Title

Metrics of Mayhem: Quantifying Capriciousness in Capital Cases

ISBN

9781108498579

Author

Thaxton, Sherod

Publication Date

2020

UCLA School of Law

PUBLIC LAW & LEGAL THEORY RESEARCH PAPER NO. 20-28

METRICS OF MAYHEM: QUANTIFYING CAPRICIOUSNESS IN CAPITAL CASES

BY

SHEROD THAXTON
PROFESSOR OF LAW

In THE EIGHTH AMENDMENT AND IT'S FUTURE IN A NEW AGE OF PUNISHMENT 266-282 (Meghan Ryan & William Berry, eds., Cambridge University Press 2020)

Metrics of Mayhem: Quantifying Capriciousness in Capital Cases

Sherod Thaxton[†]

Introduction

Social science has figured prominently in debates over the constitutional administration of capital punishment for more than a half-century, especially with respect to capricious and bias decision-making.[1] The Court's ruling in *Furman v. Georgia*,[2] responsible for ushering in the modern era of capital punishment by invalidating then-existing capital statutes as violative of the Eighth Amendment's prohibition against cruel and unusual punishment, was replete with references to social scientific evidence documenting inconsistent, irrational, and discriminatory capital charging-and-sentencing practices.[3] *Furman* contained the most extensive discussion of social science research in any decision up to that point,[4] and all five Justices comprising the *per curiam* opinion in *Furman* agreed that evidence identifying the capricious imposition of the death penalty provided sufficient justification to impose an immediate moratorium on executions in the United States and the require the commutation of all death sentences for condemned inmates. Yet the weight given to this empirical evidence (or lack thereof) in judicial decision-making has, itself, been irrational and inconsistent.

Twelve years after *Furman*, in its landmark decision in *Pulley v. Harris*,[5] the Court rejected a capitally condemned inmate's claim that California's capital punishment statute violated the Eighth Amendment because it failed to mandate comparative (i.e., inter-case) proportionality review.[6] Many legal analysts perceived the Court as backpedaling on the very procedural

[†] Professor of Law, UCLA School of Law.

[1] *See, e.g.*, U.S. General Accountability Office, *Death Penalty Sentencing: Research Indicates Patterns of Racial Disparities*, GGD-90-57 (1990) (reporting that 82% (23 of 28) of all methodologically sophisticated studies examining capital punishment processes since the *Furman v. Georgia* ruling uncovered evidence of arbitrariness and bias) [hereinafter, "GAO"]; 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) (reporting that nearly all studies of the capital punishment process reveal arbitrary and discriminatory decision-making).

[2] 408 U.S. 238, 363 (1972) (per curiam).

[3] *Id. at* 363 (1972) (Marshall, J., concurring) (identifying social scientific evidence to support the claims that (1) most murderers are not sentenced to death; (2) most condemned inmates are not ultimately executed; (3) the death is administered in a racially and economically biased fashion; (4) innocent persons have been executed; and (5) many murderers become model prisoners and are law-abiding citizens upon release from prison)

[4] CRAIG HANEY, DEATH BY DESIGN: CAPITAL PUNISHMENT AS A SOCIAL PSYCHOLOGICAL SYSTEM 10 (2005) (noting that *Furman* contained over sixty 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.)

[5] 465 U.S. 37 (1984).

[6] Comparative proportionality review entails a comparison of a death sentenced inmate's case with sentences imposed in similar cases in order to determine whether the inmate's punishment is proportionate with respect to both the circumstances of the crime and the background of the defendant. Timothy V. Kaufman-Osborn, *Proportionality Review and the Death Penalty*, 29 JUST. SYS. J. 257, 258-263 (2008).

protections it had endorsed eight years early in *Gregg v. Georgia*,[7] when it seemingly identified comparative proportionality review—required under Georgia's capital statute—as a necessary component of constitutionally valid death penalty schemes.[8] *Gregg* emphasized that the Eighth Amendment, *inter alia*, forbade disproportionate punishment and required procedures that limit the risk of capriciousness,[9] and appellate review of legal actors' discretionary choices appeared to be part and parcel of any statutory framework capable of identifying and correcting constitutional errors. *Pulley* acknowledged that comparative proportionality review was a laudable feature of Georgia's statute and such review might be necessary if a capital statute was "so lacking in other checks on arbitrariness,"[10] but reasoned that it was not an indispensable feature.[11] The Court explained that Georgia's capital statute adequately limited and directed the jury's discretion without inter-case review via other procedural protections—namely, bifurcated proceedings (i.e., separate trials for determining guilt and punishment), a limited list of crimes and special circumstances required for death eligibility, and mandatory consideration of mitigation evidence at the sentencing phase.[12] The Court's identification of what it believed to be procedurally sufficient protections was perplexing because the initial capital statutes it approved as passing constitutional muster had not been subject to empirical scrutiny by the Court[13]; rather, the Court based its decisions solely on the assumption that the procedural protections contain therein would eliminate or significantly reduce the rampant capriciousness and bias that had been identified by social scientists and provided the basis for the Court's *per curiam* opinion in *Furman*.[14] The majority in *Gregg* was careful to note that its approval of any particular capital statute did not constitute a blanket endorsement of any similarly constructed statute and "each distinct death penalty system must be examined on an individual basis;"[15] however, the Court failed to clearly articulate any metric it would use to evaluate whether a system was, in fact, operating as the government promised or the pre-*Furman* mayhem persisted.[16]

---

[7] 428 U.S.153, 189 (1976) (plurality opinion).

[8] Georgia's capital statute was closely modeled after the American Law Institute's Model Penal Code. *Compare* 1973 Ga. Laws 74; Ga. Code Ann. § 27-2534 (1973) *with* Model Pen. Code §210.6.

[9] *Gregg*, 428 U.S. at 189 ("*Furman* mandates that…discretion must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action.).

