even these studies report final results from the stable models that only include a handful or a few dozen variables.¹⁵⁷ The fatal flaw in including so many variables is the well-known "curse of dimensionality."¹⁵⁸

¹⁵⁵ See *infra* note 158 and accompanying text.

¹⁵⁶ For example, Weiss and colleagues disaggregate a defendants' and victims' ages into, respectively four and five variables. This results nine variables capturing age, whereas I use a total of two variables. Similarly, Weiss et al. use fifteen variables to capture the various defendant-victim racial combinations; whereas my model includes three variables. Robert E. Weiss et al., *Death Penalty Charging in Los Angeles County: An Illustrative Data Analysis Using Skeptical Priors*, 28 SOCIOLOGICAL METHODS & RESEARCH 91, 94–95 (1999).

¹⁵⁷ See, e.g., BALDUS ET AL., *supra* note 49 (examining models as large as 230 variables, but settling on a final model with 43 variables); Paternoster & Brame, *supra* note 144 (finding little difference in the predictive ability of models with as many as 176 variables and a few as a dozen); Issac Unah & John C. Boger, *Race and the Death Penalty in North Carolina: An Empirical Analysis, 1993-1997* (The Common Sense Foundation 2001) (examining models with upwards of 100 variables, but finding similar predictive ability with much a simpler model that included 37 variables); Weiss et al., *supra* note 156 (constructing a database of 600 variables, but including less than 20 variables in any single model specification).

¹⁵⁸ The curse of dimensionality refers to the problem arising from fitting complex models to data. Regression analysis becomes more difficult as the

The model also includes the jurisdiction of every capital case in the dataset. Evidence of regional disparities in death penalty charging and sentencing has been characterized as proof of arbitrariness, bias, or both,¹⁵⁹ but that debate need not be resolved here. It suffices that, for the purposes of these analyses, any sizable association between location and the likelihood a defendant receives a death notice that cannot be accounted for by case-level factors undermines the reliability and rationality of the system, so at a minimum, it is indicative of arbitrariness.¹⁶⁰ Professor Richard Berk has remarked, “If the concern is about illegitimate factors affecting capital cases, the impact of location needs to be studied in much greater depth. As now measured, a county or city is just a proxy for processes that are not analyzed.”¹⁶¹

2. Arbitrariness as Unreliability

Intra-Class Correlation Coefficient Type-1. An important advantage of MLMs over prior approaches to analyzing capital charging data is the ability of the model to assess the level of consistency in prosecutorial charging decisions.¹⁶² MLMs

number of estimated relationships (i.e., parameters) increases relative to the number of data points. When models contain a large number of variables, it may not be possible to estimate the parameters in the model because there is not sufficient variation on all of those variables across observations. RICHARD E. BELLMAN, *ADAPTIVE CONTROL PROCESSES: A GUIDED TOUR* 94 (1961).

¹⁵⁹ Cf. Smith, *supra* note 99 (describing inter-county variation in death penalty decision-making as being suggestive of arbitrariness); Paternoster et al., *supra* note 122, at 1978–99 (characterizing geographic differences in the administration of the death penalty as bias). See also *Glossip v. Gross*, 135 S. Ct. 2726 (2015) (Breyer, J., dissenting) (noting that geography, like race and gender, impermissibly affect the application of the death penalty).

¹⁶⁰ See *supra* Part I.B. Maintaining the analytical clarity between arbitrariness and bias assists in systematically evaluating capital charging systems, but does not negate the relationship between the two concepts.

¹⁶¹ Berk et al., *supra* note 85, at 387.

¹⁶² Katherine J. Klein & Steve W.J. Kozlowski, *From Micro to Meso: Critical Steps in Conceptualizing and Conducting Multilevel Research*, 3 *ORG. RESEARCH METHODS* 211, 224 (2000).

permit the calculation of the intra-class correlation coefficient, ICC_1 , which measures the degree of stability (i.e., reliability) of death charging decisions within judicial circuits.¹⁶³ In other words, the ICC_1 “can also be thought of as the correlation among units within the same group”¹⁶⁴ or “as an estimate of the extent to which raters are interchangeable—that is, the extent to which one rater from a group may represent all the raters within the group.”¹⁶⁵ $ICC_1 = \psi / (\psi + \theta)$, where ψ is the between-circuit variance in capital charging and θ is the within-circuit variance, and the statistic ranges from 0 to 1 (or 0% to 100%).¹⁶⁶ When explanatory variables are included in the model, the within- and between-circuit variances are “residual variances” because of the effects of those variables are removed from the variances, permitted an assessment of the degree of variability that is unaccounted for by the explanatory variables. The ICC_1 can also be interpreted as the proportion of the total variation attributable to variation between clusters.¹⁶⁷ When the ICC_1 is large, a single case from a circuit is likely to provide a reliable estimate of the other cases (i.e., the group average), and therefore one can infer a strong dependency (i.e., consistency) across cases in the circuit.¹⁶⁸ On the other hand, when the ICC_1 is small, multiple cases are necessary to provide a reliable estimate of the group average, and therefore charging decisions for individual cases can be viewed as inconsistent (or independent).¹⁶⁹ Scholars disagree about the threshold the ICC_1 statistic must meet in order for a group of ratings to be deemed consistent. Arguably, thresholds for reliability should be problem-specific and researchers should avoid adopting any

¹⁶³ Klaus Larsen & Juan Merlo, *Appropriate Assessment of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression*, 161 AM. J. EPIDEMIOLOGY 81, 82–83 (2005).

¹⁶⁴ GELMAN & HILL, *supra* note 112, at 448.

¹⁶⁵ Klein & Kozlowski, *supra* note 162, at 224.

¹⁶⁶ The ICC is sometimes denoted as *lambda* (λ) or *rho* (ρ) because it is both a correlation coefficient and a reliability measure. See also Methodological Appendix.

¹⁶⁷ GELMAN & HILL, *supra* note 112, at 448.

¹⁶⁸ *Id.* at 258.

¹⁶⁹ *Id.*

universal cut-off criterion.¹⁷⁰ A general rule of thumb, however, is that an ICC₁ value of .7 (or 70%) is indicative of a reliable system.¹⁷¹

The unadjusted model yields an ICC₁ of .14 (or 14%), indicating considerable inconsistency in capital charging behavior across prosecutors. It must be emphasized, however, that this statistic does not take into account potentially important factual differences in cases, so it is not extremely helpful in understanding capital charging dynamics, in and of itself. We expect factually dissimilar cases to be treated differently by prosecutors both within and between circuits. Nevertheless, the unadjusted ICC₁ provides a baseline to which comparisons can be made when legally legitimate case characteristics are added to the model (i.e., the adjusted model). The adjusted model, including 40 case-level predictors, produced an ICC₁ of .19 (or 19%). In other words, 81% of the variability in death noticing for factually similar cases is attributable to within-circuit dynamics.¹⁷² The increase in ICC₁ from 14% to 19% suggests slightly better within-circuit consistency in the handling of cases once legitimate case characteristics are taken into account. The within- and between-circuit variance components proportions will always sum to 1 (or 100%), so the increase in ICC₁ based on the adjusted model also means that factually similar cases are treated less consistently across circuits than factually dissimilar cases.¹⁷³ Nonetheless, the overall reliability in charging

¹⁷⁰ Charles E. Lance et al., *The Sources of Four Commonly Reported Cutoff Criteria What Did They Really Say?*, 9 ORG. RESEARCH METHODS 202, 205–7 (2006) (discussing the purported origins of the cut-off criterion for reliability measures).

¹⁷¹ The greater between-group variance relative to within-group variance, the larger the value of the ICC₁ statistic. Lance et al., *supra* note 170; Klein & Kozlowski, *supra* note 162, at 225.

¹⁷² Klein & Kozlowski, *supra* note 162, at 224.

¹⁷³ Although most of the unexplained variability in charging can be attributed to inconsistencies in the manner in which different prosecutors handle similarly situated defendants, (i.e., within-circuit variability), the between-circuit dynamics are still meaningful: a rule of thumb is that an ICC₁ over 5% is substantively meaningful and that inter-circuit based inconsistency cannot be ignored. TOM A.B. SNIJDERS & ROEL J. BOSKER,

behavior based on cases that are factually similar is extremely low.

MLMs also permit the calculation of the relative consistency of charging decisions across defendant and victim racial/ethnic groups.¹⁷⁴ Specifically, I evaluate whether inconsistency in prosecutorial discretion appears to be a function of race/ethnicity.¹⁷⁵ Recall that the ICC₁ for the adjusted model of the entire sample is 19%, suggesting a very low, but non-trivial amount of within-circuit consistency. The ICC₁ for Caucasian-defendant and non-Caucasian-defendant cases is 21% and 14%, respectively. For Caucasian and non-Caucasian victim cases, the ICC₁ values are, respectively, 20% and 14%. Although the differences between the subgroups in terms of consistency are not especially stark, these results provided evidence that the relative reliability in charging behavior is related to legally impermissible factors, and therefore indicative of an unacceptable risk of arbitrariness.

Intra-Class Correlation Coefficient Type-2. In addition to determining the level of consistency in charging behavior across cases within a circuit, I evaluate the reliability of the estimated circuit-level effects in Georgia. I use a variant of the ICC, called ICC₂, to determine whether circuits can be meaningfully differentiated in terms of their death noticing behavior—that is, are the observed differences between circuits sufficiently pronounced to suggest that the inter-circuit variability is an important feature of capital charging dynamics.¹⁷⁶ ICC₂ =

MULTILEVEL ANALYSIS: AN INTRODUCTION TO BASIC AND ADVANCED MULTILEVEL MODELING (1999).

With data drawn from a sample, one would examine whether the variance was statistically distinguishable from zero. Such an examination is unnecessary because the data analyzed in this study include the entire population of death eligible cases in Georgia during the time period under investigation, and not a mere sample. See, *supra*, note 132. If our data were drawn from some “hyper-population,” the results still suggest that the variation in capital charging is statistically significant.

¹⁷⁴ See *supra* Part II.B.2.

¹⁷⁵ Hedeker et al., *supra* note 124, at 3328.

¹⁷⁶ Klein & Kozlowski, *supra* note 162, at 225; RAUDENBUSH & BRYK, *supra* note 112, at 111.

$\psi/(\psi + \theta/\bar{n}_j)$, where ψ is the between-circuit variance, θ is the within-circuit variance, and $\bar{n}_j$ is the average number of death eligible cases per circuit.¹⁷⁷ The intuition behind the measure is that circuit-specific effects based on circuits that, on average, contain many cases are more stable and useful measures of circuit-level properties than circuit-specific effects based on circuits that have, on average, only a few cases.¹⁷⁸ Similar to the ICC₁, the range for ICC₂ is from 0 to 1 (or 0% to 100%). ICC₂ will always be larger than ICC₁ for the same model, with higher ICC₂ scores indicating significant between-circuit variability.¹⁷⁹ The ICC₂ for the unadjusted and adjusted models are, respectively, 75% and 80%—both well above the .7 (or 70%) threshold.¹⁸⁰ These statistics provide rather strong evidence that the circuit-level charging practices are highly variable (i.e., inconsistent) and the inclusion of legally relevant case-level predictors have a negligible impact on increasing consistency across circuits.

This circuit-level reliability can also be depicted graphically to provide a more intuitive presentation of the results. Circuit-specific effects were calculated for the unadjusted and adjusted models, and these circuit-specific effects were plotted.¹⁸¹ The graphical depiction provides two key pieces of information: (1) the magnitude of each circuit-specific effect and (2) the overall variability of the circuit-specific effects throughout the Georgia. Viewed collectively, the graph permits a visual assessment of the “institutional performance” of the individual circuits—that is, which circuits appeared to be “rogue” jurisdictions with respect to their charging behavior—and the level randomness or idiosyncrasy at the jurisdictional level. Figures 2 and 3 display unadjusted variation in death noticing across jurisdictions. That is, these two figures display inter-circuit

¹⁷⁷ RAUDENBUSH & BRYK, *supra* note 112, at 111.

¹⁷⁸ Klein & Kozlowski, *supra* note 162, at 225.

¹⁷⁹ Large group sizes generally result in more stable mean scores, therefore it is possible to have high ICC₂ values and low ICC₁ values. RAUDENBUSH & BRYK, *supra* note 112, at 72.

¹⁸⁰ See *supra* note 171 and accompanying text.

¹⁸¹ For a description of the calculation of circuit-specific effects, see Methodological Appendix.

differences in the probability that a death-eligible homicide will be charged capitally. Figure 2 depicts the information spatially on a map of Georgia. The legend located to the left of the map displays a color grid (in grayscale) corresponding to the magnitude of the probability of receiving a death notice for each circuit. The black horizontal line in Figure 3 represents the statewide probability of a death-eligible defendant receiving a death notice (.33) and the circles denote the probability of a death notice for each judicial circuit.¹⁸² The probability of a death-eligible defendant receiving a death notice ranges from approximately .06 (Atlanta Circuit) to .62 (Ocmulgee Circuit).

Median Absolute Deviation. I also calculated the mean absolute deviation (MAD), which is an estimate of the spread of the circuit-specific effects and is calculated by subtracting the mean of a distribution of circuit-effects (i.e., the statewide average) from each of the absolute values of the circuit-specific effects and then taking the mean of the resulting scores: $\sum |\beta_{0j} - \gamma_{00}| / J$, where β_{0j} is the circuit-level predicted probability of a death notice, γ_{00} is the statewide predicted probability, and J is the total number of circuits.¹⁸³ The MAD for the unadjusted model is .11 (or 11%). In other words, the “typical” circuit death-noticing probability differs from the state-wide probability by 11 percentage points. Both MAD and ICC₂ statistics measure interrater reliability, but do so in slightly different ways. The MAD assesses how different, on average, are the circuits from the state-wide average; whereas the ICC₂ captures whether circuits can be reliability differentiated in terms of their charging patterns.¹⁸⁴

Figures 4 and 5 display the adjusted variation in the probability of receiving a death notice across Georgia’s judicial circuits. The figures reveal that cases which are factually similar

¹⁸² Because this is the statewide probability of a death notice, based on all of the death-eligible cases in the state during the time period, this is the probability that the “average” case receives a death notice.

¹⁸³ *DICTIONARY OF STATISTICS & METHODOLOGY*, *supra* note 75, at 89. *See also* Methodological Appendix.

¹⁸⁴ *See supra* note 179 and accompanying text.

along several key dimensions, including the overall level of aggravation according to Georgia's capital statute,¹⁸⁵ are still handled very differently across Georgia's judicial circuits with respect their probability of being formally noticed for the death penalty. The expected probability of a case receiving a death notice ranges from .11 to .59. The MAD for the adjusted model is .8 (or 8%), so the inclusion of the 40 case-level factors decreases the average circuit deviation from by the state-wide average by about 3 percentage points.¹⁸⁶ Figure 6 overlays Figures 2 and 4 and indicates the relatively small impact individual-level case characteristics have on explaining circuit-level inconsistency in death-noticing in Georgia.

Median Odds Ratio. A third useful measure for assessing between-circuit variability is the median odds ratio (MOR). The MOR quantifies the variation between circuits by comparing two charging decisions in factually similar cases from two randomly chosen circuits. The MOR is the average ratio between the cases of higher propensity with the cases of lower propensity. $MOR = \exp[\sqrt{2 \times \psi} \times \Phi^{-1}(0.75)]$, where ψ is the between-circuit variance, $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution, and $\exp[\cdot]$ is the exponential function.¹⁸⁷ The MOR encapsulates the increased risk that would occur if a particular case moved from one context to another. For the unadjusted model, the MOR describes the extent to which the outcome depends on context. When covariates are included, the MOR is a measure of the variation between outcomes across circuits that is not explained by the explanatory variables. The MOR will always be greater than or equal to 1. If the MOR is 1, then there is no variation between circuits. The MOR and the ICC are complementary, but not equivalent, measures.¹⁸⁸ The MOR for the unadjusted and

¹⁸⁵ GA. CODE ANN. § 17-10-30 (2000).

¹⁸⁶ The median absolute deviation—which is more resistant to extreme circuit values—for the unadjusted and adjusted models are, respectively, .09 and .07.

¹⁸⁷ Larsen & Merlo, *supra* note 163, at 82–83.

¹⁸⁸ Some analysts believe the MOR is more interpretable than the ICC₁ because there is a parallel between the coefficients of the explanatory

adjusted models are, respectively, 2.03 and 2.32. Both of these statistics indicate that, on average, a case moving from a lower probability death charging circuit to the higher probability death charging circuit is a least twice as likely to be noticed for the death penalty. Of greater significance, however, is that factually similar cases are treated more dissimilarly across circuits. The difference is slight, but supports the earlier findings from the ICC₁ analysis—the addition of relevant explanatory variables increase between-circuit inconsistency.¹⁸⁹

These results provide rather clear evidence of legally unjustifiable inconsistency in charging behavior. While some variation across prosecutors and judicial circuits is to be expected—and perhaps even valued—the fact that the inconsistency in charging practices, at both the individual and circuit levels is so severe and largely unaffected by important case-level characteristics should give even the staunchest death penalty retentionists pause.