[10] *Pulley*, 465 U.S. at 50.

[11] *Compare id*. at 45 *with* Walker v. Georgia, 555 U.S. 979 (2008) (Stevens, J.,) (noting that the Georgia Supreme Court "carried out an utterly perfunctory review" and a "breakdown in the statutory process" of Georgia's comparative proportionality review).

[12] *Pulley*, 465 U.S. at 45.

[13] Initial statues were approved in Georgia, Florida, and Texas. Gregg, 428 U.S. at 153; Proffit v. Florida, 428 U.S. 242 (1976); Jurek v. Texas, 428 U.S. 262 (1976).

[14] STUART BANNER, THE DEATH PENALTY: AN AMERICAN HISTORY 273 (2002) (noting that the petitioners from Florida, Georgia, and Texas who initially brought challenges against the post-*Furman* capital statutes in *Gregg* emphasized that the States neglected to provide any empirical evidence that these statutes could accomplish what the *Furman* majority unequivocally stated that the Eighth Amendment demanded).

[15] *Gregg*, 428 U.S. at 195.

[16] The Court's view is akin to Justice Potter Stewart's (in)famous statement concerning the definition of obscenity: "I know it when I see it." Jacobellis v. Ohio, 378 U.S. 184, 197 (1964)

Gregg foreshadowed the Court's capital punishment jurisprudence for the next four decades.[17] *Furman*'s strong concerns about actual outcomes of death penalty cases were deemphasized; *Gregg*'s assumption that the guided-discretion statutes would result in valid and reliable death sentences was left unexamined; and social science evidence that identified and documented the persistence of the pre-*Furman* era problems was either ignored or significantly minimized.[18] The Court's misplaced confidence in the rational, consistent, and nondiscriminatory administration of the death penalty was highlighted in Justice Brennan's forceful dissent in *Pulley*. According to the Justice, the Court was ignoring available evidence and "simply deluding itself, and also the American public, when it insists that those [capital defendants] have been selected on a basis that is neither arbitrary nor capricious, under any meaningful definition of those terms."[19] The Justice also (1) noted that the results of scholarly research necessary to support claims of capriciousness and bias "are being complied into a rapidly expanding body of literature," (2) identified no less than thirteen empirical studies demonstrating the unconstitutional (or constitutionally suspect) administration of capital punishment, and (3) emphasized that disproportionate treatment of capital defendants could only be eliminated after these disparities are identified by a procedural device such as inter-case proportionality review.[20] Thirty years after *Pulley*, Justice Brennan's sentiments were echoed by Justice Breyer in his landmark dissenting opinion in *Glossip v. Gross*.[21] Justice Breyer remarked that the "arbitrary imposition of punishment is the antithesis of the rule of law" and the urged the Court to reengage social and empirical facts concerning capital charging-and-sentencing practices.[22] Describing nearly four decades of social science evidence documenting the capricious and bias operation of capital punishment systems, Justice Breyer believed that the time was ripe for the Court to reopen the question of the constitutionality of the administration of the death penalty and invite full briefing that would allow the Court to scrutinize this relevant empirical scholarship with greater care than it had done it the past.[23]

---

[17] Sherod Thaxton, *Un*-Gregg-*ulated: Capital Charging and the Missing Mandate of Gregg v. Georgia*, 11 DUKE J. CONST. L. & PUB. POL'Y 145, 158 (2016) (describing the Court's repeatedly marginalization of social scientific evidence of the unconstitutional administration of the death penalty it the aftermath of *Gregg*).

[18] *Id*. *See also* McCleskey v. Kemp, 481 U.S. 279 (1987) (holding that statistical evidence of racially disproportionate death penalty charging and sentencing, even if believed, was insufficient to deem Georgia's capital statute unconstitutional as applied); Craig Haney & Deana D. Logan, *Broken Promise: The Supreme Court's Response to Social Science Research on Capital Punishment*, 50 J. SOC. ISSUES 75 (1994) (arguing that the Court has frequently ignored psychological data in its death penalty jurisprudence).

[19] *Pulley*, 465 U.S. at 62. *See also* Charles J. Ogletree Jr., *Black Man's Burden: Race and the Death Penalty in America*, 81 OR. L. REV. 15, 34 (2002) (explaining that not only have has the Court's incremental procedural fixes to the capital punishment system provide the appearance of legality and impartiality, but have done little to correct the rampant problems of caprice and bias).

[20] *Pulley*, 465 U.S. at 70

[21] 135 S. Ct. 2726 (2015). Justice Ginsburg joined Justice Breyer's dissenting opinion.

[22] *Id.* at 2760–62 (Breyer, dissenting).

[23] *Id.* at 2759. *Cf.* Pulley, 465 U.S. at 67 (Brennan J., dissenting) ("If the Court is going to fulfill its constitutional responsibilities, then it cannot sanction continued executions on the unexamined assumption that the death penalty is being administered in a rational, nonarbitrary, and non-capricious manner.")

Three years after *Glossip*, Justice Breyer issued a statement with respect to the denial of *certiorari* in a capital case, *Hidalgo v. Arizona*,[24] that appeared to build upon his prior dissenting opinion.[25] Justice Breyer gave an indication of the types of evidence to which the Court may be receptive in future cases challenging the constitutionality of capital punishment. Of particular relevance was that Justices Kagan and Sotomayor, along with Justice Ginsburg, joined Justice Breyer's statement in *Hidalgo*, whereas only Justice Ginsburg joined Justice Breyer's dissent in *Glossip*.[26] Given the four-vote requirement to grant *certiorari*, Justices Kagan and Sotomayor's endorsement of Justice Breyer's view of the growing relevance of social science to the Court's constitutional calculus makes it significantly more likely that a constitutional challenge to the death penalty, if supported by the requisite evidence, will be heard by the Court (and possibly garner at least four votes). In *Hidalgo*, the petitioner, Abel Daniel Hidalgo, challenged Arizona's capital statute because of, *inter alia*, its inability to meaningfully narrow the class of death-eligible defendants to comport with the requirements of Eighth Amendment as set forth in *Furman* and its progeny. Petitioner Hidalgo offered empirical evidence indicating that nearly 98% of first-degree murder defendants were death-eligible under Arizona's capital statute.[27] Justice Breyer noted that Hidalgo's "evidence is unrebutted[,] points to a possible constitutional problem[,] [and] warrants careful attention and evaluation."[28] The Justice, nonetheless, agreed with the denial of *certiorari* because the petitioner's "opportunity to develop the record through an evidentiary hearing was denied [and] the issue presented in this petition will be better suited for *certiorari* with such a record."[29]

Justice Breyer focused on legislative narrowing and was dismissive of the claim that "prosecutors may perform the narrowing requirement by choosing to ask for the death penalty only in those cases which a particularly wrongful first-degree murder is at issue."[30] But the Justice's position appears in tension with *Furman*'s emphasis on ensuring the rational and consistent imposition of the death penalty. Indeed, the majority in *Furman* distinguished its Eighth Amendment ruling in *Furman* from its Fourteenth Amendment due process ruling a year prior in *McGautha v. California*[31] based on the reasoning that a constitutionally permissible process could generate a constitutionally impermissible result;[32] that is, the Eighth Amendment's focus was on

---

[24] 138 S. Ct. 1054 (2018).

[25] *Compare id.*, with Glossip, 135 S. Ct. at 2726.

[26] Justice Sotomayor wrote a separate dissenting opinion, joined by Justices Breyer, Ginsburg, and Kagan.

[27] *See also* Steven F. Shatz & Nina Rivkind, *The California Death Penalty: Requiem for Furman*?, 72 N.Y.U. L. REV. 1283 (1997) (noting that California's capital statute fails to meaningfully narrow the class of death-eligible murders and 92% of non-death judgement first degree murder cases were factually eligible for the death penalty).

[28] *Hidalgo*, 138 S. Ct. 1054.

[29] *Id*.

[30] *Id*. ("that reasoning cannot be squared with this Court's precedent—precedent that insists that States perform the "constitutionally necessary" narrowing function "at the stage of legislative definition.") (citing Arizona, 390 P. 3d 791 (2017)).

[31] 402 U.S. 183 (1971).

[32] *But see Furman*, 408 U.S. at 376 (Burger, C.J., dissenting) (arguing that the issue in *Furman* had already been decided in *McGautha* and the Court should not revisit the issue so quickly after that ruling).

actual punishments (distributive justice) and not merely the process by which the punishment was decided (procedural justice).[33] Assessments of the rationality and consistency (or lack thereof) of outcomes a system produces requires a careful inquiry into decision-making beyond statutory narrowing. This is particularly true because the practical consequence of constraining the discretion of the sentencing authority through guided-discretion statutes was to increase the discretion of front-end and back-end actors in the process.[34] Opponents of the post-*Furman* capital statutes expressly argued that the legislatures were required to craft capital statutes that imposed greater justificatory and evidentiary burdens on prosecutors (charging and plea-bargaining) and on governors and pardon and parole boards (clemency). Although the Court dismissed these concerns, claiming that charging and clemency discretion were inevitable components of capital schemes and outside of the effective control of legislatures,[35] in the four decades since the Court's ruling, social scientific evidence has underscored that the Court's confidence in the rational, consistent, and nondiscriminatory administration of the death penalty *at any stage of the process* is unwarranted.[36] The "full briefing" of the empirical evidence on the operation of the death penalty that Justice Breyer invited in his dissent in *Glossip* must also include relevant evidence on capital charging dynamics. Even assuming, as Justice Breyer does, that "prosecutorial narrowing" cannot salvage a capital statute that fails to genuinely narrow the class of death-eligible murders, *Furman* suggests that irrational and inconsistent capital charging could render a capital statute constitutionally infirm—depending on its magnitude—apart from meaningful legislative narrowing. Justice Brennan emphasized in his dissent in *Pulley* that "we must worry whether, first, we have designed procedures which are appropriate to the decision between life and death and, second, whether we have followed those procedures."[37]

---

[33] HANEY, *supra* note 4, at 3-23 (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); Sherod Thaxton, *Disciplining Death: Assessing and Ameliorating Arbitrariness in Capital Charging*, 49 ARIZ. ST. L.J. 137, 147 (2017).

[34] BANNER, *supra* note 14, at 273.

[35] Gregg, 428 U.S. at 199.

[36] *See, generally*, DAVID C. BALDUS ET AL., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS (1990); Barbara O'Brien et al., *Untangling the Role of Race in Capital Charging and Sentencing in North Carolina, 1990-2009*, 94 N.C. L. REV. 1997 (2016); 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); 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); Stephanie Hindson et al., *Race, Gender, Region and Death Sentencing in Colorado, 1980-1999*, 77 U. COLO. L. REV. 549 (2006); William Alex Pridemore, *An Empirical Examination of Commutations and Executions in Post-Furman Capital Cases*, 17 JUST. Q. 159 (2000); 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); Robert E. Weiss et al., *Assessing the Capriciousness of Death Penalty Charging*, 30 LAW & SOC'Y REV. 607 (1996);; Marvin E. Wolfgang et al., *Comparison of the Executed and the Commuted among Admissions to Death Row*, 53 J. CRIM. L. CRIMINOLOGY & POLICE SCI. 301 (1962).

[37] *Pulley*, 465 U.S. at 69 (quoting John Kaplan, *The Problem of Capital Punishment*, 1983 U. ILL. L. REV. 555 (1983)).