3. Arbitrariness as Invalidity

Coefficient of Determination. Another benefit of the MLM approach is the ability to describe how well case-level factors explain the observed variability in capital charging decisions. The coefficient of determination (R^2) describes the reduction in the proportion of residual variance based on the model of interest.¹⁹⁰ In other words, it measures the improvement in the

variables, β , and the coefficient of the between-circuit random effect, ζ_j . See, e.g., Germán Rodríguez & Irma Elo, *Intra-Class Correlation in Random Effects Models for Binary Data*, 3 STATA J. 32, 43 (2003); Larsen & Merlo, *supra* note 163, at 82–83.

¹⁸⁹ See *supra* note 173 and accompanying text.

¹⁹⁰ To be sure, the R^2 measure has been the subject of intense debate and abuses of the statistic are commonplace; nonetheless it is still one of the most widely used model summary statistics in quantitative research. The statistic is used in the present study to assess the ability of the model, which is primarily derived from Georgia's capital statute, to explain variability in capital charging behavior. The model does not purport to fully describe the

predicative ability of the adjusted model compared to the unadjusted model.¹⁹¹ The R^2 statistic can be viewed as a proxy for the “rational connection between the facts found and the choice made”—that is, it is a measure of validity of the decision-marking.¹⁹² As emphasized earlier,¹⁹³ validity (i.e., R^2) and reliability (i.e., ICC₁, ICC₂, and MAD) statistics capture distinct aspects of capital charging. A reliability analysis concerns the consistency/consensus of capital charging, and therefore relates to the interchangeability of prosecutors (and circuits). A validity analysis, on the other hand, addresses the accuracy/truthfulness of capital charging. Consistency in capital charging does not necessarily imply accuracy; however, inconsistency in capital charging will unavoidably undermine its accuracy.

Recall that the adjusted model takes into account the 40 case-level variables, whereas the unadjusted model does not include any predictors. The overall R^2 statistic for the adjusted model is .40. Unlike traditional regression models, however, MLMs also allow the R^2 statistic to be disaggregated into case-level (“within”) and circuit-level (“between”) components.¹⁹⁴ The R^2 statistics for case-level and judicial circuit-level variance components are 0.44 and 0.21, respectively, for the adjusted model. In other words, the covariates in the model explain, approximately, 44% of the variability within circuits and 21% of the variability across circuits.¹⁹⁵ Location in a particular circuit accounts for approximately 19% of the variability in

process under investigation, rather it is used to explore whether death penalty charging decisions may be deemed sufficiently rational and predictable.

¹⁹¹ See Methodological Appendix.

¹⁹² See *Motor Vehicle Mfrs. Ass’n, Inc. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983) (defining arbitrary and capricious action in the context of administrative law); accord *Burlington Truck Lines, Inc. v. United States*, 371 U.S. 156, 168 (1962).

¹⁹³ See *supra* Part I.B.

¹⁹⁴ RAUDENBUSH & BRYK, *supra* note 112, at 79–85.

¹⁹⁵ Location in a particular jurisdiction accounts for approximately 19% of the variability in death charging, see *supra*, so approximately 4% (.21 x .19 = .039) of the total variability can be explained by inter-jurisdictional differences in case-level factors.

death charging, so approximately 4% ($.21 \times .19 = .039$) of the total variability in death noticing can be explained by inter-circuit differences in case-level factors. This disaggregation of the variability in death noticing decisions into within- and between-circuit components reveals that case-level explanatory variables do a slightly better job of explaining within-circuit variability than a model that ignored the hierarchical structure of the data would suggest (44% versus 41%).¹⁹⁶

In addition to the overall explanatory power of the model based on legally relevant case-level characteristics, another indicator of the rationality of capital charging is the lack of association between legally illegitimate considerations and the predictability of charging decisions.¹⁹⁷ Specifically, the explanatory power of the case-level characteristics should not vary according to defendants' and victims' race/ethnicity. I examined the R^2 statistics for four sub-models: Caucasian defendant, non-Caucasian defendant, Caucasian victim, non-Caucasian victim.¹⁹⁸ The R^2 for the entire sample was 43%; whereas the R^2 was 59% for Caucasian defendant cases, 42% for non-Caucasian defendant cases, 51% for the Caucasian-victim cases, 37% for non-Caucasian victim cases.¹⁹⁹

I examined a "trimmed" model that only includes the number statutorily defined aggravating circumstances, defendant's criminal history, number of victims and defendants,

¹⁹⁶ RAUDENBUSH & BRYK, *supra* note 112, at 109–10 ("The estimates of the proportion of variance explained from a hierarchical analysis may be quite different from those generated in conventional level-1 or level-2 analysis and may lead to different conclusions.").

¹⁹⁷ See *supra* Part II.B.3; ALAN AGRESTI & BARBARA FINLAY, *STATISTICAL METHODS FOR THE SOCIAL SCIENCES* 409 (3d ed. 1997) (discussing comparison of models based on the coefficient of determination).

¹⁹⁸ In order to compare R^2 across sub-groups, the statistic had to be re-scaled in order to take into account the different total variance of the intercept only model. JOOP J. HOX, *MULTILEVEL ANALYSIS: TECHNIQUES AND APPLICATIONS* (2d ed. 2010).

¹⁹⁹ There is, admittedly, significant overlap between Caucasian-defendant and Caucasian-victim because the vast majority of those cases are intra-racial, so the higher R^2 statistics in those sub-models are likely a function of both of those factors.

and the relationship between the defendant and victim. This model captures the “legal core” of the case that should, theoretically, drive capital charging.²⁰⁰ The trimmed model explains 29% of the overall variance, 31% of the variance within jurisdictions and 12% of the variance between jurisdictions. Stated differently, legally legitimate case factors that purportedly guide discretion in the capital punishment process account for less than one-third of the variation in prosecutors’ charging decisions.

Tjur’s “D” Statistic. The R^2 statistic may not be especially intuitive when analyzing models with a dichotomous variable, as is the case in these analyses. By construction, the total variance, and how much of that variance is explained by the model, is based on an underlying continuous latent variable,²⁰¹ so the R^2 statistic relates to a transformation of the dependent variable rather than the actual variable.²⁰² A more interpretable measure for assessing the validity of capital charging decisions, Tjur’s D , was also calculated for the full model, the trimmed model, and the different racial/ethnic subgroups.²⁰³ Tjur’s D , also called the “coefficient of discrimination,” compares the predicted probability of observing an outcome when the outcome is actually observed to the predicted probability of observing an outcome when the outcome is not observed: $D = \hat{\mathbb{P}}(y = 1 | y = 1) - \hat{\mathbb{P}}(y = 1 | y = 0)$, where $\mathbb{P}$ is the probability operator, the first term of the right-hand side of the equation is defined as the true positive rate, and second term is the false

²⁰⁰ Donald Black, *Social Control as a Dependent Variable*, in 1 TOWARD A GENERAL THEORY OF SOCIAL CONTROL: FUNDAMENTALS 1, 20 (Donald J. Black ed., Studies on Law and Social Control, 1984) (describing the legal core of a case as “the rules in the face of the evidence that can be meaningfully analyzed in the jurisprudential tradition”).

²⁰¹ For a detailed description of the R^2 statistic, see Methodological Appendix

²⁰² The transformation is the natural logarithm of the odds of observing the event (i.e., formal charging of a death notice). See Methodological Appendix.

²⁰³ Tue Tjur, *Coefficients of Determination in Logistic Regression Models—a New Proposal: The Coefficient of Discrimination*, 63 AM. STAT. 366, 369 (2009).

positive rate.²⁰⁴ Tjur's D has a range from 0 to 1 (or 0% to 100%). The larger the Tjur's D statistic, the more rational the decision-making process is because the predicted probability of a positive outcome will increase for cases with a positive outcome and decrease for cases with a negative outcome.²⁰⁵ In other words, the rationality of the decision-making process is directly proportional to its ability to minimize both false positive and false negatives. Tjur's D for the entire sample was 39% for the full model and 30% for the trimmed model.²⁰⁶ Unfortunately, Tjur's D cannot be disaggregated into within and between-circuit components like R^2 . Tjur's D was 42% for Caucasian defendant cases, 38% for non-Caucasian defendant cases, 43% for the Caucasian-victim cases, 29% for non-Caucasian victim cases.

These results suggest that cases involving Caucasian defendants and victims are handled more rationally, and therefore less arbitrary when compared to the cases involving non-Caucasian defendants and victims or the entire population of cases. According to the R^2 measure, the difference in validity

²⁰⁴ Tjur, *supra* note 203.

²⁰⁵ *Id.*

²⁰⁶ The accuracy of a model can also be assessed by the improvement it provides in classifying the cases over simple chance. JACOB COHEN ET AL., APPLIED MULTIPLE REGRESSION/CORRELATION ANALYSIS FOR THE BEHAVIORAL SCIENCES 516 (3d ed. 2003). In the Georgia data, prosecutors sought the death penalty in 33% of death-eligible cases. If one were to predict the likelihood that a case would receive the death penalty without knowing anything else about the case, one could classify all of the cases as not being noticed for the death penalty and be correct 67% of the time. The usefulness of the model in explaining capital charging, then, could be measured assessing the improvement in classification when the explanatory variables are included in the model. The full model predicts the correct response approximately 82% of the time, so the model improves classification over pure chance by 15 percentage points. The trimmed model improves classification by 5.4 percentage points.

The major drawback of this model is that a threshold must be chosen, *ex ante*, in order to classify cases. The default cut-off value is .5, so cases with a predicted probability greater than or equal to .5 will be classified as positives. As a consequence, the predictive power of the model greatly depends on the cut-off value the analyst chooses. The R^2 and Tjur's D statistics avoid this problem.

of capital charging decisions based on the defendant's race/ethnicity and victim's race/ethnicity was, respectively, 17 and 14 percentage points. Calculations based on Tjur's *D* yielded similar results: a 5 percentage difference based on the defendants' race/ethnicity and 14 percentage point difference according to victims' race/ethnicity.

Some scholars have challenged using the predictability of death penalty charging and sentencing decisions as a measure of the arbitrariness of the capital punishment system, noting the unpredictability may not only stem from capricious behavior, but also imperfections in the data and underlying model.²⁰⁷ While it is true that the criminal justice system may not lend itself to highly accurate statistical modeling, the Court's heightened reliability requirement under its capital punishment jurisprudence demands a higher standard that would be typically expected of non-capital criminal justice decision-making.²⁰⁸ Moreover, dozens of methodologically rigorous studies of capital charging-and-sentencing decisions employing different model specifications and examining different time periods and jurisdictions have been unable to predict the discretionary choices of prosecutors and jurors with much accuracy.²⁰⁹ As Professor Berk and colleagues have explained, "It is difficult to imagine that a few covariates exist that if included as predictors would lead to clear and justified distinctions between defendants who are charged with a capital crime and defendants who are not...if idiosyncrasies associated with the case, the defendant, or the adjudication process seem to determine a substantial part of the outcome, the adjudication

²⁰⁷ *E.g.*, Stephen P. Klein et al., *Race and the Decision to Seek the Death Penalty in Federal Cases* (RAND 2006).

²⁰⁸ *See generally* Woodson v. North Carolina, 428 U.S. 280, 305 (1976) ("Because of that qualitative difference [between the death penalty and other severe punishments], there is a corresponding difference in the need for reliability in the determination that death is the appropriate punishment in a specific case."); Thompson v. Oklahoma, 487 U.S. 815, 856 (1988) ("Under the Eighth Amendment, the death penalty has been treated differently from all other punishments.").

²⁰⁹ Berk et al., *supra* note 85.

process is suspect.”²¹⁰ Irrespective of the shortcomings inherent in data and statistical models, when the statutorily defined culpability factors predict capital charging decisions only slightly better than chance alone,²¹¹ then the death penalty cannot be functioning in a rational manner.

4. Arbitrariness as Disproportionality

In their concurring opinions in *Furman*, Justices Douglas, Marshall, and Stewart all expressed concern that the arbitrary administration of the death penalty also invited discrimination based on race/ethnicity and poverty.²¹² Elsewhere, the Court has defined arbitrary and capricious government action as the reliance on factors that legislative bodies have not intended the government actor to consider, or the failure of the government actor to consider an important aspect of the problem.”²¹³ My prior analyses have revealed that race/ethnicity, of both the defendant and the victim, is related to inconsistency and irrationality in capital charging decisions. Those analyses did not estimate the direct impact of race/ethnicity on capital charging decisions—they examined whether the degree of inconsistency and irrationality in charging practices varied by race/ethnicity.

I conducted additional analysis to determine whether race/ethnicity had a direct effect on the likelihood that a defendant received a formal death notice, as well as whether this impact varied across circuits. When the race/ethnicity of

²¹⁰ *Id.* at 387.

²¹¹ See *supra* note 206 and accompanying text.

²¹² See *supra* Part I. Justices Brennan, Powell, and Stewart also hinted that an equal protection challenge might prevail considering the racial subtext of the cases. The defendants in the consolidated cases comprising *Furman*—William Furman (Georgia), Lucius Jackson (Georgia), and Elmer Branch (Texas)—were all African American. Moreover, Furman’s homicide appeared to be accidental and Jackson and Branch were sentenced to death for non-homicidal rape.

²¹³ *Motor Vehicle Mfrs. Ass’n, Inc. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

the defendant or the victim has a direct impact on the probability of a defendant being formally charged with the death penalty, then the system produces racially/ethnically disproportionate outcomes. In other words, a system that either imposes a penalty or confers a benefit based upon membership (as a defendant or victim) in a particular racial/ethnic group, then by definition, that system is producing (potential) punishments that are not “graduated and proportioned to the offense.”²¹⁴ And, as emphasized earlier, it is unnecessary to posit that racial/ethnic disparities are motivated by intentional bias because compelling evidence of the effects of unconscious/implicit bias in capital decision-making may still support a constitutional challenge under the Eighth Amendment.²¹⁵ In other words, a racial disparities claim under the Eighth Amendment’s Cruel and Unusual Punishment Clause is doctrinally and analytical distinct from a racial discrimination claim under the Fourteenth Amendment’s Equal Protection Clause.²¹⁶ My focus, here, is on the Eighth Amendment framework and the type of evidence that is illustrative of a violation of that specific constitutional provision.

The results indicate that both the race/ethnicity of the defendant and victim influence the probability that a defendant is charged with the death penalty. At minimum, this suggests that there is a high risk of arbitrariness irrespective of discriminatory intent. The probability of a death notice being filed against a defendant increased by 7 percentage points if the defendant was Caucasian and 15 percentage points if the victim was Caucasian, all else equal. The race-of-defendant effect should be interpreted with caution, however, because the overwhelming majority of homicides involving Caucasian defendants are intra-racial (92%); whereas homicides committed by non-Caucasian defendants are less racially-homogeneous (71%). Forty-five percent of death-eligible cases involved Caucasian victims, but 65% of cases actually noticed for the death penalty involved Caucasian victims. Similarly,

²¹⁴ See *supra* Part I.A

²¹⁵ See *supra* Part II.A.

²¹⁶ *Id.*

Caucasian defendants comprised 25% of death eligible cases, but 37% of cases receiving a death notice. These results are consistent with prior research—cases involving Caucasian victims are much more likely to be charged with the death penalty, all else equal.²¹⁷

It is not clear whether these results demonstrate intentional racial/ethnic bias because, by definition, the statistical models represent simplifications of the underlying data-generating process and there may be unobservable factors that account for the observed racial/ethnic differences in capital charging.²¹⁸ It is worth noting that nearly every methodologically rigorous examination of the death penalty has uncovered similar results: cases involving white victims are the most likely to be noticed for the death penalty and sentenced to death, all else equal.²¹⁹ And it is especially interesting (and informative) that methodologically rigorous studies are more likely to discover racial/ethnic disparities than studies that are less methodologically rigorous in the very same jurisdictions.²²⁰ However, as explained, *supra*, it suffices that these results strongly suggest that charging behavior is arbitrary via its relationship to a legally arbitrary factor: race/ethnicity.²²¹

An important advantage of MLMs is the ability to examine the variability in the effects of race/ethnicity—for both the defendant and the victim—across judicial circuits. Prior research on capital charging decisions that was attentive to jurisdictional-level variability was limited by models that

²¹⁷ See, e.g., Baldus & Woodworth, *Race Discrimination in the Administration of the Death Penalty*, *supra* note 140.