When evidence of, *inter alia*, capricious capital charging decisions was presented to the Court three years after *Pulley*, in *McCleskey v. Kemp*,[38] the five-Justice majority rejected the petitioner's arbitrariness claim. Justice Powell, who authored the majority opinion in *McCleskey*, based his decision, in part, on (1) the perceived benefits that prosecutorial discretion offers criminal defendants in the form of individualized justice and (2) the unprecedented procedural safeguards designed to eliminate arbitrariness and bias.[39] Justice Powell also deemed the statistical evidence of race-based arbitrariness as insufficiently strong to rise to the level of a constitutional violation,[40] although he would subsequently remark that he had an extremely limited understanding of statistical analysis and regretted his decision in that case after he retired from the Court.[41] Yet as Justice Breyer emphasized in *Glossip*, it is equally true that individualized *injustice* is antithetical to the rule of law.[42] The Court has long held that discretionary choices cannot be left to a legal actors' inclinations, but to their judgement which must be guided by sound legal principles.[43] Claims of capriciousness and bias are predicated on the demonstration that legal actor's judgement is unconstrained (or inadequately constrained) by sound legal principles.

A key obstacle for litigants and scholars raising empirically anchored claims of capriciousness has been the absence of a workable definition from both capital statutes and the courts' analysis of them. How is this phenomenon to be measured and what threshold showing must be made before these claims of constitutional error become cognizable to the Court? The American Law Institute's explicit recognition that *Furman* and its progeny have been ineffective at guaranteeing a constitutionally permissible capital punishment system largely stems from the fact that neither the courts nor legislatures have described, intelligibly, how these systems may satisfy or fail the prevailing constitutional standards.[44] But the Court or legislatures need not be the sole source of guidance on this issue. The science of statistics, at its core, is concerned with drawing inferences from data in the face of various types of uncertainty, and therefore provides a useful framework for identifying and quantifying the types of errors relevant to the assessment of the constitutionality of the operation of death penalty systems. In other words, statistical social science facilitates the measurement of the extent of mayhem in capital charging-and-sentencing that the Court has, in theory, labored to minimize for the past five decades. The purpose of this essay is to describe

---

[38] 481 U.S. 279 (1987).

[39] *Id*. at 311. *But see* Pulley, 465 U.S. at 67 (Brennan J., dissenting) ("Simply to assume that the procedural protections mandated by this Court's prior decisions eliminate the irrationality underlying application of the death penalty is to ignore the holding of *Furman* and whatever constitutional difficulties may be inherent in each State's death penalty system.")

[40] Petitioner Warren McCleskey's race-based Eighth Amendment arbitrariness claim was distinct from his Fourteenth Amendment (equal protection) race-based discrimination; nonetheless the majority in *McCleskey* rejected both claims.

[41] JOHN C. JEFFRIES JR., JUSTICE LEWIS F. POWELL, JR.: A BIOGRAPHY 451 (1994).

[42] *Glossip v. Gross*, 135 S. Ct. 2726 (2015).

[43] United States v. Burr, 25 F Cases 30, 35 (Cir. Ct. Va. 1807). ("[A] motion to [the court's] discretion is a motion, not to its inclination, but to its judgment; and its judgment is to be guided by sound legal principles.").

[44] AMERICAN LAW INSTITUTE, REP. OF THE COUNCIL TO THE MEMBERSHIP OF THE AMERICAN LAW INSTITUTE ON THE MATTER OF THE DEATH PENALTY (2009) (withdrawing its endorsement of the Model Penal Codes capital punishment provisions).

several statistical techniques that are capable providing doctrinally-relevant and empirically-informed assessments of the degree of capriciousness in capital charging-and-sentencing systems that can be "scrutinized [by the Court] with greater care than it had done it the past."[45]

The ensuing discussion is organized into two parts. Part I maps the Court's Eighth Amendment jurisprudence on a conceptual and operational framework that is amendable to social scientific inquiry. Part II describes and applies a suite of statistical techniques well-suited to quantify the level of capricious in capital charging-and-sentencing decisions. In the interest of space, the application of these approaches focuses, specifically, on capital charging decisions, but the techniques are relevant for assessing *both* capital and non-capital charging-and-sentencing decisions.

I. Complementary Conceptions of Capriciousness

A. Capriciousness as Unreliability

From a statistical point of view, capriciousness may refer to the "degree of unpredictability or randomness in the output of any social system, *even if the same inputs are consistently applied*."[46] In the context of capital charging, "inputs" are the legally relevant defendant and crime characteristics, and the outputs are prosecutors' decisions to seek the death penalty in a case. Under this definition, capriciousness can take two general forms: similarly situated defendants are treated dissimilarly *or* dissimilarly situated defendants are treated similarly. A common term for this type of capriciousness is *unreliability* or *inconsistency*. Even though prosecutors are permitted to exercise their reasoned discretion in seeking the death penalty, *Furman* emphasized that a highly inconsistent death penalty system cannot be "fair and evenhanded," and is therefore unconstitutional.[47] The guided discretion statutes approved by the Court in *Gregg* were designed to rationalize the capital punishment process via constraining frontline legal actors—namely prosecutors, judges, and jurors.

It must be noted that the unreliability of a capital charging system may have causes that do not call into question the constitutionality of the death penalty. Even under a guided discretion regime, the application of rules to facts will always be subject to interpretation, so decision-makers may reasonable differ as to what constitutes similarly situated cases. So, for example, most capital statutes encompass murders that are "heinous, atrocious, and cruel" or "wantonly vile, horrible, and inhuman," but often there is no obvious line of demarcation and the Court's prior attempts to provide greater clarity to these terms appears to have only muddied the waters even more.[48] As a consequence, seemingly identical cases may be treated differently not simply because prosecutors are exercising their discretion in an inappropriate manner, but because of the problems inherent in the application of legal rules/categories to facts. Cases may also appear factually similar on all key

---

[45] *Glossip*, 135 S. Ct. 2726.