²¹⁸ George E.P. Box, *Science and Statistics*, 71 J. AM. STAT. ASSOC. 791, 792 (1976) (explaining that statistical models are, invariably, under-inclusive).

²¹⁹ See, e.g., Carol S. Steiker & Jordan M. Steiker, *The American Death Penalty and the (In)Visibility of Race*, 82 U. CHI. L. REV. 243 (2015) (summarizing the empirical literature on race and capital punishment); Baldus & Woodworth, *Race Discrimination in the Administration of the Death Penalty*, *supra* note 140.

²²⁰ David C. Baldus et al., *Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia*, 83 CORNELL L. REV. 1638, 1661–62 (1998).

²²¹ See *supra* Part I.A.

precluded the empirical examination of heterogeneous case-level effects, so this question has not been adequately explored in the literature. As previously noted, evidence of racial/ethnic disparities as proof of arbitrariness was presented to the court in *McCleskey* and rejected as insufficient.²²² What evidence, then, would be sufficiently compelling to deem to capital charging unconstitutionally arbitrary because of pronounced racial/ethnic disparities? Simple answers to this question are elusive, but exploring variability in the effect of defendants' and victims' race/ethnicity on capital charging decisions across the state may be particularly illuminating because it is capable of providing a statewide baseline against which all jurisdictions can be assessed.²²³

Figures 7 and 8 display the variable effects of the race/ethnicity of the defendant across circuits. Figures 9 and 10 provide similar information for race-of-victim effects. For Figures 8 and 10, the vertical axis is the magnitude of the effect and the horizontal axis lists the circuit. These figures reveal that the influence of these aforementioned case-level characteristics can vary considerably across jurisdictions, indicating significant inconsistency in the manner in which prosecutors treat race/ethnicity in their charging decisions, and therefore a strong indication of disproportionate, and therefore arbitrary, decision-making. In concrete terms, the race-of-defendant effect ranged from -.12 to .31 (the statewide average was .07). This means that the probability that case involving a Caucasian defendant receives a death notice, relative to a case with a non-Caucasian defendant, ranges from 12 percentage points lower or 31 percentage points higher, all else equal. Figure 7 displays the spatial distribution of this effect on a map of Georgia. Limiting the analysis to circuits that filed at least ten death notices produces a range from -.08 to .31, and .16 to .29 for circuits with at least 20 death notices filed.

The race-of-victim effect ranged from -.07 to .41 (the

²²² See, *supra*, note 66 and accompanying text.

²²³ Kyle Graham, *Overcharging*, 11 OHIO ST. J. CRIM. L. 7, 710–11 (2014) (noting the difficulty of defining a widely accepted baseline by which to compare prosecutors' charging decisions).

statewide average was .15), with a similar interpretation as the aforementioned race-of-defendant effect. Figure 9 displays the spatial distribution and Figure 10 reports the race-of-victim effect for each circuit. The range was -.07 to .41 for circuits with at least ten death notices filed and .16 to .40 for circuits filing at least 20 death notices. Clearly, circuits with race/ethnicity effects that are markedly different from the statewide average create cause for concern about arbitrariness.²²⁴

5. Summary

My analyses of the Georgia data provide considerable evidence of an arbitrary capital charging process. The inability of legally legitimate case characteristics to meaningfully improve consistency, rationality, and racial/ethnic proportionality—along with the strong association of race/ethnicity with inconsistency and irrationality—should be particularly alarming considering Georgia’s death penalty system has been subject to serious scrutiny for more than forty years. The information obtained about Georgia’s capital charging process is extremely valuable in its own right, but its usefulness extends beyond extracting patterns of behavior for past decisions—the model can also assist us in making assessments about future potentially capital cases. The next section describes exactly how this can be done.

²²⁴ The direct race/ethnicity-effect estimates of arbitrariness (via disproportionality) reported in this section should be viewed as conservative because results obtained from circuits with only a small number of cases were weighted towards the statewide average. These “shrunk” estimates (also called “partially pooled” or empirical Bayes estimates) exhibit less overall variability than what would be obtained using completely un-pooled data, but permit more reasonable inferences from the data. The partially pooled estimate, $\tilde{\beta}_k$, is calculated as: $[1/(1 + \lambda_j)] \times \hat{\beta}_k$; where $\lambda_j = \psi_j/[\psi_j + (\theta/n_j)]$, where n_j is the sample size of judicial circuit, j , ψ is the between-circuit variance, and θ is the variance within circuits. RAUDENBUSH & BRYK, *supra* note 112, at 46 (discussing advantages of empirical Bayes estimation for hierarchical data). The completely un-pooled estimates for race-of-defendant range from -.21 to .41, whereas the shrinkage estimates range from -.12 to .31. For the race-of-victim estimates, the un-pooled estimates range is -.24 to .48 compared to the shrinkage estimate range of -.07 to .41.

C. Model Predictions for Case-Specific Outcomes

The parameter estimates obtained from the empirical model described in the previous section can be used to predict unobserved (yet observable) case outcomes.²²⁵ After analyzing the death penalty data within the MLM framework, one can obtain educated guesses about a defendant's probability of receiving a death notice even though the defendant was not included in the previously analyzed data.²²⁶ One of the key advantages of prediction in the framework I have adopted is that it also permits contextually dependent predictions. In other words, the analyst can make counterfactual predictions based on the location of the actual case, other locations throughout the state, and predictions based on the statewide baseline.²²⁷ These predictions can provide greater understanding of the relative culpability of the case at issue, vis-à-vis all other death-eligible cases previously processed in the system.

By way of illustration, assume a prosecutor is considering pursuing a capital charge against a defendant.²²⁸ A prediction of the probability the case would receive a formal death notice can be obtained by inserting the values of the variables specified in the model.²²⁹ Imagine the prosecutor presents a case

²²⁵ The terms *prediction* and *forecast* are often used interchangeably in the statistics literature. Technically speaking, forecasts only involve out-of-sample predictions, whereas predictions may include either in-sample or out-of-sample predictions. Gary King & Langche Zeng, *The Dangers of Extreme Counterfactuals*, 14 POL. ANAL. 131 (2006). For the purposes of this discussion, I will use the term *prediction* to refer to both in-sample and out-of-sample predictions.

²²⁶ See, e.g., David Afshartous & Jan de Leeuw, *Prediction in Multilevel Models*, 30 J. EDUC. & BEHAV. STAT. 109 (2005).

²²⁷ *Id.*

²²⁸ A detailed discussion of the statistical theory underlying prediction in multilevel models is beyond the scope of this paper, but thorough descriptions are readily available. See *id.*

²²⁹ Prediction uncertainty can also be quantified, so both the expected value and a range in which the expected value may fall can be calculated. Gary King et al., *Making the Most of Statistical Analyses: Improving Interpretation*

representative of a “typical” death-eligible case in the state based on the data: the defendant is a African American male, 27 years old, two contemporaneous felony charges (in addition to the murder charge), one prior felony conviction, used a firearm, monetary motive for the killing, non-gang related, has at least one child, employed at the time of the killing, unmarried, without a high school diploma or equivalent, no military service, history of alcohol and drug use, 1.3 (out of 5) on the troubled family background index, an IQ of 89, no or minimal psychiatric impairment, and resided in the same county where the killing occurred.²³⁰ The sole victim was an African American male, 37 years old, and unacquainted with the defendant prior to the murder. The prosecutor also alleges two statutorily-defined aggravating circumstances. Based on the statistical model, there is a 33 percent chance that a defendant would be noticed for the death penalty.

It is important to keep in mind this prediction is also based on the defendant being tried in the “typical” jurisdiction in the state. When we take into account differences in death-noticing behavior for similar cases across jurisdictions in the state, the expected likelihood of a death notice for this defendant may be as large as 59 percent or a low as 12 percent.²³¹ Figures 4 and 5

and Presentation, 44 AM. J. POL. SCI. 341, 350 (2000) (discussing sources of uncertainty in data simulation).

²³⁰ The consideration of factors such as a defendant’s marital status, number of children, employment status, etc., may be legally relevant for the purposes of the defendant’s mitigation case, so it is important that these factors are expressly modeled. See generally Stephen P. Garvey, *Aggravation and Mitigation in Capital Cases: What Do Jurors Think?*, 98 COLUM. L. REV. 1538 (1998) (discussing jurors’ assessment of mitigation evidence in capital trials); Robert J. Smith et al., *The Failure of Mitigation?*, 65 HASTINGS L.J. 1221 (2014).

²³¹ Prediction using both hierarchical and non-hierarchical data structures depends on the proper estimation of the model parameters. In the case of multilevel models, the fixed effects (β) and variance components (ψ and θ) are assumed to be “true” when calculating expected outcomes. Admittedly, all statistical models are simplifications of much more complex dynamics and, therefore, no model is ever true. Box, *supra* note 218, at 792 (“Since all models are wrong, the scientist cannot obtain a ‘correct’ one by excessive elaboration.”). Nonetheless, the closer the proposed model is to the true model, the more accurate the parameter estimates and, by extension,

depict the jurisdictional variability in probability of receiving a death notice based on the aforementioned case characteristics. This predicted probability provides a rough sense of the death-worthiness of a case based on prosecutors' patterns of actual charging decisions throughout the state. To be clear, this information cannot indicate whether the case *should* be authorized for the death penalty; it merely identifies how we would expect the case to be handled based on prior death penalty noticing activity.

Now imagine a prosecutor presenting a similar case that differs only with respect to the number of statutory aggravating circumstances present in the case, which is now four. In this case, the predicted probability that the same defendant is noticed for the death penalty increases to 57 percent.²³² Again, this is the expected probability of a death notice in the typical jurisdiction, and this expected value would range from 27 percent to 81 percent depending on the jurisdiction. Averaging over jurisdictions, the difference in the expected probabilities of receiving a death notice is nearly 27 percentage points, but the difference could be as large as 54 points (99 versus 33) depending on the judicial circuit.

Keeping with this example, now suppose that two cases are factual similar, except for the race of the victim. The model reveals that a case involving a Caucasian victim has an expected probability of receiving a death notice that is 15 percentage points greater than a case with a non-Caucasian victim. Assuming that the 15 percentage point difference in the

the more plausible the predictions derived from those estimates. The number of cases in a group (n_j), as well as the number of groups (J), may profoundly impact the quality of the parameter estimates, so one must be careful when drawing inferences from the statistical model. Afshartous & de Leeuw, *supra* note 226.

²³² This example assumes a linear effect of the number of statutory aggravating factors on the probability that a case receives a death notice. It is possible that the actual effect is nonlinear, in which case the predicted probability could be larger or smaller than 57 percent. *See, e.g.*, SIMON N. WOOD, GENERALIZED ADDITIVE MODELS: AN INTRODUCTION WITH R 316–24 (2006) (describing multilevel models that relax the linearity assumption for the effects of predictor variables).

expected probability is not attributable to any other legitimate case characteristics not captured in the statistical model, this evidence may assist in the assessment of the defendant's culpability—especially in borderline cases. Returning to our example of the typical case in Georgia, which has an African American victim, the expected probability of 33% of receiving a death notice (i.e., one out of three cases). A factually identical case, with the exception of having a Caucasian victim, would have an expected probability of 48% of being noticed for death.

This admittedly simplified example assumes that the race-of-victim effect is constant throughout the state, but as explained, *supra*, the data reveal this assumption is unwarranted, and the race-of-victim may be stronger or weaker depending on the jurisdiction. In some jurisdictions the race-of-victim effect may be over twice as strong,²³³ while in a handful of others white-victim cases are less likely to be noticed for death.²³⁴ In jurisdictions where the race-of-victim effect is significantly larger, in absolute value, than the statewide average, an inference of an unacceptably high risk of arbitrariness in death charging in that particular jurisdiction may be justified.

The real world, of course, is more complicated because cases typically differ along multiple important dimensions. A crucial shortcoming of existing practices is the inability of appellate courts to develop general culpability measures to identify comparable cases irrespective of factual differences.²³⁵ The calculation of the expected probability of receiving a death notice becomes extremely useful, perhaps essential, when analysts cannot identify a sufficient number of similarly situated defendants.²³⁶ This expected probability constitutes a “charging

²³³ The effect of the race-of-victim coefficient ranges from -.07 to .40.

²³⁴ The race-of-victim effect resulting in a lower probability that a defendant was charged with the death penalty Caucasian-victim cases was only present in jurisdictions that sought the death penalty on fewer than ten occasions.

²³⁵ See, *infra*, Part III.

²³⁶ An additional advantage of this approach is that the systematic collection of data and uniform coding of variables will greatly assist courts in

propensity score” and cases can be organized into various groupings/tiers based on similar propensity scores.²³⁷ The predictions can be “normalized” with respect to race/ethnicity and geography, by either adding or subtracting the influence of those factors from the prediction.²³⁸ This approach permits comparisons based upon the similarity predicted scores, rather than only focusing on cases that have similar facts.²³⁹

D. Generalizability of the Findings and Exportability of the Model

1. Generalizability

There is strong evidence of arbitrariness in capital charging decisions in Georgia during the time period investigated in this study. Admittedly, the model provides a simplification of a more complex decision-making process, but consistent with prior research, it captures the essential features of the process—i.e., relevant aggravation and mitigation evidence. The current analysis focused solely on eight years of capital charging data in Georgia, but there are two compelling reasons to believe that the

identifying cases that are factual similar among relevant dimensions. Cases need not necessarily have the exact same values on these key variables, and the analyst can identify a range of permissible values that would satisfy the query. *See, e.g.,* Stefano M. Iacus et al., *Causal Inference without Balance Checking: Coarsened Exact Matching*, POL. ANAL. 1 (2012) (describing a method for identifying similarly situated units in social science research).

²³⁷ Paul R. Rosenbaum & Donald B. Rubin, *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, 70 BIOMETRIKA 41 (1983) (defining a propensity score as the conditional probability of assignment into a group given a set of observed covariates).

²³⁸ The choice to add or subtract the influence of race/ethnicity would depend on which group is chosen as the baseline.

²³⁹ To improve the accuracy of the model predictions, it is also important that prior cases that were deemed excessive on appeal and reversed must be excluded from the analysis. BALDUS ET AL., *supra* note 49, at 282 (noting flaws in proportionality review systems); Rankin et al., *supra* note 131 (discovering the Georgia Supreme Court’s routine use of overturned cases when conducting its mandatory proportionality review).

findings may be generalizable to other states. First, following the Court's approval of Georgia's capital punishment statute in *Gregg*, many states developed statutes that were very similar to Georgia's statute. As noted, *supra*, Georgia's statute was closely modeled after the ALI's model death penalty statute.²⁴⁰ The demonstrated failure of Georgia's statute to eliminate or sufficiently reduce arbitrariness and bias in capital charging seriously brings into the question the ability of similarly structured statutes to accomplish that goal. Second, the empirical results obtained in the current study are very similar to the findings reported in other states. For example, the marginal effect of the race-of-victim on the probability of a death notice was 15 percentage points. Recent studies of from Connecticut and Maryland report similar marginal effects—respectively, 20 and 13 percentage points.²⁴¹ This provides some evidence that similar dynamics are occurring across jurisdictions.

2. Exportability

The usefulness of the diagnostic tool developed in this Article extends beyond its applicability to Georgia. This is important because developing legally cognizable claims about arbitrariness in capital charging will require empirically grounded research that is reproducible.²⁴² The common structure of capital statutes, along with the Court's governing capital punishment jurisprudence, allows the model to be applied in all capital jurisdictions with only minor modifications. The key doctrinally and empirically relevant

²⁴⁰ See *supra* Part I.A

²⁴¹ Donohue, *supra* note 106; Paternoster & Brame, *supra* note 144.

²⁴² David Zuckerman, Building a Capital Arbitrariness Claim from the Ground up: A "How To" Primer Based on the Pennsylvania Experience, presented at Capital Punishment Training Conference 1 (Airlie, Va. Aug. 1997) (emphasizing the need for social science research in developing arbitrariness challenges to capital punishment systems); Smith, *supra* note 99 (positing that the systematic collection and analysis of data showing regional variation in capital charging is essential for challenges to capital punishment systems).

factors influencing capital charging-and-sentencing decisions have been well-documented. Prosecutors, themselves, routinely identify the very same factors that researchers include in their predictive models.²⁴³ The model I develop in this Article is capable of incorporating additional legally relevant (as well as empirically relevant) factors necessary to properly understanding capital charging processes. All current capital statutes enumerate aggravating circumstances; some expressly identify mitigating factors while others do not.²⁴⁴ The presence or absence of statutorily defined mitigating factors was made largely irrelevant by the Court's ruling in *Lockett v. Ohio* because courts are not allowed to place many restrictions on the types of mitigation evidence defendants can present at trial.²⁴⁵

Identifying jurisdictional variability in the expected probability of receiving a death notice and the influence of case-level factors is straightforward in my model. Death charging decisions are made at the county-level (or similar sub-state unit level) and information about the locality of capital prosecutions is readily available. My model did not attempt to explain variability across the judicial circuits in Georgia using circuit-level factors (e.g., social and economic variables), but these relationships have been explored elsewhere.²⁴⁶ I did discover, however, that case-level variables explain only a very small proportion of the between-circuit variability in capital charging.

The use of the model is dependent on sufficient data from which to draw reasonable inferences about capital charging practices. While it is true that data will be more available in

²⁴³ See generally Ehrhard, *supra* note 105 (discussing factors that prosecutors list as relevant and irrelevant to their capital charging decisions); Ehrhard-Dietzel, *supra* note 105 (same); Paternoster & Brame, *supra* note 144 (same).

²⁴⁴ Jeffrey L. Kirchmeier, *Casting a Wider Net: Another Decade of Legislative Expansion of the Death Penalty in the United States*, 34 PEPP. L. REV. 1 (2006) (describing capital statutes in every state).

²⁴⁵ See *supra* Part I.

²⁴⁶ Sherod Thaxton, *The Social Geometry of Death: Social Structure and Capital Punishment in Georgia, 1993-2000* (2009) (unpublished Ph.D. dissertation, Emory University).

some states than in others, data limitations need not be an insurmountable obstacle. Preliminary results, based on models lacking sufficiently detailed case-level information, may still be informative. To the extent that these models provide tentative evidence of arbitrariness, a strong claim can be made to local and state governments to systematize the collection of relevant data, or at least impose a temporary moratorium on capital charging until their death penalty systems can be explored in greater depth.²⁴⁷

E. Summary

The utility of social scientific methods and evidence in both understanding the administration of capital punishment and remedying persistent problems is beyond serious dispute in the scholarly community,²⁴⁸ yet many courts and legislatures remain resistant to fully translating this body of knowledge into meaningful death penalty doctrines and reform efforts.²⁴⁹ In order to gain traction with courts and legislatures, social/behavioral science models of death penalty decision-making must be packaged in such a way that emphasizes value

²⁴⁷ See, e.g., Charles S. Lanier & James R. Acker, *Capital Punishment, the Moratorium Movement, and Empirical Questions: Looking beyond Innocence, Race, and Bad Lawyering in Death Penalty Cases*, 10 PSYCHOL. PUB. POL. & L. 577 (2004) (discussing death penalty moratoriums across the nation in response to concerns over the improper administration of capital punishment); Jeffrey L. Kirchmeier, "Another Place Beyond Here": *The Death Penalty Moratorium Movement in the United States*, 73 U. COLO. L. REV. 1 (2002) (same).

²⁴⁸ David C. Baldus, *The Death Penalty Dialogue between Law and Social Science*, 70 IND. L.J. 1033 (1995) (discussing the influence of social science on nearly every aspect of the administration of capital punishment); Michael L. Radelet & Marian J. Borg, *The Changing Nature of Death Penalty Debates*, 26 ANN. REV. SOCIOLOGY 43 (2000) (same).

²⁴⁹ See generally James R. Acker, *A Different Agenda: The Supreme Court, Empirical Research Evidence, and Capital Punishment Decisions, 1986-1989*, 27 LAW & SOC'Y REV. 65 (1993) (noting that the U.S. Supreme Court justices were much more likely to discredit social scientific evidence and announce principles detached from the evidence of the actual administration of capital punishment); Baldus et al., *supra* note 122 (documenting legislators' resistance to statistical evidence of racial bias in the administration of the death penalty).

in their adoption, relative simplicity of their implementation, and ease in their justification to the public. In the next section, I briefly describe how my diagnostic model can be used to improve accuracy and consistency in capital charging, and thereby impose greater discipline on the use of the death penalty. I do not attempt to articulate a fully developed proposal, rather my aim is to lay a foundation upon which a more elaborate model can be built. Key to laying this foundation, I argue, are meaningful mechanisms to discourage inadequate charging screening by prosecutors that legislatures are likely to seriously consider.²⁵⁰

III. POLICY IMPLICATIONS

The majority of proposed and enacted death penalty reforms offer the appearance of greater regularity, but do very little to reduce arbitrariness and bias.²⁵¹ Professor James Liebman aptly notes that death penalty reforms must avoid becoming mere “window-dressing” for change that further disserve capital defendants by legitimating a broken system and reducing incentives for adopting subsequent meaningful reforms.²⁵² Combating arbitrariness in front-end charging decisions is an indispensable step in any effort to improve the rationality and transparency of the capital punishment process.²⁵³ Numerous proposals aimed at either explicitly restricting the breadth of prosecutorial discretion or better illuminating and policing

²⁵⁰ Liebman, *supra* note 7 (arguing that most death penalty reforms are unlikely to be adopted by courts or legislatures because they fail to provide adequate incentives).

²⁵¹ Carol S. Steiker & Jordan M. Steiker, *Should Abolitionists Support Legislative Reform of the Death Penalty*, 63 OHIO ST. L. 417 (2002) (describing the tendency of criminal justice officials to adopt “toothless” reforms).

²⁵² Liebman, *supra* note 7; see also Charles J. Ogletree Jr., *Black Man's Burden: Race and the Death Penalty in America*, 81 OR. L. REV. 15 (2002) (arguing that the Court's procedural regulation of the death penalty has further entrenched an inconsistent and racially discriminatory system).

²⁵³ See *supra* Part I.

prosecutors' discretionary choices already exist.²⁵⁴ Detailed descriptions and critiques of these existing proposals can be found elsewhere.²⁵⁵ The purpose of this section is to sketch a foundation for a general framework for death penalty reform that can be both effective in reducing arbitrariness, capriciousness, and excessiveness, but also sufficiently attractive to legislators to have an honest chance at being implemented. My framework draws inspiration from some of these existing proposals; nonetheless, there are multiple important points of departure that may hold the promise of offering unique improvements over these other models.

My key intervention involves front-end proportionality review of death charging heavily guided by social scientific inquiry into arbitrariness and disproportionality. This intervention would likely need to be accompanied by meaningful financial, administrative, and reputational disincentives for foreseeable charging errors because useful reforms must raise the costs of making foreseeable errors in capital charging.²⁵⁶ Current capital charging-and-sentencing

²⁵⁴ See, e.g., Joseph L. Hoffman et al., *Plea Bargaining in the Shadow of Death*, 69 FORDHAM L. REV. 2313 (2000–2001); Liebman, *supra* note 3; F. Thomas Schornhorst, *Preliminary Screening of Prosecutorial Access to Death Qualified Juries: A Missing Constitutional Link*, 62 IND. L.J. 295 (1986–1987). Rachel Barkow, *Institutional Design and the Policing of Prosecutors: Lessons from Administrative Law*, 61 STAN. L. REV. 869 (2009); Oren Bar-Gill & Omri Ben-Shahar, *The Prisoners' (Plea Bargain) Dilemma*, 1 J. LEGAL ANAL. 737 (2009); Stephanos Bibas, *Regulating the Plea-Bargaining Market: From Caveat Emptor to Consumer Protection*, 99 CAL. L. REV. 1117 (2011); Lucian E. Dervan, *The Surprising Lessons from Plea Bargaining in the Shadow of Terror*, 27 GA. ST. U. L. REV. 239 (2010–2011); Oren Gazal-Ayal, *Partial Ban on Plea Bargains*, 27 CARDOZO L. REV. 2295 (2005–2006); Russell M. Gold, *Promoting Democracy in Prosecution*, 86 WASH. L. REV. 69 (2011); Robert L. Misner, *Recasting Prosecutorial Discretion*, 86 J. CRIM. L. & CRIMINOLOGY 717 (1996); William J. Stuntz, *Plea Bargaining and Criminal Law's Disappearing Shadow*, 117 HARV. L. REV. 2548 (2004); Josh Bowers, *Punishing the Innocent*, 156 U. PA. L. REV. 1117 (2008); Roger A. Jr Fairfax, *Delegation of the Criminal Prosecution Function to Private Actors*, 43 U.C. DAVIS L. REV. 411 (2009).

²⁵⁵ See, e.g., Steiker & Steiker, *supra* note 251; Liebman & Clarke, *supra* note 118.

²⁵⁶ Misner, *supra* note 254 (“The current flaw in the evolving power of the prosecutor is the failure to force her to face the full costs of prosecutorial

systems make trial error virtually costless to prosecutors—and to a lesser extent the capital defense bar²⁵⁷—because proper feedback mechanisms from appellate and post-conviction stages to trial actors are nearly non-existent.²⁵⁸ By adequately discouraging poor screening decisions with respect to guilt and punishment, not only can high error correction costs be substantially reduced,²⁵⁹ but also the greater expense associated with pre-trial and trial stages of capital cases. In the interest of space, however, I only focus on the front-end proportionality review component in this section. The necessary details for the second component are fleshed-out elsewhere.²⁶⁰

decisions.”); William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505 (2001) (“Unless the trial system imposes costs for making mistakes, they will make too many. Broad liability rules are a way of evading the adjudication system, and therefore of making mistakes cheaper.”).

²⁵⁷ Due to the fact that the pool of capital charged cases is so large, the capital defense bar tends to neglect capital pre-trial, trial, and direct appellate proceedings, and focuses on procedural issues that are winnable in habeas proceedings rather than the substantive issues that are in play at the earlier stages. Liebman, *supra* note 102, at 2076. This is particularly true because condemned inmates’ access to state-compensated attorneys shrinks, so it becomes easier for outside attorneys to intervene during the post-conviction stages. *Id.*; cf. Emily Garcia Uhrig, *A Case for a Constitutional Right to Counsel in Habeas Corpus*, 60 HASTINGS L.J. 541, 541 (2008–2009) (commenting on the absence of a constitutional guarantee of counsel in post-conviction proceedings and arguing that the liberty interests of condemned inmates improperly sentenced to death require such a guarantee). See also *Ross v. Moffitt*, 417 U.S. 600, 611–12 (1974) (no right to counsel in certiorari proceedings after direct appeal); *Murray v. Giarratano*, 492 U.S. 1, 10 (1989) (no right to counsel in state-post conviction proceedings).

²⁵⁸ Liebman, *supra* note 7, at 324–25.

²⁵⁹ Bowers, *supra* note 254, at 1179 (suggesting that systemic miscarriages of justice are not due to the practice of plea bargaining, but result from overreaching in cases that technically meet the statutory requirements for the highest possible charge, but fall outside of systemic and communal norms); Gazal-Ayal, *supra* note 254, at 2306 (arguing that poor charge screening decisions are responsible for miscarriages of justice, and not plea bargaining itself).

²⁶⁰ Thaxton, *supra* note 30 (describing a system of financial, administrative, and reputational disincentives for poor charge screening in potentially capital cases).

A. *Designing Front-End Proportionality Review*

The feature of post-*Furman* capital statutes approved in *Gregg* holding the most promise for eliminating, or at least significantly reducing, arbitrariness, bias, and disproportionality in the administration of capital punishment was meaningful appellate review of all death sentences imposed at trial.²⁶¹ The *Gregg* court focused much attention on the alleged narrowing function that the newly drafted capital statutes performed, but even at the time of the Court's ruling, the aggravating circumstances enumerated in the capital statutes encompassed a very wide range of murders. Thus from the very beginning, it was highly unlikely that those circumstances, in and of themselves, could sufficiently narrow the death-eligible pool to make the administration of capital punishment less arbitrary or discriminatory. This was particularly true at the time of *Gregg* because the death penalty was still permissible for a wide range of non-homicide offenses (e.g., rape, kidnapping, armed robbery, and arson). If the revised statutes were to have any reformatory force, then their impact would have been primarily—if not exclusively—through appellate review of death sentences.

Critics of the statutes approved by the Court in *Gregg* voiced two key objections. First, they argued that the new laws were incapable of ensuring the constitutionally permissible administration of capital punishment required under *Furman*.²⁶² The statutes merely shifted the “unbridled discretion” of the pre-*Furman* era statutes to other parts of the process—namely, prosecutorial charging, plea-bargaining, and executive clemency.²⁶³ Chief Justice Burger commented during oral argument in *Gregg* that front-end (charging) and back-end (clemency) discretion were inevitable components of any capital scheme and outside of the effective control of legislatures.²⁶⁴ According to the Chief Justice, if the critics' arguments were taken to their logical conclusion, no statute

²⁶¹ See *supra* Part I.A.

²⁶² BANNER, *supra* note 49, at 273.

²⁶³ *Id.*

²⁶⁴ *Id.*

would meet the standards that critics of the current schemes advocated, no matter how narrowly death eligibility was defined. But the Chief Justice appears to have missed the critics' underlying point. The revised statutes were insufficiently expansive to implement the heightened reliability and accuracy standards required under *Furman's* "death is different" logic. Potential abuses of executive branch power needed to be monitored and, when appropriate, remedied by the courts. It was within the legislature's power to craft capital statutes that imposed greater justificatory and evidentiary burdens on prosecutors when pursuing the death penalty, thus impacting the front-end discretionary process. Regulating clemency decisions, admittedly, would be more difficult, but the relative infrequency and heightened transparency of these decisions made the practice less of a concern to these critics.

The second major objection to the revised statutes was Court's unsupported belief that the "heightened procedural regulation approach" to capital punishment would satisfy the Eighth Amendment. Even assuming, *arguendo*, that the revised capital statutes could accomplish what they purported, neither the states defending their statutes nor an independent party provided evidence that their new regimes were non-arbitrary and unbiased. For example, Georgia's statute was closely modeled on "an untested innovation in 1962" by the American Law Institute (ALI).²⁶⁵ Neither the ALI, Georgia, nor the Court had any reliable evidence as to whether the guided-discretion statutes presented before the Court in *Gregg* and its companion cases were capable of eliminating or sufficiently reducing the rampant arbitrariness (and potential bias) in death penalty charging-and-sentencing practices that the Court deemed violated the Eighth Amendment.²⁶⁶ In the nearly forty years

²⁶⁵ American Law Institute, *supra* note 9, at 4.

²⁶⁶ Statistical evidence of racial bias in the administration of the death penalty was presented to the Court in *Furman*, but this evidence did not form the basis for the Court's ruling. Only Justices Douglas and Marshall cited statistical evidence in their opinions. HAINES, *supra* note 49.

Interestingly, the Court, in *Gregg*, also criticized the petitioners' challenges to the revised capital statutes for failing to provide evidence that those schemes did not (or could not) satisfy the constitutional mandate of

since *Gregg*, the Court has deemphasized *Furman*'s strong concerns about actual outcomes of death penalty cases, refused to test whether *Gregg*'s assumption that the guided-discretion statutes would result in accurately and consistently imposed death sentences,²⁶⁷ and ignored social science evidence on the operation of the death penalty.²⁶⁸ Numerous scholars have attributed the Court's reluctance to embracing the social scientific evidence of the realities of capital charging-and-sentencing practices to its lack of expertise in evaluating statistical evidence of arbitrariness, bias, and excessiveness.²⁶⁹ As Professors Carol Steiker and Jordan Steiker note, the Court's avoidance of direct engagement in the statistical evidence provided by petitioners and respondents in capital cases was routine and "many justices may have felt that their personal legitimacy as jurists was threatened in cases involving statistical proof."²⁷⁰ These two criticisms—the absence of constraints on front-end/back-end decision-makers in the revised capital statutes and the Court's unwillingness to acknowledge and respond to applicable social science—foreshadowed problems

Furman. HANEY, *supra* note 31.

²⁶⁷ Smith, *supra* note 99, at 249. It would have been possible for the Court to periodically grant review for a group of cases and determine whether the state's statute was effectively distinguishing the worst-of-the-worst cases from those cases that were technically death-eligible but, nonetheless, received a sentence less than death. *Id.*

²⁶⁸ *McCleskey v. Kemp* clearly highlighted the fact that the promise of *Gregg* had not been fulfilled—at least in Georgia. BALDUS ET AL., *supra* note 49.

²⁶⁹ Steiker & Steiker, *supra* note 219, at 277.

²⁷⁰ *Id.* at 282. According to Professor Scott Sundby, Justice Powell's aversion to quantitative reasoning was not evident across all cases. In a series of antitrust opinions, Justice Powell employed sophisticated economic reasoning guided by mathematical models. Scott E. Sundby, *The Loss of Constitutional Faith: McCleskey v. Kemp and the Dark Side of Procedure*, 10 OHIO ST. J. CRIM. L. 5, 13 (2012). Although econometric (i.e., statistical) analyses differs from economic reasoning, Professor Sundby noted that "one might wonder whether Powell's comfort in employing sophisticated economic analysis reflects that a person's 'numberphobia' to some extent tracks one's ideological priors. Complex theories look helpful and clear when they lead to conclusions that are congenial with our views, but appear confusing and incomplete when leading to conclusions less favorable to our predispositions." *Id.*; See generally THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* 97–98 (2d ed. 1970) (explaining that disconfirming facts vary in their importance depending on the point in which the data are observed).

for the next forty-plus years.