[46] Weiss et al., *supra* note 36, at 609 (emphasis added).

[47] *Furman*, 408 U.S. at 309 (Stewart, J., concurring).

[48] *Compare* Godfrey v. Georgia, 446 U.S. 420 (1980) *with* Arave v. Creech, 507 U.S. 463 (1993).

crime and defendant characteristics, but differ with respect to unique (and often formally irrelevant facts) that determine the outcome. This will, too, undermines consistency, but the permissibility of such factors will likely turn on the degree of overall inconsistency and, of course, whether these unique factors do not run afoul of the law (e.g., race/ethnicity, national orientation, sexual orientation, etc.). To be clear, a system that is highly consistent is not necessary a system that is fair and rational. The Court emphasized this point when it invalidated the mandatory death penalty statutes enacted by Louisiana and North Carolina.[49] These capital statutes violated *Furman* because, *inter alia*, they prohibited juries from considering mitigation evidence, which in the Court's view, made those sentences irrational.

B. Capriciousness as Irrationality

Capriciousness can also refer to a system's truthfulness or accuracy—i.e., its level of *rationality*/*validity*.[50] A rational system produces what it is purported and designed to produce. A system typically generates irrational results when it is incorrectly calibrated or actors implementing the system use different procedures and considerations.[51] A rational system, on the other hand, produces outcomes that are supported by theory and evidence.[52] In *Furman*, the majority stressed that constitutionally permissible capital punishment systems must be fair and even-handed, and the rationality of any assessment is closely related to its fairness.[53] This point was underscored by Justice Brennen in *Pulley* when he argued for the necessity of comparative case review.[54] Prosecutors' assessments of defendants' culpability should be consonant with the standards set forth in the applicable capital statutes because the statutes provide the prosecutor the criteria upon which culpability assessments are be made in order to comport with *Furman* and its progeny. Specifically, when the factors enumerated in a capital statute fail to account for prosecutorial charging decisions (within an acceptable range), then one can conclude that the system is operating irrationally. Improper crime and defendant characteristics that influence the functioning of a system also undermine its rationality because prosecutors, judges, and jurors are prohibited from considering those factors. In other words, when legally legitimate factors fail to sufficiently explain outcomes *or* legally illegitimate factors do explain outcomes (after accounting for the legally legitimate factors), there is evidence of an irrational decision-making process. Under an Eighth Amendment analysis, conscious discrimination on the part of decision-makers need not be demonstrated in order to substantiate a claim of an unconstitutionally capricious system.[55] Also,

---

[49] Woodson v. North Carolina, 428 U.S. 280 (1976); Roberts v. Louisiana, 428 U.S. 325 (1976).

[50] BRIAN FORST, ERRORS OF JUSTICE: NATURE, SOURCES, AND REMEDIES 33 (2004).

[51] Id. at 55.

[52] Burlington Truck Lines, Inc. v. United States, 371 U.S. 156, 168 (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)

[53] In the educational testing literature, assessments are deemed fair when they assess what is "taught." W. Steve Lang & Judy R. Wilkerson, Accuracy vs. Validity, Consistency vs. Reliability, and Fairness vs. Absence of Bias: A Call for Quality 13 (Feb. 2008) (unpublished manuscript) ("[V]alidity means that assessors are making justifiable interpretations about their data and good decisions.").

[54] *Pulley*, 465 U.S. at 60.

[55] *McCleskey*, 481 U.S. at 279.

as mentioned, *supra*, an irrational system does not necessarily imply an inconsistent system. Indeed, the imposition of the death penalty for non-homicidal rape was extremely consistent throughout American history: African American men accused of raping Caucasian women were almost invariably charged with the death penalty and sentenced to death.[56] On the other hand, an unreliable system will almost invariably imply an irrational system because inconsistent undermines accuracy.

II. Quantifying Unreliability and Irrationality

A. *Metrics of Inconsistency*

A metric of inconsistency captures the degree of instability of legal decision-makers across prosecutors. Numerous measures are available, but three are particularly useful in the capital charging context: *intraclass correlation coefficient; mean absolute deviation,* and *median odds ratio*.[57] Given the decentralized and county-centric nature of death penalty charging authority, these metrics are particularly useful for investigating hierarchically structured data (e.g., individual cases grouped by county), and the Court has emphasized the importance of explicitly examining both intra- and inter-jurisdiction processes when assessing the constitutionally of criminal punishments.[58] The intraclass correlation coefficient (ICC) measures the degree of stability of death charging decisions within a specific jurisdiction (i.e., county or similar administrative unit). When key case factors are taken into account, the ICC measures the total *residual* variability—that is, variability left unexplained by those factors included in the model as determinative of capital charging decisions. The ICC ranges from 0 to 1, and when the ICC is large, death noticing decisions are more consistent within jurisdictions. Conversely, when the ICC is small, charging decisions within a jurisdiction are inconsistent. There are no bright-line rules for interpreting the magnitude of the ICC statistic, but a general rule of thumb is that an ICC value above 0.7 is indicative of a highly consistent system; alternatively, a value below 0.4 suggests that the system is inconsistent.[59]

My examination of capital charging decisions in Georgia from 1993-2000, which included nearly 1,300 death-eligible homicides and considered more than 40 case-level variables pertaining

---

[56] M. Watt Espy & John Ortiz Smykla, *Executions in the United States, 1608-2002: The Espy File* (2004) (revealing that from 1608-1964, African American men constituted 90% of the 947 executions for non-homicidal rape); *see also see also William C. Bailey, Rape and the Death Penalty: A Neglected Area of Deterrence Research*, *in* CAPITAL PUNISHMENT IN THE UNITED STATES 226 (Hugo Adam Bedau & Charles M. Pierce eds., 1976) (explaining that, from 1930–1971, 50% of arrests for rape were among non-whites, but 90% of executions for rape (407 of 455) were of non-whites); Office of Planning and Analysis, Georgia Department of Corrections, *A History of the Death Penalty in Georgia, Executions by Year 1924-2014*, 5-13 (2015) (noting that from 1924-1961, 97% of executions for rape were African American men); *accord* Marvin E. Wolfgang & Marc Riedel, *Rape, Race, and the Death Penalty in Georgia*, 45 AM. J. ORTHOPSYCHIATRY 658 (1975) (African American males were much more likely to receive the death penalty than similarly situated Caucasian males even after taking into account 14 legally relevant case characteristics).