But the appellate court's review could, in theory, correct errors of inadequate charge screening by identifying factors in the cases it reviewed that warranted a punishment less than the death penalty, irrespective of the defendant's eligibility under the governing statute. The appellate court was not limited to reviewing cases for trial error, so it was free to engage in a more comprehensive assessment of the totality of circumstances in each case. General standards by which defendants and criminals could be assessed was a necessary component to guide this evaluation. Comparative proportionality review, which entailed a systematic inquiry into similar and dissimilar cases, appeared to provide the vehicle through which these culpability assessments could be made. Arbitrariness and bias, on a systemic level, could be reduced by rigorous (dis)proportionality assessments at the individual level. Individual punishments that were appropriately calibrated based on the disciplined consideration of legitimate defendant and crime factors could increase overall consistency and rationality.

The ability of comparative proportionality review to accomplish this daunting, yet doable, task was cut short by the Court a mere eight years after *Gregg* in *Pulley v. Harris*²⁷¹—a case that has been viewed by many analysts as especially counterproductive to the stated goals articulated in *Furman* and *Gregg*.²⁷² In *Pulley*, the constitutionality of California's death penalty statute was challenged, in part, because the California Supreme Court refused to conduct comparative review of the defendant's case with sentences imposed in similar capital cases to determine whether the defendant's death sentence was

²⁷¹ 465 U.S. 37 (1984).

²⁷² See, e.g., Timothy V. Kaufman-Osborn, *Capital Punishment, Proportionality Review, and Claims of Fairness (with Lessons from Washington State)*, 79 WASH. L. REV. 775 (2004) (discussing the debate surrounding the constitutionality and effectiveness of comparative proportionality review); Barry Latzer, *The Failure of Comparative Proportionality Review of Capital Cases (with Lessons from New Jersey)*, 64 ALB. L. REV. 1161 (2001) (same).

proportionate.²⁷³ The Ninth Circuit Court of Appeals had previously ruled that comparative proportionality review was constitutionality required,²⁷⁴ but the Court reversed and held that comparative proportionality review was not an indispensable feature of constitutional death penalty statutes.²⁷⁵ The Court did not completely negate the possibility that comparative proportionality review might be required for a particular death penalty statute, but it reasoned that California's statute was not "so lacking in other checks on arbitrariness that it would not pass constitutional muster without comparative proportionality review."²⁷⁶ The Court also acknowledged that "any capital sentencing scheme may occasionally produce aberrational outcomes, [but] such inconsistencies are a far cry from the major systemic defects identified in *Furman*."²⁷⁷ While the Court could have been correct, in principle, it failed to reference concrete evidence supporting its assertion that those major pre-*Furman* defects were, in fact, relics of the past. In fact, there was a growing body of evidence to the contrary and Justice Brennan highlighted this discrepancy, among others, in his dissenting opinion.²⁷⁸ Also troubling to Brennan was the Court's

²⁷³ *Pulley v. Harris*, 465 U.S. at 39–40.

²⁷⁴ *Harris v. Pulley*, 692 F.2d 1189 (9th Cir. 1982).

²⁷⁵ *Pulley v. Harris*, 465 U.S. at 45 ("Examination of our 1976 cases makes clear that they do not establish proportionality review as a constitutional requirement").

²⁷⁶ *Id.* at 52.

²⁷⁷ *Id.* at 54.

²⁷⁸ Studies highlighted by Justice Brennan in his dissenting opinion, in draft or published form, included: David C. Baldus & George Woodworth, *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 J. CRIM. L. & CRIMINOLOGY 661 (1983); William J. Bowers & Glenn L. Pierce, *Arbitrariness and Discrimination under the Post-Furman Capital Statutes*, 26 CRIME & DELINQUENCY 563 (1980); Linda A. Foley & Richard Powell, *The Discretion of Prosecutors, Judges, and Juries in Capital Cases*, 7 CRIM. JUST. REV. 16 (1982); Samuel R. Gross & Robert Mauro, *Patterns of Death: An Analysis of Racial Disparities in Capital Sentencing and Homicide Victimization*, 37 STAN. L. REV. 27 (1984); Michael L. Radelet & Glenn L. Pierce, *Race and Prosecutorial Discretion in Homicide Cases*, 19 LAW & SOC'Y REV. 587 (1985); Marc Riedel, *Discrimination in the Imposition of the Death Penalty: A Comparison of the Characteristics of Offenders Sentenced Pre-Furman and Post-Furman*, 49 TEMP. L. Q. 261 (1976); Hans Zeisel, *Race Bias in the Administration of the Death Penalty: The Florida Experience*, 95 HARV. L. REV.

refusal to consider whether “comparative proportionality review should be required in order to ensure that the irrational, arbitrary, and capricious imposition of the death penalty invalidated by *Furman* does not still exist.”²⁷⁹

The implications of the Court’s ruling in *Pulley* were far-reaching. Several states that adopted Georgia’s model of proportionality review either explicitly repealed the applicable provisions or expressed views that the process was unnecessary.²⁸⁰ Detailed assessments of comparative proportionality review practices in states retaining the procedural safeguard uncovered a common characteristic: “an apparent inability or unwillingness to monitor their capital-sentencing in a sufficiently consistent, comprehensive, and principled manner to identify excessive or discriminatory sentences when they occur.”²⁸¹ The National Center for State Courts (NCSC) Project on Comparative Proportionality Review in Death Sentence Cases released a report explaining that the

456 (1981).

²⁷⁹ *Pulley v. Harris*, 465 U.S. at 74.

²⁸⁰ BALDUS ET AL., *supra* note 49, at 280, 290.

²⁸¹ *Id.* at 280; *See also* Rankin et al., *supra* note 131 (describing problems with Georgia’s system of comparative proportionality review); Kaufman-Osborn, *supra* note 272 (identifying fatal shortcomings in Washington State’s comparative proportionality review system); Donald H. Wallace & Jonathan R. Sorensen, *Comparative Proportionality Review: A Nationwide Examination of Reversed Death Sentences*, 22 AM. J. CRIM. JUST. 13 (1997) (conducting a systematic review of comparative proportionality review nationwide and discovering that appellate courts routinely failed to conduct meaningful review of death sentences); Rhonda G. Hartman, *Critiquing Pennsylvania’s Comparative Proportionality Review in Capital Cases*, 52 U. PITT. L. REV. 871, 872 (1990–1991) (noting that Pennsylvania’s comparative proportionality review protocols are inconsistent with the principles of fairness and uniformity); Donald Wallace & Jonathan Sorensen, *Missouri Proportionality Review: An Assessment of a State Supreme Court’s Procedures in Capital Cases*, 8 NOTRE DAME J. L. ETHICS, & PUB. POL’Y 281, 310–13 (2012) (explaining that the Missouri Supreme Court employs an enfeebled proportionality review process that exacerbates a system that is not operating to minimize the possibility of arbitrary and capricious sentencing). *But cf.* Ken Driggs, *The Most Aggravated and Least Mitigated Murders: Capital Proportionality Review in Florida*, 11 ST. THOMAS L. REV. 207, 275 (1998–1999) (praising Florida’s proportionality review as meaningful and superior to most other jurisdictions, including neighboring Georgia).

only conceivable way to achieve an effective proportionality review process was for reviewing courts to: (a) identify both life- and death-sentence cases comparable to the case being reviewed, (b) determine the proportion of cases resulting in a death sentence, and (c) make a legal judgment as to whether the relative frequency of death sentences within a group is insufficiently large to warrant affirming the sentence.²⁸² Currently, no court appears to be conducting the type of comparative proportionality review consistent with the NCSC model.²⁸³

The diagnostic model I developed in Part II facilitates front-end, empirically informed, comparative review of death penalty charging decisions. This front-end review can be used to compliment, and not replace, back-end review. Moreover, my diagnostic model need not be the only, or even the most important, feature of the front-end review, but it would be an indispensable component of this review process because of its social scientific rigor and replicability. The use of my proposed model would also avoid three persistent shortcomings of current back-end proportionality review practices: (1) the failure to develop general measures of culpability that enable courts to identify comparable cases irrespective of factual differences; (2) failure to make available to the parties in litigation the information on the cases considered by the court in its proportionality review; and (3) the inability (or unwillingness) of courts to identify evidence of racial/ethnic discrimination in the imposition of death sentences.²⁸⁴

First, my statistical model provides a general measure of culpability, based on the relationship between numerous case-level factors and actual capital charging outcomes. As such, it avoids the problems of comparison methods that require the identification of factually identical (or at least very similar) cases.²⁸⁵ Second, the data and methods used for the assessment of the capital charging system and future individual cases would

²⁸² BALDUS ET AL., *supra* note 49, at 281.

²⁸³ *Id.* at 282.

²⁸⁴ *Id.* at 286.

²⁸⁵ *See, supra*, Part II.C.

be available to both all parties involved in the litigation. This facilitates an open inquiry into the data and model used to draw inferences about general patterns of charging behavior. Third, and finally, the model allows a more nuanced understanding of racial/ethnic disparities in capital charging. By identifying a statewide “baseline race/ethnicity effect” and the variability of the effect across sub-state units, one can draw more reliable inferences about the influence of race/ethnicity in capital charging as an indicator of arbitrariness. Even if it is the case that the overall race/ethnicity effect (for defendants or victims) appears to be negligible or an artifact of the model misspecification of deficiencies with the data, it is much less likely that race/ethnicity effects that are two or three orders of magnitude greater than the statewide average are merely the product of potential shortcomings of the model and data.

Front-end comparative proportionality review may hold genuine promise for disciplining capital charging. It removes much of the mystery that has continued to plague the back-end process—namely, inadequate or inappropriate comparisons and lack of transparency. The process would entail some start-up costs, but the intellectual and financial resources necessary to get the project operational should be relatively minimal. In fact, there are numerous examples of commissions tasked with gathering and analyzing data on the operation of the death penalty in their respective jurisdictions. At the federal level, attorneys at the Department of Justice (DOJ) collect data on all potential federal death penalty cases and front-end charge screening is performed by a committee with input from attorneys for both parties.²⁸⁶ Two state legislatures, Kentucky and North Carolina, even enacted legislation requiring not only the collection and analysis of capital punishment litigation data, but also providing legal causes of action for defendants raising certain claims, such as racial/ethnic discrimination in charging or sentencing, with the statistical evidence.²⁸⁷

²⁸⁶ U.S. Department of Justice, *The Federal Death Penalty System: Supplementary Data, Analysis and Revised Protocols for Capital Case Review* (Government Printing Office 2001).

²⁸⁷ Kentucky Racial Justice Act, Ky. Rev. Stat Ann. § 532.300 (Ky. 1998);

Legislators and prosecutors in other jurisdictions have been aware of the extensive empirical literature documenting arbitrariness and bias for decades, often in their own counties and states, but this evidence has had little impact, if any, on general charging patterns. The material and psychological benefits of the death penalty for elected officials and frontline prosecutors may simply override any commitment to truly rationalizing the process at a systemic level. Social scientific evidence identifying problems with the administration of capital punishment has never been self-implementing. Actually disciplining prosecutorial behavior requires more. As I mentioned, *supra*, genuine feedback mechanisms from appellate and post-conviction stages to prosecutors that force them (or their counties) to internalize the costs of their mistakes are essential. While my diagnostic model is useful in identifying potential charging error, the insights gained from it may not become fully realized unless embedded in a larger framework that forces prosecutors to confront the consequences of foreseeably poor choices in capital charging.²⁸⁸ Again, the development and description of this larger framework is beyond the scope of this project.²⁸⁹

B. Summary

The purpose of this section was to provide a fairly rough sketch of how a rigorous social scientific model of capital charging can be incorporated into a more robust front-end comparative proportionality review process to incentivize more carefully and empirically informed capital charge screening. I have only described what I believe are some (not necessarily all)

North Carolina Racial Justice Act of 2009, N.C. Gen. Stat. § 15A-2010 (N.C. 2010) (repealed 2013). Under the North Carolina statute, for example, a capital defendant can have his or her sentence reduced to life imprisonment without parole if there is evidence proving "that race was a significant factor in decisions to seek or impose the sentence of death in the county, the prosecutorial district, the judicial division, or the State at the time the death sentence was sought or imposed."

²⁸⁸ Liebman, *supra* note 7, at 324-25.

²⁸⁹ Thaxton, *supra* note 30.

of the indispensable features of this front-end reform, but hopefully, this can provide a springboard for future research; nonetheless, two important caveats are in order. First, future proposals that are likely to gain traction with courts, legislatures, and the general public are unlikely to constitute radical departures from the corpus of proposals articulated by other scholars also concerned with reducing/eliminating both arbitrariness and bias in the capital charging-and-sentencing process. This conventionality is understandable, in many respects, given the nature of the doctrinal, political, and structural constraints confronting death penalty reformers, either abolitionist or retentionists. Any feasible proposal must be developed with full awareness of the (im)practicalities accompanying the administration of the government's ultimate sanction.²⁹⁰

Second, death penalty reformers should recognize that there are important, yet often underappreciated, limitations to *technocratic* thought—that is, the widely held view among lawyers and legal scholars that problems of law can be reduced to problems of technique.²⁹¹ It remains an open question whether the capital punishment process is amenable to the types of “fixes” that have been successfully implemented by the federal government to substantially reduce (to varying degrees) the arbitrariness and racial/ethnic bias in previously problematic areas such as voting, housing, employment, and

²⁹⁰ FROM LYNCH MOBS TO THE KILLING STATE: RACE AND THE DEATH PENALTY IN AMERICA (Charles J. Ogletree Jr. & Austin Sarat eds., 2006); Stephen B. Bright & Patrick J. Keenan, *Judges and the Politics of Death: Deciding between the Bill of Rights and the next Election in Capital Cases*, 75 B.U. L. REV. 759 (1995); Stephen B. Bright, *In Defense of Life: Enforcing the Bill of Rights on Behalf of Poor, Minority and Disadvantaged Persons Facing the Death Penalty*, 57 MO. L. REV. 849 (1992); Stephen B. Bright, *The Politics of Crime and the Death Penalty: Not “Soft on Crime,” but “Hard on the Bill of Rights,”* 39 ST. LOUIS U. L.J. 479 (1995); Ogletree, *supra* note 252; Steiker & Steiker, *supra* note 219; ANDREW WELSH-HUGGINS, NO WINNERS HERE TONIGHT: RACE, POLITICS, AND GEOGRAPHY IN ONE OF THE COUNTRY'S BUSIEST DEATH PENALTY STATES (1st ed. 2009).

²⁹¹ Donald Black, *The Boundaries of Legal Sociology*, 81 YALE L.J. 1086, 1090–91 (1972).

public education.²⁹² Although broad consensus exists among death penalty analysts concerning the sources of arbitrariness and bias,²⁹³ there is considerable disagreement over whether any set of proposed remedies will produce the desired result.²⁹⁴ According Carol Steiker and Jordan Steiker, “the body of doctrine produced by the Court is enormously complex and its applicability to specific cases [is] difficult to discern; yet it remains unresponsive to the central animating concerns that inspired the Court to embark on its regulatory regime in the first place. Indeed, most surprisingly, the overall effect of [forty]-odd years of doctrinal head-banging has been to substantially reproduce the pre-*Furman* world of capital sentencing.”²⁹⁵ All proposals aimed at repairing the “broken system” of capital punishment must fully acknowledge that inherent limitations may exist with respect altering doctrines, policies, and practices in order to achieve the even-handed administration of the death penalty. Establishing “super due process” for criminal defendants on the one hand, and encouraging “tighter” monitoring and regulation of charging and plea-bargaining practices in district attorney offices, on the other hand, may only provide minor fixes to longstanding problems of the improper influence of race/ethnicity, gender, social class, and geography.

²⁹² Howe, *supra* note 90; Ogletree, *supra* note 252.

²⁹³ These factors include overly-broad capital statutes, inadequate legal representation, decentralized decision-makers, broad prosecutorial and sentencer discretion, and overly restrictive merits review in state and federal habeas proceedings.

²⁹⁴ Cf. Baldus et al., *supra* note 18 (refuting the claim that discrimination in the imposition of the death penalty is inevitable and impossible to prevent) with William J. Bowers & Wanda D. Foglia, *Still Singularly Agonizing: Law's Failure to Purge Arbitrariness from Capital Sentencing*, 39 CRIM. L. BULL. 51 (2003) (demonstrating that constitutionally mandated requirements to guide jury discretion and eliminate arbitrariness in sentencing are not working); Howe, *supra* note 90 (suggesting that even an increase in the due process protections -- “super due process” -- for capital defendants is unlikely to significantly reduce or eliminate racial bias in the capital punishment process); Smith et al., *supra* note 230 (demonstrating that the Court’s mitigation-facilitating doctrines have largely failed to benefit capital defendants with compelling evidence for mercy).

²⁹⁵ Carol S. Steiker & Jordan M. Steiker, *Sober Second Thoughts: Reflections on Two Decades of Constitutional Regulation of Capital Punishment*, 109 HARV. L. REV. 355, 359 (1995).

The reality may be that it is unlikely that more “tinkering” is all that is needed to satisfy the still unfilled promise of *Furman*.²⁹⁶

CONCLUSION

Following a series landmark rulings in the late-1970s and early-1980s,²⁹⁷ the U.S. Supreme Court appeared to craft a capital punishment jurisprudence that unambiguously mandated that frontline criminal justice officials reserve the death penalty for the worst crimes and worst criminals.²⁹⁸ The Court announced the arbitrary and discriminatory administration of capital punishment would no longer be tolerated. The guided-discretion statutes that emerged in the aftermath of *Furman* offered the promise of constraining the hyper-discretion that existed in the pre-*Furman* era, but the Court’s faith in these doctrines to effectively guard against caprice and bias in the capital charging-and-sentencing was misplaced.²⁹⁹ Much to the dismay of the various current and former U.S. Supreme Court Justices and cautiously optimistic death reformers,³⁰⁰ rigorous empirical research on the capital

²⁹⁶ See *Callins v. Collins*, 510 U.S. 1141 (1994) (Blackmun, J., dissenting) (stating that he would no longer “tinker with the machinery of death”).

²⁹⁷ See also *Furman v. Georgia*, 408 U.S. 238 (1972); *Gregg v. Georgia*, 428 U.S. 153 (1976); *Coker v. Georgia*, 433 U.S. 584 (1977) (death penalty for non-homicidal rape is unconstitutional); *Eberheart v. Georgia*, 433 U.S. 917 (1977) (death penalty is unconstitutional for non-homicidal kidnapping); *Lockett v. Ohio*, 438 U.S. 586 (1978) (statutory restrictions on mitigation evidence is unconstitutional); *Presnell v. Georgia*, 439 U.S. 14 (1978) (death sentence must be based on a statutorily defined factor); *Godfrey v. Georgia*, 446 U.S. 420 (1980) (vaguely defined aggravating factors are unconstitutional); *Enmund v. Florida*, 458 U.S. 782 (1982) (death sentence unconstitutional when defendants are neither killers nor had the intention to kill).

²⁹⁸ In the seven years immediately following the Court’s decision in *Gregg*, it ruled in favor of 14 out of 15 death-sentenced inmates whose appeals were fully heard. HAINES, *supra* note 49.

²⁹⁹ BALDUS ET AL., *supra* note 49.

³⁰⁰ Multiple U.S. Supreme Court Justices have made statements in recent years either expressly condemning the practice of capital punishment, or raising serious concerns as to its fair administration. Moreover, several Justice who voted to uphold the constitutionality of the death penalty while

charging-and-sentencing process almost unequivocally reveals that states have failed to purge the process of the arbitrariness and bias that the Court believed to be particularly rampant pre-*Furman*.³⁰¹ Prosecutors' charging decisions remain highly arbitrary both within and across jurisdictions.³⁰² Effective death penalty reform, if possible, must begin with the gatekeepers of the system.

METHODOLOGICAL APPENDIX

The multilevel models (MLMs) utilized in this Article offer four key improvements over prior research investigating capital charging dynamics.³⁰³ These models (1) correctly take into

on the Court publically criticized the death penalty after retiring from the Court. *See, supra*, note 6.

³⁰¹ Baldus & Woodworth, *Race Discrimination in the Administration of the Death Penalty*, *supra* note 140; Baldus & Woodworth, *Race Discrimination in America's Capital Punishment System Since Furman v. Georgia* (1972), *supra* note 140; Richard C. Dieter, *Twenty Years of Capital Punishment: A Re-Evaluation* (Death Penalty Information Center 1996); Steiker & Steiker, *supra* note 295.

³⁰² In his concurring opinion in *Furman*, Justice Stewart famously wrote that "death sentences are cruel and unusual in the same way that being struck by lightning is cruel and unusual." *Furman v. Georgia*, 408 U.S. at 309. More recently, Washington Supreme Court Justice Charles Johnson reiterated Justice Stewart's concern in his dissenting opinion: "Reviewing the history of this court's proportionality review reveals how the administration of capital cases defies any rational analysis. [...] These cases exemplify the arbitrariness with which the penalty of death is exacted. They are symptoms of a system where statutory comparability defies rational explanation. The death penalty is like lightening, randomly striking some defendants and not others. [...] No rational explanation exists to explain why some individuals escape the penalty of death and others do not." *State v. Cross*, 132 P.3d 80, 110–15 (Wash. 2006).

From 2006–2013, 355 condemned inmates were executed and 261 individuals were killed by lightning in the United States—a ratio of approximately 1.4 executions for every lightning strike death. Richard C. Dieter, *Facts About the Death Penalty* (Death Penalty Information Center), Apr. 2015; National Weather Service, U.S. Lightning Fatalities, 2006–2015 (2015), <http://www.lightningsafety.noaa.gov/fatalities.shtml>.

³⁰³ *See* Sherod Thaxton, *Un-Gregg-Ulated: Capital Charging and the Missing Mandate of Gregg v. Georgia*, 11 DUKE J. CONST. L. & PUB. POL'Y 145, 166

account the non-independence of cases nested in the same judicial circuit;³⁰⁴ (2) allow for the partitioning of variation in the case-level charging decisions into within- and between-circuit effects;³⁰⁵ (3) provide better estimates of the effects of case-level explanatory variables by combining information on both the with- and between-circuit effects of those variables;³⁰⁶ and (4) produce sensible (and statistically defensible) calculations of circuit-specific effects that facilitate the assessment and ranking of the institutional performance of the circuits.³⁰⁷

Prior studies have *recognized* that between-circuit processes may be an important source of variability in capital

(2016) (describing the virtues of multilevel modeling to investigate the constitutionality of death penalty dynamics).

³⁰⁴ Inferences drawn from analytical frameworks that do not explicitly account for the fact that death eligible cases are nested in different jurisdictions are often misleading because relationships measured at one level of analysis (e.g., between cases) do not necessarily hold at another level of analysis (e.g., between circuits). Interpreting associations at the higher level as pertaining to the lower level is known as an *ecological fallacy*. The opposite of the ecological fallacy is an *atomistic fallacy*, and this occurs when one draws inferences about the relationships between group-level variables based on information about individual-level relationships. These fallacies are problems of *inference*, not of measurement. DOUGLAS A. LUKE, MULTILEVEL MODELING 5–6 (Quantitative Applications in the Social Sciences, volume 143, 2004).

³⁰⁵ Multilevel models have a complex error structure because the total variability in individual outcomes is comprised of two components: the within-cluster variance and the between-cluster variance. Decomposing the random part of the multilevel model into unit-specific and cluster-specific effects allows the analyst determine how much variability in the outcome can be attributed to each level. ANDERS SKRONDAL & SOPHIA RABE-HESKETH, GENERALIZED LATENT VARIABLE MODELING: MULTILEVEL, LONGITUDINAL, AND STRUCTURAL EQUATION MODELING 51 (2004).

³⁰⁶ The case-level estimates of explanatory variables in MLMs is the weighted average of the within- and between-circuit effects. GELMAN & HILL, *supra* note 112, at 478 (explaining the use of pooling factors for weighting individual coefficients in a multilevel model).

³⁰⁷ Estimates of the variability of the within- and between-circuit influences on the outcome, net of the case-level explanatory factors, are used to predict the circuit-specific effects. SKRONDAL & RABE-HESKETH, *supra* note 305, at 225; Duncan et al., *supra* note 25, at 111.

charging and sentencing, yet they have not properly incorporated this information in their analyses.³⁰⁸ The unfortunate result of this oversight has been an incomplete description and understanding of the operation of capital punishment systems. Rather than taking inter-jurisdictional variability as an important object of study, prior research as treated inter-jurisdictional variability as a nuisance that needed to be minimized or corrected in an effort to properly analyze intra-jurisdictional variability.³⁰⁹

The traditional approach adjusts for differences across jurisdictions through the use of “fixed effects”—that is, the estimation of a set of jurisdiction-specific regression coefficients intended to capture differences between jurisdictions for similar cases.³¹⁰ But the fixed effects modeling framework is ill-suited for the investigation of death penalty charging data for at least four reasons.³¹¹ First, information from jurisdictions that

³⁰⁸ The importance of separating variability into within- and between-circuit effects was brilliantly underscored by two statisticians who famously remarked, “One statistician’s error term is another’s career!” DON HEDEKER & ROBERT D. GIBBONS, *LONGITUDINAL DATA ANALYSIS* 56 (2006).

³⁰⁹ Carroll, *supra* note 124, at 211–15 (noting the strong tendency of statistical analyses to treat variability in effects as uninteresting rather than an object of study).

³¹⁰ Even studies focusing on geographical differences fail to employ the MLM framework, which is specifically designed to permit a more nuanced assessment of jurisdictional variability. *See, e.g.*, Katherine Barnes et al., *Place Matters (Most): An Empirical Study of Prosecutorial Decision-Making in Death-Eligible Cases*, 51 ARIZ. L. REV. 305 (2009); Donohue, *supra* note 106; Paternoster et al., *supra* note 122; Songer & Unah, *supra* note 115; David Weisburd & Joseph Naus, *Report to Special Master David Baime: Re Systemic Proportionality Review* (New Jersey Judiciary 2001). Although these studies show that significant differences remain between jurisdictions even after taking account numerous case-level explanatory variables, the analytical framework they employ cannot provide answers to questions that I specifically address in my analyses by using the MLM framework.

³¹¹ Cf. PAUL D. ALLISON, *FIXED EFFECTS REGRESSION MODELS 2* (Quantitative Applications in the Social Sciences, volume 160, 2009) (explaining that prior studies have preferred the fixed effects approach, in part, because it can take into account any jurisdiction-level unobserved effects on case outcomes that may be potentially correlated with case-level explanatory variables) *with* Bell & Jones, *supra* note 116 (“[W]e take the strong, and rather heterodox,

either have a single case (called *singleton clusters*) or from jurisdictions where all cases received the same outcome (i.e., no death-eligible case was noticed for the death penalty or all death eligible cases were noticed for the death penalty) must be discarded.³¹² MLMs, on the other hand, utilize information from circuits with a single observation and from circuits containing cases that are identical with respect to their death noticing outcomes. Although the circuit-level effect will be imprecisely estimated, the information from the single case contributes to the estimation of the coefficients and variance parameters of the individual and circuit-level regressions.

Second, under the fixed-effects framework, cases must exhibit substantial within-circuit variation along multiple case-level characteristics in order to reliably explore case-level dynamics. If a substantial proportion of the variation in the case-level characteristics is between-circuit and not within-within circuit, then the traditional approach will give imprecise estimates because the case-level estimates only deal with a small subsection of the variance of the case-level characteristic. As two scholars have recently noted, “[I]n controlling out context, [fixed effects] models effectively cut out much of what is going on—goings-on that are usually of interest to the researcher, the reader and the policy maker[—]and offer overly simplistic and impoverished results that can lead to misleading interpretations”³¹³ By explicitly modeling circuit-level heterogeneity in capital charging, MLMs can sensibly

view that there are few, if any, occasions in which [fixed effects] modeling is preferable to [MLM]”) (alterations added).

³¹² The fixed effects framework removes the between-jurisdiction variability from the model, so all that remains to be examined is within-jurisdiction variability. As a consequence, singleton clusters or jurisdictions in which all death-eligible cases either receive or do not receive a death notice are dropped from the analysis. There can be no within-jurisdiction variation with a single case or multiple cases that are identical across the variables examined in the model. ADRIAN COLIN CAMERON & PRAVIN K. TRIVEDI, MICROECONOMETRICS: METHODS AND APPLICATIONS 796–97 (2005).

³¹³ Bell & Jones, *supra* note 116, at 134 (“[I]n controlling out context, [fixed effects] models effectively cut out much of what is going on—goings-on that are usually of interest to the researcher, the reader and the policy maker.”).

incorporate information about within- and between-circuit variability in the effects of explanatory factors, and therefore provide reasonable answers about the general effect of a variable even when a very large proportion of variability in explanatory variable is between-circuit.³¹⁴

Third, fixed effects models require the data to contain a moderate to large number of cases in each jurisdiction in order to provide an accurate measure of the jurisdiction-specific effect. This results from the fact that the framework treats each jurisdiction separately from the other jurisdictions in order to capture the within-jurisdiction variability. As one scholar aptly noted, the fixed effects approach is “amnesic” because every new cluster is treated like a new world and does not benefit from the information about other clusters.³¹⁵ Whereas fixed effects models are hindered by their naïve treatment of circuit-specific effects that give evidence from small circuits undue influence, MLMs avoid this problem by weighting small circuits that exhibit high within-circuit variability (i.e., unreliable information) towards the overall average, thereby minimizing the influence of these small circuits on the determination of cluster-specific effects.³¹⁶

Lastly, the fixed effects approach does not permit inferences about the between-jurisdiction variation, including whether the variability is substantively meaningful. This is especially true when the outcome variable is binary, such as the decision to seek the death penalty against a defendant, because even the naïve estimates of the circuit-specific effects are precluded.³¹⁷

³¹⁴ The variance of the parameter estimate will also be impacted by the weighting because the uncertainty around the effect of any particular variable related to the (dis)similarity of the jurisdictions. RAUDENBUSH & BRYK, *supra* note 112, at 47.

³¹⁵ RICHARD MCELREATH, *STATISTICAL RETHINKING: A BAYESIAN COURSE WITH EXAMPLES IN R AND STAN 355* (2015).

³¹⁶ *Id.*

³¹⁷ CAMERON & TRIVEDI, *supra* note 312, at 796 (explaining that fixed effects models for binary outcomes, such as noticing decisions, cannot produce estimates of the circuit-specific effects). Some scholars have employed a linear probability model (LPM) to estimate the cluster-specific

All of these aforementioned problems originate from the inability of the fixed effects framework to explicitly model context and heterogeneity, and therefore simultaneously consider within- and between-jurisdiction variability. Bell and Jones have argued that the fixed effect technique is “used it too readily as a default option without a full understanding of what is being estimated and what is being lost by doing so.”³¹⁸

The intuition behind MLMs is the estimation of the mean and variance of the *distribution* of the circuit-specific effects—but not the actual circuit-specific effects—via the imposition of a modest constraint on the variability of the between-circuit effects: a probability model that assumes the circuit-specific effects arise from deviations from a typical circuit in the state. The constraint allows the model to utilize all of the information from available cases to provide better estimates of case-level effects and circuit-level variability. MLMs offer a sensible compromise between within- and between-circuit effects (called *partial pooling*) because the effect of a case-level factor is neither purely a between-circuit effect (because the case-level factor can vary across cases within jurisdictions) nor purely within-circuit effects (because the case-level factor may be constant across cases within a particular circuit).³¹⁹

MLMs can be written in two parts: a Level-1 model and a Level-2 model. The Level-1 model is, essentially, a series of sub-models for each Level-2 unit (e.g., judicial circuit). The unit of

fixed-effects, called “unconditional fixed effects,” on a binary outcome. *E.g.*, David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145, 1168 (2007). Not only is the LPM subject to the same shortcomings as the traditional fixed effects models for a non-binary outcome, but the unconditional effects are also inconsistent (i.e., they fail to converge to the true effect as the sample size increases, so the LPM suffers from additional shortcomings. See Ethan Katz, *Bias in Conditional and Unconditional Fixed Effects Logit Estimation*, 9 POL. ANAL. 379, 380 (2001).

³¹⁸ Bell & Jones, *supra* note 116, at 134 (alterations added).