[57] Thaxton, *supra* note 33, at 173.

[58] Solem v. Helm, 463 U.S. 277, 290-92 (1983); Enmund v. Florida, 458 U.S. 782, 794-96 (1982).

[59] Domenic V. Cicchetti, *Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology*, 6 PSYCHOL. ASSESSMENT 284 (1994).

to the crime, defendant, and the victim,[60] revealed an ICC of 0.19 (or 19%)—well below the 0.4 cut-off indicating an inconsistent system. That is, 81% of the residual variability in death noticing for factually similar cases is attributable to within-jurisdiction dynamics and only 19% results from cases being tried in different jurisdictions.

The mean absolute deviation (MAD) provides the estimate of the variability in the probability that a similar situated case is noticed for the death penalty across jurisdictions (e.g., across counties).[61] The intuition behind this measure is that the overall inconsistency of a capital charging system can be obtain by determining the average magnitude of the difference in charging behavior across counties within a state for a typical case. Similar to the ICC, the MAD captures how (dis)similar prosecutors treat similarly situated defendants, but it does so in a slightly different manner. The MAD is especially useful when comparing a baseline model that does not include factors determinative of prosecutors' charging-decision (i.e., a "baseline") and an adjusted model that includes the relevant inputs. An adjusted model should yield a lower MAD than the baseline under the assumption that charging decisions in counties across the state will converge towards the statewide average for similarly situated defendants.

Based on my analysis of the Georgia data, the MAD for the baseline model was 0.11 (11%), and the adjusted model was 0.8 (8%). In other words, the jurisdictions in Georgia differed in their charging behavior for factual similar cases, on average, by 8 percentage points, and the inclusion of relevant case-level factors decreased the average statewide deviation by 3 percentage points. This finding is consistent with the results from the ICC analysis—most of the variability in charging behavior is attributable to within-jurisdiction charging inconsistency. Slightly more informative than the MAD is the actual range jurisdiction effects. The expected likelihood of a similarly situated case receiving a death notice ranged from .11 to .59 across jurisdictions in Georgia. In other words, the *probability* of receiving the death notice can be 5.36 times higher depending on the jurisdiction and, equivalently; the *odds* can be 11.5 times greater. This, of course, represents the extreme situation: going from the least punitive to the most punitive jurisdiction.

[60] The crime related factors include circumstances of murder (commission of felony, domestic altercation, other altercation, gang related, drug-related, sex-crime related, etc.), murder weapon, statutorily defined aggravating factors, motive for killing, confession evidence, weapon evidence, video evidence, and whether the defendant was ultimately convicted of murder. Defendant related factors are number of defendants, race/ethnicity, sex, age, level of education, employment status, marital status, number of children, military service, history of drug use, psychiatric status, IQ, Wide Range Achievement Test (WRAT), troubled family history, prior violent felony, prior felony conviction, prior murder conviction, trigger-person. Victim related factors included the number of victims, race/ethnicity, sex, age, relationship with defendant (stranger, intimate partner, family, friend). Thaxton, *supra* note 33, at 176.

Prior studies have noted in the inclusion of 200+ variables. DAVID C. BALDUS ET AL., *supra* 36, at 42-46 (1990); Raymond Paternoster et al., *Justice by Geography and Race: The Administration of the Death Penalty in Maryland, 1978-1999*, 4 MD. L.J. RACE, RELIGION, GENDER & CLASS 1, 57 (2004). It must be emphasized that I use a conservative counting method for the total number of variables. 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 count them as a single variable that indexes the total number of statutory aggravating circumstances present in the case.

[61] Thaxton, *supra* note 33, at 176.

The next measure I describe, the median odds ratio, provides an assessment for the typical situation.[62]

The final inconsistency measure, the median odds ratio (MOR), quantifies the variability between counties by comparing two charging decisions from two factually similar cases that are randomly chosen from two different counties.[63] Some analysts believe the MOR is more interpretable than the ICC or the MAD. The MOR is the typical (median) ratio between the cases with a higher and lower probability of receiving a death notice from different jurisdictions. Intuitively, the MOR captures the increased risk of death penalty notice that would occur if a case moved from one county to another. As with the MAD, comparing the baseline and adjusted measures of the MOR assists the analyst in determining the improvement in consistency when relevant case-level factors are considered. The MOR will always be greater than or equal to 1. If the MOR is 1, then there is no variation between jurisdictions because likelihood of receiving a death notice based on specific set of case characteristics is equal across counties. The MOR for the baseline and adjusted models for the Georgia data are, respectively, 2.03 and 2.32. So, on average, a case moving from a lower probability death charging jurisdiction to the higher probability death charging jurisdiction more than doubles the odds of receiving a death notice. Of greater significance, however, is that factually similar cases are treated more dissimilarly across circuits, as indicated by the higher, albeit slightly, MOR for the adjusted model.