³¹⁹ The multilevel estimate, β_k^{ML} , is calculated as: $\lambda_j \beta_j^B + (1 - \lambda_j) \times \beta_k^W$; where $\lambda_j = \psi_j / [\psi_j + (\theta / n_j)]$, n_j is the sample size of judicial circuit, j , ψ is the between-circuit variance, θ is the variance within circuits, β_k^B is the between-circuit effect, and β_k^W is the within-circuit effect. GELMAN & HILL, *supra* note 112, at 477–78.

analysis is the death-eligible murder case and the sample size for each regression is number of death-eligible cases for each particular judicial circuit. This model captures variability in death noticing among cases within the judicial circuit. Formally, the Level-1 model can be written as: $\Pr(y_{ij} = 1) = g^{-1}(\beta_{0j} + \beta_k X_{kij} + \epsilon_{ij})$, where the subscripts i and j index the i th defendant and j th judicial circuit, respectively, y_{ij} is a binary outcome indicating “1” if the defendant is noticed for the death penalty and “0” if otherwise, $\Pr(y_{ij} = 1)$ is the probability that the defendant i th in circuit j th is noticed for the death penalty, β_{0j} (*beta*) is the circuit-level probability that a defendant is noticed for the death penalty (conditional on all explanatory variables being equal to zero), X_k are k explanatory variables with β regression coefficients, and ϵ_{ij} (*epsilon*) are Level-1 errors (case-level the deviation from the expected probability for the i th defendant in the j th judicial circuit).³²⁰ In order to meaningfully interpret β_{0j} , explanatory variables, X_k , are centered at their overall (i.e., statewide) average values, $\bar{X}_k = 0$, so the intercept is the probability of a “typical” case in Georgia receiving the death notice in the j th circuit.

The unit of analysis for the Level-2 model is the judicial circuit, not the individual death-eligible cases, and the outcome variable is the circuit-specific probability (β_{0j}). Formally, the Level-2 model is: $\beta_{0j} = \gamma_{00} + \zeta_{0j}$, where β_{0j} is the same as described above, γ_{00} is the probability of a death notice for the typical circuit (i.e., the statewide average across circuits, not cases), and ζ_{0j} (*zeta*) is a circuit-specific deviation from the

³²⁰ Here $g^{-1}(\cdot)$ is the *inverse link function* (also called the logistic function), so $\Pr(y_{ij} = 1) = \frac{\exp(\beta_{0j} + \beta_k X_{kij} + \epsilon_{ij})}{1 + \exp(\beta_{0j} + \beta_k X_{kij} + \epsilon_{ij})}$. Technically speaking, there is no Level 1 residual error (ϵ_{ij}) in this mathematical expression, but ϵ_{ij} appears in another equivalent formulation: $y_{ij}^* = \beta_{0j} + \beta_k X_{kij} + \epsilon_{ij}$, where y_{ij}^* represents the propensity to notice a case for the death penalty, such that $y_{ij} = 1$ if $y_{ij}^* > 0$ and $y_{ij} = 0$ if $y_{ij}^* \leq 0$. The equivalence of the two equations can be shown: $\Pr(y_{ij} = 1) = \Pr(\beta_{0j} + \beta_k X_{kij} + \epsilon_{ij} > 0)$. Nevertheless, I include ϵ_{ij} in the prior equation in order to make the interpretation of the regression coefficients more intuitive and relationship of the variance components more apparent in the MLM framework.

statewide average.³²¹ The total variance of the circuit-specific intercepts, $Var(\beta_{0j}) = Var(\zeta_{0j}) + Var(\epsilon_{ij}) = \psi + \theta$, where ψ (*psi*) is the between-circuit variance, $Var(\zeta_j)$, and θ (*theta*) is the within-circuit variance, $Var(\epsilon_{ij})$.³²² When case-level explanatory variables, X_k , are included in the model, ψ and θ are *residual variances*—i.e., variability left unexplained after taking into account the explanatory variables. The ζ 's are not model parameters, but are quantities of interest predicted from the estimated parameters ($\hat{\beta}$, $\hat{\psi}$, and $\hat{\theta}$) which are treated as known. The ζ 's are crucial for making inferences for the circuits in the data (e.g., assessment of institutional performance) and can be

³²¹

The Level-2 model cannot be estimated on its own because the random intercept, β_{0j} , is not observed. Instead, the Level-2 model must be substituted in the Level 1 model to obtain a reduced form model for the observed responses: $y_{ij}^* = \gamma_{00} + \zeta_{0j} + \beta_k X_{kij} + \epsilon_{ij}$, where $(\gamma_{00} + \zeta_{0j}) = \beta_{0j}$.

³²² For the logistic regression model, θ has a fixed variance that is specified, *a priori*, by the logistic distribution: $\theta = \frac{\pi^2}{3} \approx 3.29$. The use of a linear probability model (LPM), which treats a binary outcome variable as continuous, to examine clustered data will give misleading results because θ will be incorrectly estimated from the data, and therefore all inferences based on those statistics will be unreliable.

The LPM suffers from two additional limitations that makes it ill-suited for the current project. First, the LPM assumes that the relationship between the explanatory variables and the binary outcome variable is linear, which is an unrealistic assumption for this project because the explanatory variables attempt to index a defendant's culpability level. For example, it is improbable that the impact of an increase in the number of victims in a homicide case on the probability that a defendant receives the death penalty is the same when the number increases from one to two as it would be from five to six. The logistic regression model explicitly takes this nonlinearity into account to properly estimate the relationships between explanatory variables and the probability of receiving a death notice. Second, with respect to predicting the probability that a case is noticed for the death penalty, particularly cases not included in the estimation sample, the LPM is much more likely give probabilities that are less than "0" and greater than "1". These out-of-range predictions are caused, in part, by the erroneous assumption of a linear relationship between the explanatory variables and the binary outcome variable. Although it is possible to round the predictions up or down to obtain probabilities bounded at zero and one, the out-of-range predictions are strong evidence that data do not meet the assumptions of the model.

used to compare the various circuits in terms of their punitiveness (or leniency) with respect to death noticing behavior because ζ 's are residual deviations (i.e., the deviations take into account the case-level characteristics included in the model). Under the assumption that the key legal features of the death noticing process have been included in the model (see Table 1) or have been proxied by other variables included in the model, these two variance components can be used to measure different aspects of arbitrariness in death noticing decision-making, such as those detailed in Parts II.B.2, II.B.3, and II.B.4.

Unreliability Measures. Part II.B.2 described four measures of unreliability/inconsistency in capital charging based on the aforementioned variance components: within-circuit unreliability (ICC_1),³²³ between-circuit unreliability (ICC_2),³²⁴ the mean absolute deviation (MAD) of the circuits from the statewide average,³²⁵ and the median odds ratio (MOR).³²⁶ The (un)reliability of capital charging behavior was also explored for subsets of cases that varied according to the race of the defendant and race of the victim. As noted, *supra*, it is important to examine whether cases differ in variability along legally impermissible dimensions and not just the average level of an outcome when assessing the level of arbitrariness impacting legal decision-making.³²⁷ This approach, referred to as a "heterogeneous variance analysis," is useful for exploring whether certain classes of cases appear to be handled more haphazardly than others.

Invalidity Measures. The invalidity/irrationality analysis

³²³ See, *supra*, note 163.

³²⁴ See, *supra*, note 177.

³²⁵ See, *supra*, note 183. The mean absolute deviation (MAD) uses the circuit-specific deviation to calculate the average difference of the circuit-level probabilities from the state-wide probability for a factual similar case.

³²⁶ See, *supra*, note 187. The median odds ratio relies on the between-circuit residual variance to quantify the variation between circuits by comparing two charging decisions in factually similar cases from two randomly chosen clusters. The MOR is the average ratio between the cases of higher propensity with the cases of lower propensity.

³²⁷ Hedeker et al., *supra* note 124, at 3328.

discussed in Part II.B.3 was based on the *coefficient of determination* (R^2) and the *coefficient of discrimination* (Tjur's D). Both of these statistics measure of how well the specified model predicts the outcome. The R^2 statistic quantifies the proportional reduction in prediction error variance comparing the model without covariates (the "null" or "unadjusted" model) with the model of interest (the "adjusted" model).³²⁸ As noted, *supra*, the error variance has a complex structure in the multilevel context: $Var(\zeta_j + \varepsilon_{ij}) = \psi + \theta$. The coefficient of determination for two-level models is the proportional reduction in the estimated total error variance comparing the unadjusted model without covariates with adjusted model: $R^2 = [(\hat{\psi}_0 + \hat{\theta}_0) - (\hat{\psi}_1 + \hat{\theta}_1)]/(\hat{\psi}_0 + \hat{\theta}_0)$, where $\hat{\psi}_0$ and $\hat{\theta}_0$ are estimates for the unadjusted model, and $\hat{\psi}_1$ and $\hat{\theta}_1$ are the estimates for the adjusted model.³²⁹ The proportional reduction in each of the variance components can be evaluated: $R_1^2 = (\hat{\theta}_0 - \hat{\theta}_1)/\hat{\theta}_0$ and $R_2^2 = (\hat{\psi}_0 - \hat{\psi}_1)/\hat{\psi}_0$, where R_1^2 and R_2^2 are, respectively, the proportional reduction in the within- and between-circuit residual variances.³³⁰ Similar to the aforementioned unreliability analyses, I disaggregate the measure of the rationality of death penalty charging by race of the defendant and race of the victim and provide an assessment of whether prosecutors death charging decisions appear to be more or less rational/valid depending on the particular subclass of case.

Unlike models analyzing continuous predictors, the level-1 residual variance in the logistic regression model, θ , is fixed so it cannot decrease when adding other variables to the model.³³¹ As a consequence, the variance estimates of the random effects

³²⁸ Analysts disagree as to whether the coefficient of determination is an accurate measure of "model fit" because the R^2 will tend be small when the "true" model has a large residual variance, and therefore it would be erroneous to interpret a small R^2 as indicating model misspecification. See Gary King, *How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science*, 30 AM. J. POL. SCI. 666, 675 (1986).

³²⁹ SNIJDERS & BOSKER, *supra* note 173, ch. 7.

³³⁰ RAUDENBUSH & BRYK, *supra* note 112, ch. 4.

³³¹ HOX, *supra* note 198, at 133–39.

become larger when explanatory variables are included and, therefore, lack a straightforward interpretation.³³² The random effects must be rescaled to permit the calculation of the R^2 statistics for the different model specifications.³³³ The rescaling procedure includes: (a) calculating the total variance of the unadjusted model; (b) calculating the total variance of adjusted model; (c) calculating the scale correction factor (i.e., the square root of the ratio of the variances of unadjusted model to the adjusted model); and (d) rescaling the random effects by using the scale correction factor.³³⁴

A complimentary measure to the R^2 statistic was also calculated: Tjur's D . This statistic assesses the fit of a model on observed data by comparing the predicted probability that the prosecutor filed a death penalty notice when a death notice was actually file to the predicted probability that a death notice was filed when there was actually no noticed filed by the prosecution. The intuition underlying the Tjur's D is that the rationality of death noticing decision-making is directly proportionality to its ability to minimize both false positive and false negatives. When the predictive model, which includes many of the legally (and empirically) relevant factors purportedly driving death penalty charging, fails to differentiate cases with an acceptable degree accuracy, then the rationality/validity of the decision-making is highly questionable. Tjur's D often provides similar answers as R^2 , but is often deemed to be a more interpretable measure when the decision is binary, as is the case with death charging decisions (i.e., yes/no).

Disproportionality Measures. Part II.B.4 presented results from the racial disproportionality component of the arbitrariness analysis. This analysis relied on another key

³³² There is no direct analog to the coefficient of determination for logistic regression, but several "pseudo- R^2 " measures have been developed. J. SCOTT LONG, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES 102–8 (Advanced Quantitative Techniques in the Social Sciences, volume 7, 1997) (describing goodness-of-fit measures for binary regression models).

³³³ Hox, *supra* note 198.

³³⁴ *Id.*

feature of MLMs—the ability to measure the variability of the effects of case-level explanatory factors across judicial circuits. These models, sometimes called “random coefficient” or “heterogeneous effects” models, capture differences in case-level effects across circuits: $\Pr(y_{ij} = 1) = g^{-1}[(\beta_0 + \zeta_{0j}) + (\beta_k + \zeta_{kj})]$, where β_0 and ζ_{0j} were defined earlier, β_k is the average effect of the case-level explanatory variable for the state, and ζ_{kj} is the judicial circuit-specific deviation from the average effect of β_k .³³⁵ For the purposes of the current study, the coefficients of interest were the race-of-defendant and race-of-victim effects.³³⁶

³³⁵ SKRONDAL & RABE-HESKETH, *supra* note 305, at 50.

³³⁶ *Id.*

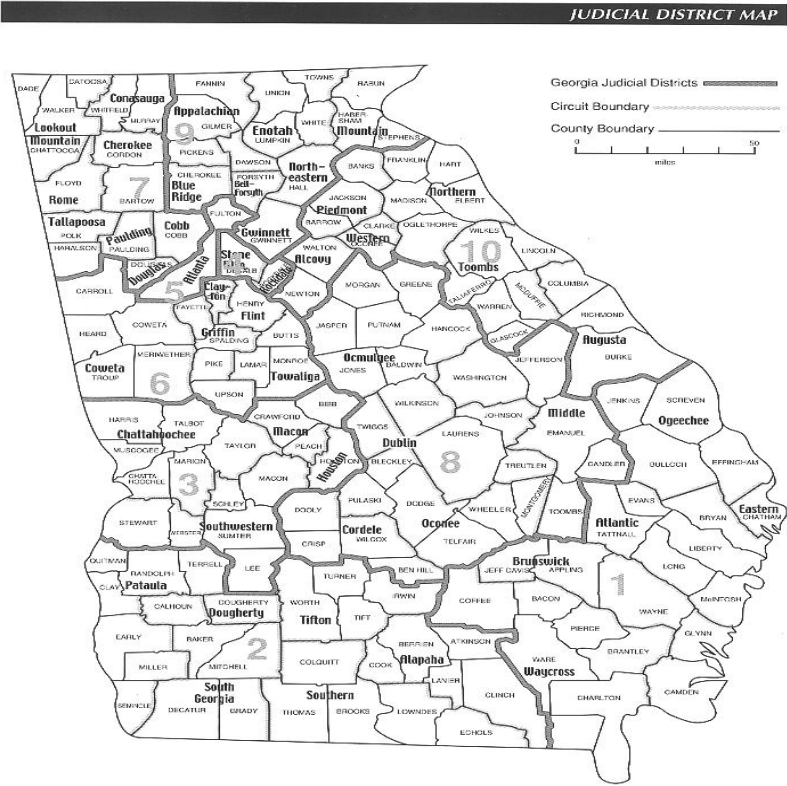


Figure 1: Map of Georgia

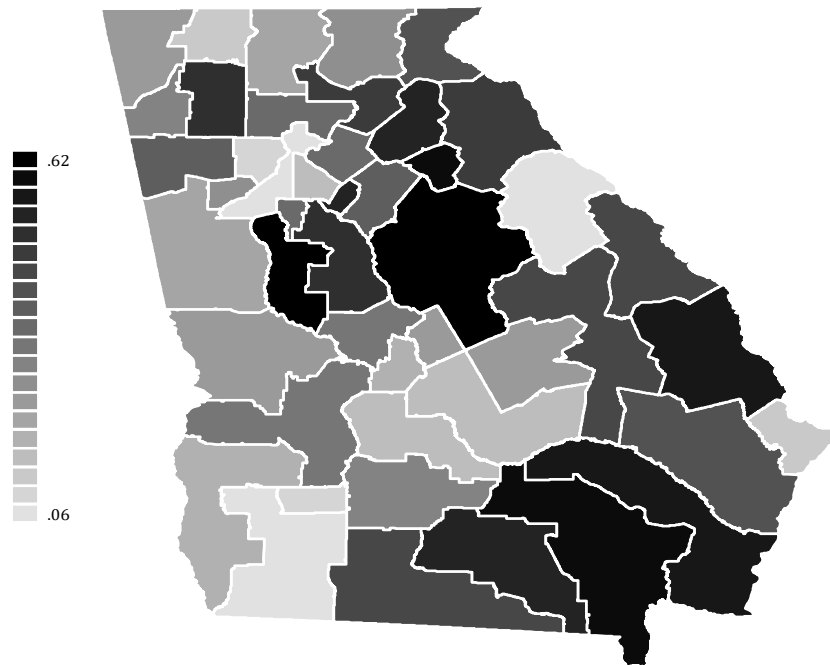


Figure 2: Jurisdictional Variability in the Probability of Receiving a Death Notice (Unadjusted)

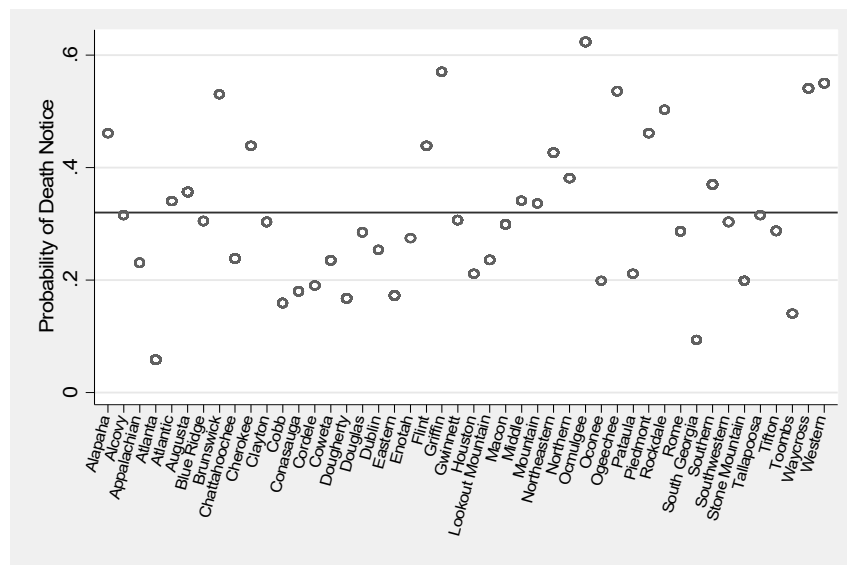


Figure 3: Jurisdictional Variability in the Probability of Receiving a Death Notice (Unadjusted)

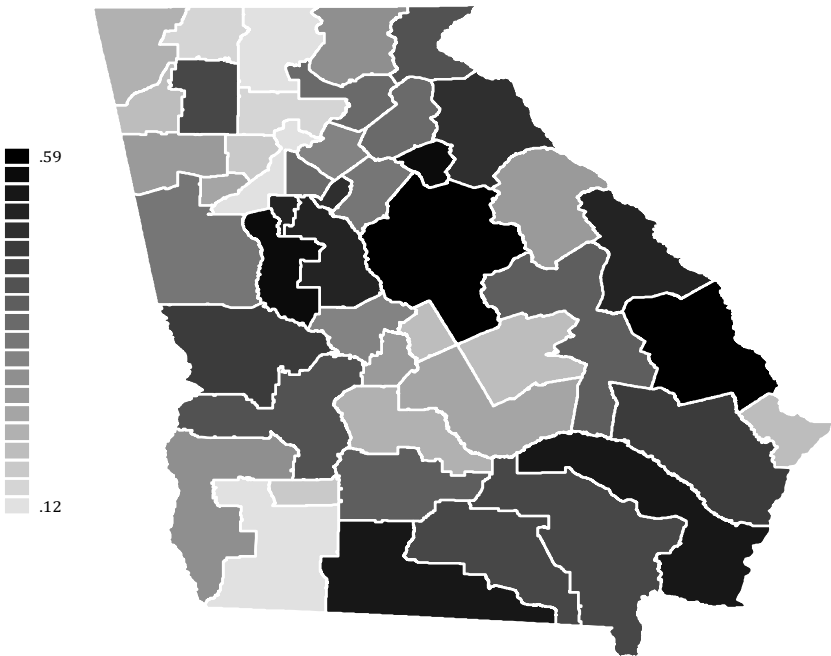


Figure 4: Jurisdictional Variability in the Probability of Receiving a Death Notice (Adjusted)

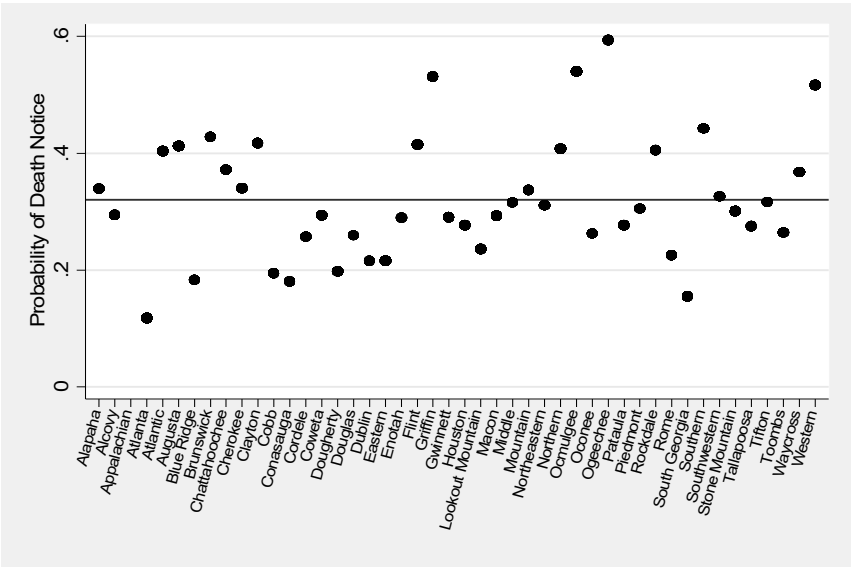


Figure 5: Jurisdictional Variability in the Probability of Receiving a Death Notice (Adjusted)

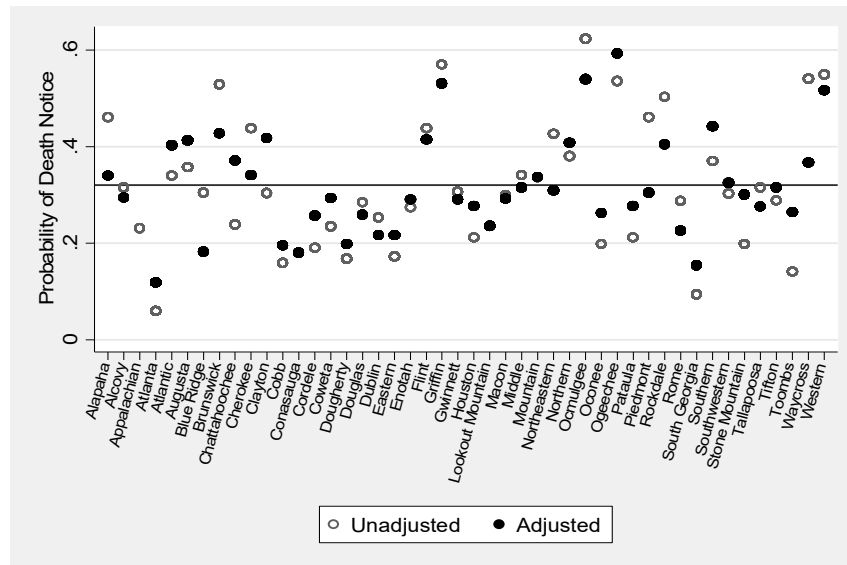


Figure 6: Jurisdictional Variability in the Probability of Receiving a Death Notice (Unadjusted & Adjusted Estimates)

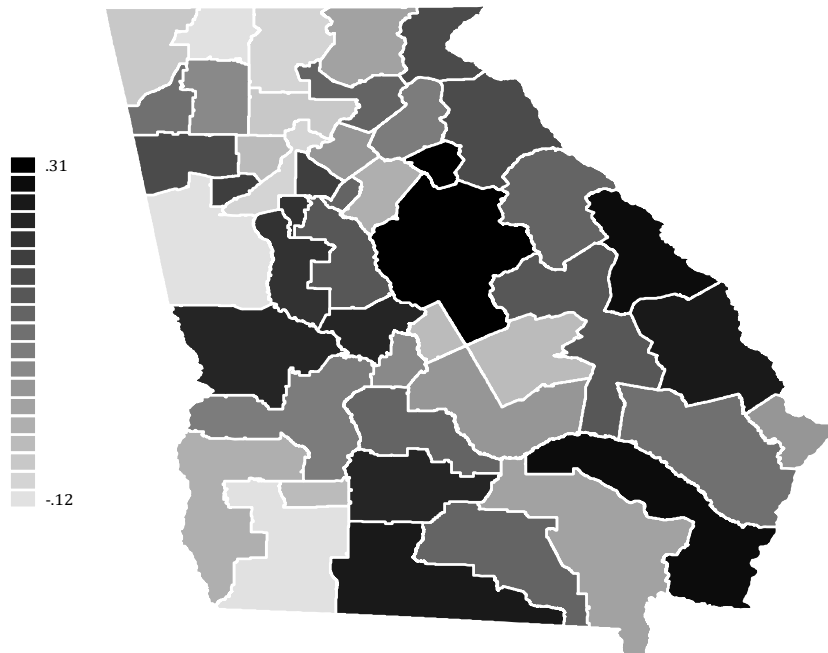


Figure 7: Jurisdictional Variability in the Race of Defendant Effect (Caucasian)

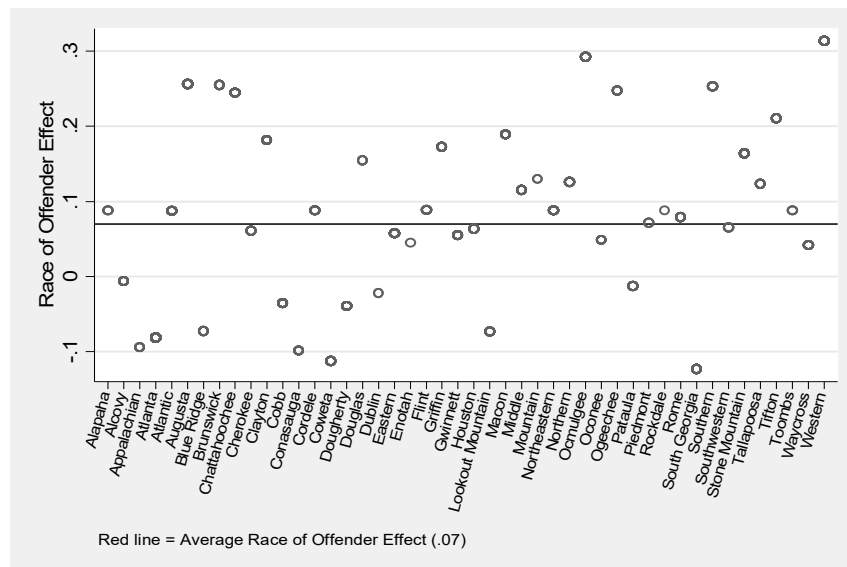


Figure 8: Jurisdictional Variability in the Race of Defendant Effect (Caucasian)

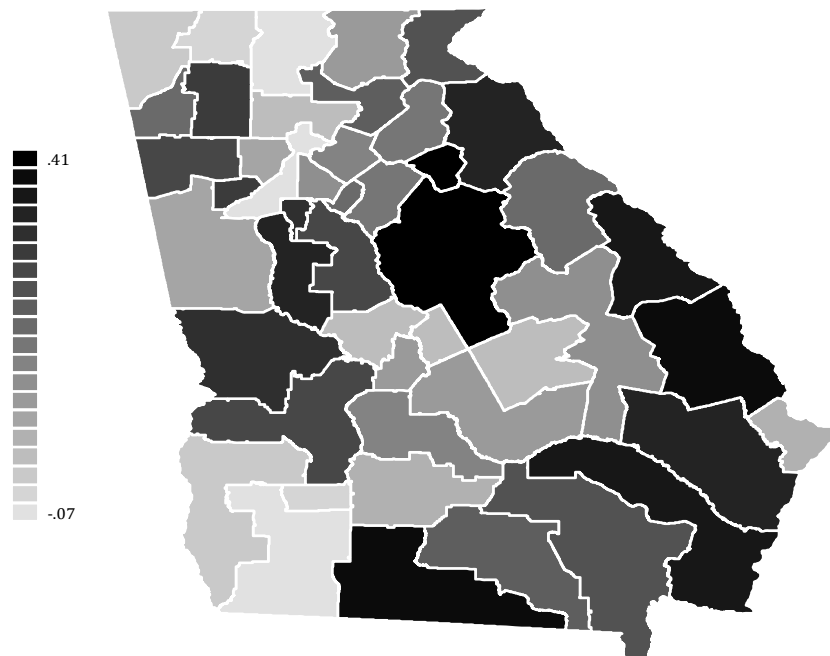


Figure 9: Jurisdictional Variability in the Race of Victim Effect (Caucasian)

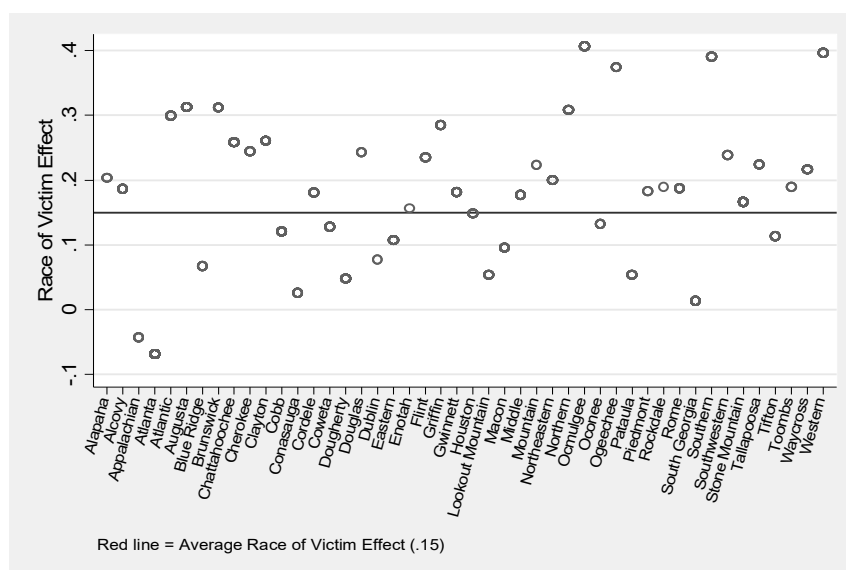


Figure 10: Jurisdictional Variability in the Race of Victim Effect (Caucasian)

Table 1: Summary Statistics

Variables	Mean/ Proportion	Std. Dev.	Min	Max
DP Notice Filed	0.322	--	0	1
Total Statutory Aggs	2.276	1.091	1	7
Year of Offense	--	--	1993	2000
# of Defs	1.793	1.109	1	7
Def White	0.248	--	0	1
Def Black	0.728	--	0	1
Def Latino	0.018	--	0	1
Def Male	0.946	--	0	1
Def Age	27.15	9.935	17	69
Def Contemp Fels	1.724	1.602	0	9
Def Prior Fels	0.514	1.332	0	10
Def # of Children	0.583	--	0	1
Def Employed	0.562	--	0	1
Def Married	0.179	--	0	1
Def HS Grad	0.262	--	0	1
Def Military Service	0.084	--	0	1
Def Drug Use	0.506	--	0	1
Def Psych Status ³³⁷	1.219	0.508	1	4
Def IQ (Culture Fair)	100	14.713	50	151
Def WRAT ³³⁸	7.594	2.802	1	13
Def Fam History ³³⁹	1.298	1.224	0	5
Monetary Motive	0.577	--	0	1
Sex-Crime Motive	0.053	--	0	1
Gang Related Motive	0.003	--	0	1
Def is "Trigger Person"	0.853	--	0	1
Firearm Homicide	0.644	--	0	1
# of Vics	1.185	0.504	1	6
Vic White	0.442	--	0	1
Vic Black	0.504	--	0	1
Vic Latino	0.032	--	0	1
Vic Female	0.360	--	0	1
Vic Age	36.72	18.20	0	97
Vic Stranger	0.350	-	0	1
Interracial Crime	0.283	--	0	1
Total Cases	1,238			

³³⁷ Defendant's psychiatric status: no impairment, minimal impairment, serious impairment, and severe impairment.

³³⁸ Wide Range Achievement Test (reading, math, and spelling).

³³⁹ Summary measure of how many risk factors for criminality were present in the defendant's family environment during childhood: alcohol/drug abuse, emotional/psychological abuse, physical abuse, family criminality, and "broken home."

Table 2: Death Notices by Judicial Circuit (1993-2000)

Judicial Circuit	Death Notices	Percent of Total Death Notices
Alapaha	3	0.8
Alcovy	6	1.5
Appalachian	1	0.3
Atlanta	21	5.3
Atlantic	11	2.8
Augusta	29	7.2
Blue Ridge	3	0.8
Brunswick	16	4.0
Chattahoochee	16	4.0
Cherokee	9	2.3
Clayton	19	4.8
Cobb	11	2.8
Conasauga	1	0.3
Cordele	3	0.8
Coweta	6	1.5
Dougherty	7	1.8
Douglas	3	0.8
Dublin	1	0.3
Eastern	10	2.5
Flint	9	2.3
Griffin	17	4.3
Gwinnett	13	3.3
Houston	2	0.5
Lookout Mountain	3	0.8
Macon	8	2.0
Middle	8	2.0
Mountain	2	0.5
Northeastern	11	2.8
Northern	9	2.3
Ocmulgee	26	6.5
Oconee	2	0.5
Ogeechee	10	2.5
Pataula	3	0.8
Paulding	5	1.3
Rockdale	4	1.0
Rome	5	1.3
South Georgia	2	0.5
Southern	12	3.0
Southwestern	4	1.0
Stone Mountain	25	6.3
Tallapoosa	6	1.5
Tifton	8	2.0
Waycross	12	3.0
Western	18	4.5
<i>Total Death Notices: 400</i>		
<i>Percent of all judicial circuits filing a death notice: 96%</i>		