B. Metrics of Irrationality

A useful and widely accepted statistic for assessing rationality is the *coefficient of determination* (R^2). This metric describes the improvement in the predictive ability of an adjusted model compared to a baseline model, and be interpreted as a proxy for the "rational connection between the facts found and the choice made."[64] Discretionary choices may not lend themselves to highly accurate statistical modeling, irrespective of the comprehensiveness of the model, because of idiosyncrasies associated with charging decisions.[65] Rather than indicating that a statistical model is improperly specified (e.g., the model excludes important determinants of capital charging decisions), a low R^2 may simply underscore the fact that a process is inherently capricious.[66] As noted, *supra*, capital charging decisions are jurisdiction-specific, so similar to metrics of consistency, appropriate metrics of rationality must account for the hierarchical nature of the data. This can be accomplished by disaggregating the variability in capital charging decisions into within- and between-jurisdiction components and assessing the level of rationality

[62] A very similar measure, the median absolute deviation, which is more resistant to extreme values from counties, may also be used. According to this metric, the unadjusted and adjusted models are, respectively, 0.09 and 0.07.

[63] Germán Rodriguez & Irma Elo, *Intra-Class Correlation in Random Effects Models for Binary Data*, 3 STATA J. 32, 43 (2003).

[64] *See* Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins., 463 U.S. 29, 43 (1983) (quoting Burlington Truck Lines v. United States, 371 U.S. 156, 168 (1962)) (defining arbitrary and capricious action in the context of administrative law).

[65] Sherod Thaxton, *Disentangling Disparity: Exploring Racially Disparate Effect and Treatment in Capital Charging*, 45 AM. J. CRIM. L. 95, 147 (2018).

[66] Id.

of each component.[67] Whereas the within-jurisdiction R^2 measures the proportion of variance accounted for in the case-level outcome by the case-level explanatory variables, the between-jurisdiction R^2 assesses the proportion of variance accounted for by the jurisdiction-level averages of those explanatory variables. So, e.g., in a rationally operating capital charging system, jurisdictions that tend to have cases which are more highly aggravated (i.e., greater defendant culpability) should be more likely, on average, to seek the death penalty. R^2 can also be calculated for subgroups and be used to assess whether the rationality of the charging decisions differs across these classifications.[68] Death noticing behavior might be said to be irrational, unfair, and uneven when associated with legally illegitimate considerations, such as race/ethnicity and gender.

The R^2 statistics for case- and jurisdiction-level dynamics in Georgia data, adjusting for relevant case characteristics, are 0.44 and 0.21, respectively. That is, the explanatory variables in the model explain, approximately, 44% of the variability within jurisdictions and 21% of the variability across jurisdictions. The ICC measure, discussed *supra*, revealed that location accounts for 19% of the variability in death charging, so approximately 4% ($.21 \times .19 = .039$) of the total variability in death noticing can be explained by between-jurisdiction differences in case-level factors. Several of the case-level characteristics included in the analysis may be deemed legally suspect, particularly for defendant related factors. One could argue, e.g., that it is improper for a prosecutor to consider defendants' race/ethnicity, gender, level of education, marital status, and employment status (to name a few) when deciding to seek the death penalty. The same may also be said for a victim's gender and race/ethnicity. In an attempt to determine how much Georgia's capital statute was constraining the behavior of prosecutors, a trimmed model was examined that only included (a) the number statutorily defined aggravating circumstances, (b) defendant's criminal history, (c) number of victims and defendants, and (d) the relationship between the defendant and victim. The premise of this analysis is that, in a rationally functioning system, the "legal core" of the case should, theoretically, drive capital charging.[69] The trimmed model explains 31% of the variance within jurisdictions and 12% of the variance between jurisdictions.

Finally, given the persistence of race-of-victim effects in capital charging-and-sentencing behavior—and to a lesser extent, race-of-defendant effects[70]—the R^2 was calculated for four groupings of cases: Caucasian defendants, non-Caucasian defendants, Caucasian victims, non-Caucasian victims.[71] The R^2 for the combined sample (i.e., all defendants) was 43%; whereas the R^2 was 59% for Caucasian defendant cases, 42% for non-Caucasian defendant cases, 51% for the

---

[67] STEPHEN W. RAUDENBUSH & ANTHONY S. BRYK, HIERARCHICAL LINEAR MODELS: APPLICATIONS AND DATA ANALYSIS METHODS 109-110 (2d ed. 2002) ("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 analyses and may lead to different conclusions.").

[68] DON HEDEKER & ROBERT D. GIBBONS, LONGITUDINAL DATA ANALYSIS 158, 195-96 (2006).

[69] *See* DONALD J. BLACK, SOCIOLOGICAL JUSTICE 20 (1989) (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.").

[70] Thaxton, *supra* note 65, at 120 (summarizing empirical literature on race-of-defendant and race-of-victim effects in capital charging and sentencing).

[71] The R^2 statistic needed to be re-scaled before comparisons across sub-groups in order for the comparison to be valid. *See* JOOP J. HOX, MULTILEVEL ANALYSIS: TECHNIQUES AND APPLICATIONS 133-134 (2d ed. 2010).

Caucasian-victim cases, 37% for non-Caucasian victim cases.[72] These results suggest that the identical case-level factors explain approximately 40% *more* of the overall variability in capital charging when the defendant or the victim is Caucasian as opposed to non-Caucasian.[73] In the Georgia data, approximately 98% of the death-eligible cases had either an African American or Caucasian defendant, and nearly 95% of the death-eligible cases had either an African American or Caucasian victim, so the "non-Caucasian" defendant and victim cases are, essentially, African American defendant and victim cases.

C. Some Qualifications about Quantification

Any assessment of capriciousness will depend on both the quality of the data and analytical model. The process of quantifying unreliability and irrationality will be impacted by (1) improperly measured inputs; (2) incorrectly specified relationships between the inputs; or (3) important omitted inputs. With respect to the first condition, there is a legitimate concern that the overall culpability of a defendant may be mismeasured because, as noted earlier, prosecutors may reasonably disagree over whether a factor is present in a case. One would expect this concern to be more relevant when assessing charging decisions inter-jurisdictionally rather than intra-jurisdictionally due to greater uniformity in the assessment of cases within a particular office, largely due to collaboration between attorneys when evaluating cases or office policies governing the assessment of cases. It is with emphasizing that 81% of the overall variability in capital charging in Georgia was attributable to within-jurisdiction dynamics, after accounting for important crime and defendant characteristics. This suggest that impact of mismeasured inputs is not driving the results.

The second condition pertains to the way the analyst posits the inputs impact the outcome (i.e., functional form). Is the relationship uniform through the range of a particular input (linear versus nonlinear)? Does the relationship between the input and the outcome depend on the value of another input (additive versus nonadditive)? Concerns over functional form can be addressed, to a large extent, by specifically examining how assumptions about (non)linearity and (non)additivity impact the measurement. Mild departures from the "true" functional form do not negative influence the assessment, but more dramatic departures may. Most of the inputs in the statistical model are binary (yes/no), so concerns over nonlinearity are not especially relevant. As for the non-binary inputs (e.g., number of co-defendants, victims, statutory aggravating circumstances, etc.), the Georgia data was reanalyzed using models that relaxed the linearity assumption for these variables and the results were similar. The impact of the non-additivity assumption was also examined for certain pairing of variables and, again, the results remained unchanged. It must be acknowledged, however, that investigating non-additivity can quickly become exceedingly complex when considering the interaction between more than three variables and these high-dimensional relationships were not explored, so this issue merits further investigation.

[72] Given that the vast majority of death-eligible cases are intra-racial, the higher R^2 statistics for the Caucasian-defendant cases and Caucasian-victim cases are likely a function of both of those factors.

[73] Race-of-defendant: $[(59 - 42) \div 42] \times 100 = 40.4\%$; race-of-victim: $[(51 - 37) \div 37] \times 100 = 37.8\%$.

The third condition may pose a significant problem for evaluating capital charging decisions if the model fails to include of all relevant variables that legitimately influence a prosecutor's decision. This concern is more pronounced for sentencing decisions than charging decisions because the important predictors of a death penalty charging have been well-established[74] and it has been demonstrated that the likelihood of finding novel and meaningful predictors is very low.[75] It is also worth noting that empirical investigations of capital charging across numerous states and time periods, employing various methodological approaches, and considering different inputs, have produced remarkably similar results.[76] Nevertheless, all of these caveats should be kept in mind not only when adopting the aforementioned metrics, but when engaging in any empirically-anchored assessment of capriciousness in capital charging-and-sentencing.[77]

Conclusion

Justice Breyer's recent invitations, in *Glossip* and *Hidalgo*, to potential litigants to develop and present social scientific evidence for the Court to assess when deciding on the constitutionality of the death penalty stands in stark contrast to Justice Scalia's (in)famous memorandum to his colleagues on the Court expressing his intention to vote against a constitutional challenge to Georgia's death penalty over thirty years ago in *McCleskey*.[78] *McCleskey* provided a vehicle for the Court to consider the most sophisticated and comprehensive statistical study ever conducted on the influence of race on the capital punishment, and the evidence overwhelmingly indicated rampant capriciousness and bias at every stage of the process. Justice Scalia informed his colleagues that he did not agree that *any* statistical showing of capriciousness and bias—irrespective of its magnitude—would justify ruling Georgia's capital statute unconstitutional. In the Justice's view, "irrational sympathies, including racial, on jury deliberations and prosecutorial decisions is real, acknowledged in the decisions of this court, and ineradicable, I cannot honestly say that all I need is more proof."[79] Even if some members of the current Court share Justice Scalia's belief that capriciousness and bias are inevitable and ineradicable features of capital punishment systems, it is also abundantly clear that at least three other members of the Court are in alignment with Justice Breyer with respect to the relevance of social scientific evidence for the Court's assessment of the constitutionality of the death penalty. Moreover, the U.S. Supreme Court is not the only plausible site of contestation—state supreme courts appear more willing to dismantle the death penalty based on statistical evidence of capriciousness and bias.[80]

---

[74] GAO, *supra* note 1.

[75] Weiss et al., *supra* note 36, at 617. *See supra* notes 65-66 and accompanying text (explaining that the low predictive power of a statistical model may not be indicative of a problem with the model, but instead reveals a process that is inherently capricious).

[76] Thaxton, *supra* note 65, at 122-123 (summarizing results from over a dozen studies of capital charging behavior and noting strikingly similar results).

[77] *See* McCleskey v. Zant, 580 F. Supp. 338, 361-62 (N.D. Ga. 1984) (describing potential deficiencies with statistical models of capital charging and sentencing decisions due to, *inter alia*, measured and omitted variables).

[78] Memorandum to the Conference from Justice Antonin Scalia in No. 84-6811, *McCleskey v. Kemp* (Jan. 6, 1987), *Thurgood Marshall Papers*, The Library of Congress, Washington, D.C.

[79] *Id*.

[80] *Compare* State v. Gregory, 427 P.3d 621, 636 (Wash. 2018) (ruling Washington State's death penalty unconstitutional in light of statistical evidence demonstrating that it was administered in an "arbitrary and racially biased manner") *with* State v. Loftin, 724 A.2d 129 (N.J. 1999) (reasoning that statistical evidence demonstrating

Consequently, how much capriciousness exists, whether the amount is intolerable, and what efforts should be taken to reduce it, remain questions of monumental importance.[81]

significant bias would provide adequate grounds for ruling New Jersey's death penalty unconstitutional). *See also* Note, "Washington State Supreme Court Declares Death Penalty Unconstitutional in Washington," 132 Harv. L. Rev. 1764 (2019) (noting that the Washington Supreme Court "became the first American court to declare the death penalty unconstitutional based primarily on statistical evidence of racial bias in sentencing").

[81] Weiss et al., *supra* note 36, at 625.